

From DEPARTMENT OF LEARNING, INFORMATICS,  
MANAGEMENT AND ETHICS  
Karolinska Institutet, Stockholm, Sweden

# **MEASURING HEALTH CARE PERFORMANCE**

**VARIATIONS IN CARE PROCESS, RESOURCE USE AND  
OUTCOMES IN CHILDBIRTH CARE**

Johan Mesterton



**Karolinska  
Institutet**

Stockholm 2019

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

Cover photo by Ann Båth, Akademiska University Hospital, shows the notebook in which birth care at the hospital was followed up between 1956 and 1987.

Printed by Eprint AB 2019

© Johan Mesterton, 2019

ISBN 978-91-7831-334-1

# MEASURING HEALTH CARE PERFORMANCE: VARIATIONS IN CARE PROCESS, RESOURCE USE AND OUTCOMES IN CHILDBIRTH CARE

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

**Johan Mesterton**

*Principal Supervisor:*

Professor Peter Lindgren  
Karolinska Institutet  
Department of Learning, Informatics,  
Management and Ethics  
Medical Management Centre

*Co-supervisor(s):*

Associate Professor Isis Amer-Wåhlin  
Karolinska Institutet  
Department of Learning, Informatics,  
Management and Ethics  
Medical Management Centre

Professor Mats Brommels  
Karolinska Institutet  
Department of Learning, Informatics,  
Management and Ethics  
Medical Management Centre

*Opponent:*

Associate Professor Thomas Bergholt  
University of Copenhagen  
Department of Obstetrics

*Examination Board:*

Professor Ove Axelsson  
Uppsala University  
Department of Women's and Children's Health

Associate Professor Birger Forsberg  
Karolinska Institutet  
Department of Public Health Sciences

Senior Lecturer Anna Glenngård  
Lund University  
Department of Business Administration



*To Arvid and Emil, by far the most amazing outcomes of Swedish birth care.*



## ABSTRACT

**Introduction:** Measurement of health care performance is important for quality assurance and improvement of health services. With regular monitoring and follow-up, unwarranted variations in performance can be reduced. For performance measurement to have the desired effects, meaningful indicators of performance need to be selected, methods that ensure comparability between hospitals need to be employed, and information must adequately be fed back to providers and professionals. Birth care is one of the most common causes of hospitalisation and large variations in practice have been observed which highlights the relevance of performance measurement in this area. More knowledge is needed on how to use indicators and appropriate methods for measurement of performance and how this information can be used to improve clinical practice.

**Aim:** The overall aim of this thesis is to show how routinely collected data can be used to measure health care performance in the area of birth care and to assess how such measurement can support quality improvement.

**Method:** Three of the four studies are based on quantitative analyses of a research database with extensive information on patient characteristics, care process, resource use, and health outcomes for almost 140 000 women giving birth. Regression analyses were employed to investigate the importance of case mix adjustment and the variations in performance between Swedish birth clinics. The fourth study is based on interviews with managers and staff in a hospital department to understand how they perceived the use of technology for feedback of performance data in improvement efforts.

**Findings:** Patient characteristics have a significant effect on birth care performance indicators and adjustment for differences in patient populations is a prerequisite for meaningful performance measurement. There are large variations in case mix adjusted performance between birth clinics in Sweden in terms of both the care process and the health outcomes achieved. If all clinics performed as the top 20%, around 2200 caesarean sections would be avoided annually in the regions studied. Similarly, almost 900 perineal tears of grade 3 or 4 and 1500 post-partum infections would be avoided. There are a number of factors that facilitate or hinder the adoption of technology for timely feedback of relevant performance data. Managers and staff perceive that such feedback of data supports quality improvement.

**Conclusions:** Adjustment for patient characteristics is a prerequisite for meaningful comparisons of performance between hospitals and can be used to analyse unwarranted variations. Analysis of case mix adjusted variations in performance between Swedish birth clinics reveals significant potential for improvement of outcomes and reduced costs. Continuous use of performance data can support quality improvement and lead to reduced variations in performance.

## LIST OF SCIENTIFIC PAPERS

- I. **Mesterton J**, Lindgren P, Ekenberg Abreu A, Ladfors L, Lilja M, Saltvedt S, Amer-Wahlin, I. Case Mix Adjustment of Health Outcomes, Resource Use and Process Indicators in Childbirth Care: a Register-based Study. *BMC Pregnancy and Childbirth*. 2016;16(1):125
- II. **Mesterton J**, Ladfors L, Ekenberg Abreu A, Lindgren P, Saltvedt S, Weichselbraun M, Amer-Wahlin, I. Case Mix Adjusted Variation in Cesarean Section Rate in Sweden. *Acta Obstetrica et Gynecologica Scandinavica*. 2017;96(5):597-606.
- III. **Mesterton J**, Brommels M, Ladfors L, Lindgren P, Amer-Wahlin I. Inter-hospital Variations in Health outcomes in Childbirth Care in Sweden: a Register-based Study. *International Journal for Quality in Health Care*. 2018;doi: 10.1093/intqhc/mzy153 [Epub ahead of print]
- IV. Tolf S, **Mesterton J**, Söderberg D, Amer-Wahlin I, Mazzocato P. How Can Technology Support Quality Improvement? Lessons Learned from an Obstetric Unit. *Manuscript*



# CONTENTS

1	Background.....	1
1.1	The challenge for health care .....	1
1.2	Performance management.....	2
1.3	Health care performance measurement and improvement.....	3
1.3.1	Accreditation and recertification .....	3
1.3.2	Public reporting .....	3
1.3.3	Pay for performance.....	4
1.3.4	Clinical audit, feedback and quality improvement .....	4
1.3.5	Potential risks associated with performance measurement.....	5
1.4	Methodological aspects of performance measurement .....	7
1.4.1	Selection and classification of indicators .....	7
1.4.2	Adjustment for differences in case mix .....	8
1.5	Challenges to the use of technology for performance measurement .....	8
1.6	Performance measurement and variations in childbirth care .....	10
1.6.1	Selection of indicators.....	11
1.6.2	Analyses of case mix impact and variations in caesarean section rate .....	12
1.6.3	Analyses of case mix impact and variations in other performance indicators .....	13
1.6.4	Performance measurement and improvement in childbirth care .....	14
1.7	Sveus.....	15
2	Aim .....	17
3	Methods .....	19
3.1	Data and study population.....	19
3.2	Definitions of indicators.....	19
3.2.1	Baseline characteristics (case mix factors).....	19
3.2.2	Performance indicators .....	20
3.3	Statistical analysis.....	21
3.4	Methods used in Study IV .....	21
3.5	Ethical considerations.....	23
4	Results.....	24
4.1	Descriptive statistics (Study I-III).....	24
4.2	Impact of case mix factors on important performance indicators (Study I) .....	26
4.3	Case mix adjusted variations in caesarean section rate (Study II) .....	27
4.4	Interhospital variations in health outcomes in childbirth care (Study III) .....	28
4.5	How technology can support quality improvement in birth care (Study IV).....	30
5	Discussion.....	32
5.1	Interpretation of findings.....	32
5.2	Strengths and limitations .....	35
5.3	Suggestions for further research.....	37

6	Conclusion .....	39
6.1	Study conclusions .....	39
6.2	Policy recommendations for performance measurement .....	39
7	Acknowledgements .....	41
8	References .....	43
9	Appendix .....	49

## LIST OF ABBREVIATIONS

BMI	Body Mass Index
CABG	Coronary artery bypass grafting
CS	Caesarean section
EBCOG	European Board of Gynaecology and Obstetrics
EU	European Union
ICD-10	International Classification of Diseases, 10 <sup>th</sup> edition
ICHOM	International Consortium for Health Outcomes Measurement
MBR	Medical Birth Register
NCQA	National Committee for Quality Assurance
NHS	National Health Service
OASIS	Obstetric Anal Sphincter Injuries
OLS	Ordinary Least Squares
OR	Odds ratio
PAS	Patient Administrative System
PREM	Patient Reported Experience Measure
PROM	Patient Reported Outcome Measure
QI	Quality Improvement
ROC	Receiver operating characteristic
UK	United Kingdom
WHO	World Health Organization



# 1 BACKGROUND

## 1.1 THE CHALLENGE FOR HEALTH CARE

Human health has improved tremendously over a long period of time, with life-expectancy and quality of life increasing dramatically world-wide: “The 20th century has seen a global transformation in human health unmatched in history.”[1] This development is in part due to rapid advancements in medicine and the provision of health care during the last hundred years. While health care is still continuously improving at a high pace, there are concerns over large variations in care practice and the increasing costs associated with health care provision. Today, we have a situation where costs of health care have been rising in most countries for several decades and are projected to increase further during the coming years. This development has been explained by factors such as demographic trends, new technologies, and increasing demands but the increase in health care costs has also been attributed to inappropriate use of medical services and unwarranted variations in clinical practice [2, 3].

Recent evidence suggests that both overuse and underuse of medical services are fundamental problems in health care delivery worldwide. For example, the rate of inappropriate total knee replacement rates has been estimated at 26% in Spain and 34% in the United States, and rate of inappropriate percutaneous coronary interventions and coronary angiography in Italy has been estimated to 22% and 30%, respectively [4]. The WHO has estimated that each year 6.2 million excess caesarean sections are performed worldwide [5]. Underuse of available medical services also contributes to substantial costs and patient harm. This is caused by limited access and availability of services, as well as failure of providers to deliver services and failure of patients to use them [6]. Research suggests that the average time for research evidence to reach clinical practice is 17 years [7]. One widely cited example of the slow diffusion of best practice in health care is a study by Eisenberg et al., which observed large variations in use of evidence-based medical therapy in patients undergoing percutaneous coronary revascularization in both the United States, Canada and Europe, and which found significant deviations from treatment guidelines [8].

The term “unwarranted variation” in health care has been attributed to John Wennberg [9]. As opposed to warranted variation, which are differences that can be attributed to the needs of the population served, unwarranted variation is defined as variations in the utilisation of health care services that cannot be explained by differences in patient characteristics or patient preferences. Unwarranted variations in health care have been pointed out as a major obstacle to achieving an effective health system where resources are put to the best possible use: “Unwarranted variation is unacceptable: it wastes resources, and it is the hallmark of poor-quality and lower-value healthcare. Investigating the causes of variation offers the opportunity of identifying and eliminating lower-value activity.” ([10], p 20)

## 1.2 PERFORMANCE MANAGEMENT

The fundamental goal of health systems is to improve the health of patients and citizens. The objective of performance measurement in health care is to evaluate and communicate to what extent the health care systems meet these key objectives. This often includes aspects such as health status achieved, responsiveness to patient preferences, and efficiency in the use of health care resources [11]. Ideas about organizational performance are largely rooted in a rationalist and mechanistic model of organizations. Given the large complexity of health systems and the number of stakeholders involved, analysis and management of organizational performance is challenging. However, performance management in health care has been gaining momentum due to a multitude of different factors such as decreased information asymmetry between different stakeholders in the systems, changes in social norms, the enormous development of IT systems for collection and analysis of data, and the increasing awareness of unacceptable variations in performance [12]. The importance of systematic analysis of performance and health system functioning is well phrased by the frequently quoted words of Donald Berwick: “Every system is perfectly designed to achieve the results it achieves” [13].

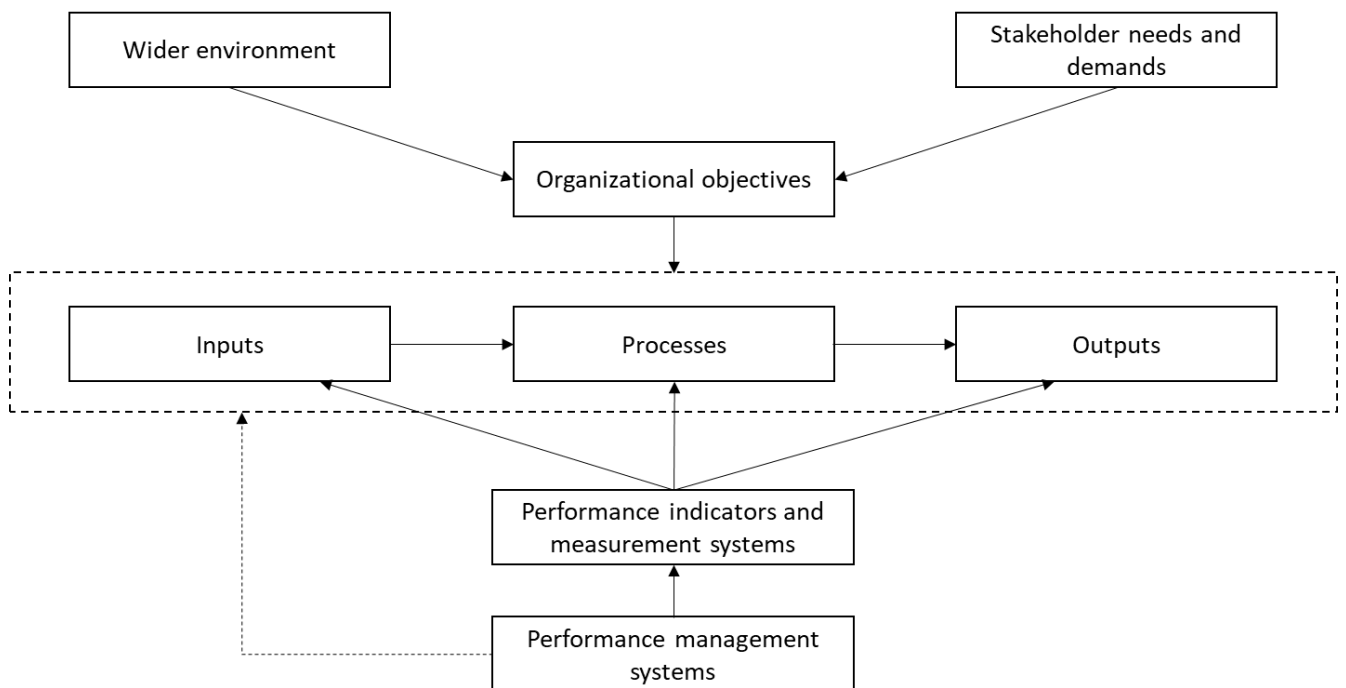


Figure 1 An analytic framework for performance management. Source: Adapted from Walshe [12]

Figure 1 presents a framework for performance management in health care. The organizational objectives are shaped both by the organization’s stakeholders (such as patients, providers and payers) and by the wider environment (such as societal norms, economic, and political drivers). The dotted box describes the organization’s functioning which consists of inputs (including facilities, human capital, knowledge, and information), processes (the way in which health care is delivered), and outputs (both in terms of number of patients treated, the volume of patient contacts, and also the outcomes achieved for those patients – such as

reduced morbidity and mortality, and improved health status and functioning). The Donabedian model for evaluating quality of care, which will be used in this thesis to classify performance indicators, categorizes performance measures into indicators of structure (Inputs), care processes (Processes), and outcomes (Outputs) [14]. Performance indicators and measurement systems serve to analyse the organizational performance in terms of both inputs, processes, and outputs, while performance management systems are mechanisms put in place to use indicators to influence the organization to change and improve.

### **1.3 HEALTH CARE PERFORMANCE MEASUREMENT AND IMPROVEMENT**

The practice of performance measurement is not a new phenomenon. There is evidence of collection of patient outcome data more than 250 years ago and already in the early 1900s Ernest Codman proposed a detailed system of follow-up, identification of best surgeons based on the actual results of their care, and interhospital comparisons [15]. However, given current demographic trends, increasing availability of technology, and increasing complexity of health systems, the need for improved performance measurement and transparency is as great as ever. Examples of large-scale initiatives for monitoring of performance and variations in health care include the Dartmouth atlas of health care in the United States [16], the NHS Atlas of Variation in Healthcare in the UK [10], and also Open comparisons (Öppna jämförelser) in Sweden [17].

Systematic analysis of hospital performance offers opportunities for identification of best practice and for clinical improvement. However, there is no automatic link between performance measurement and improvement in care provided by health care professionals. Several different possible methods (which can obviously be combined) for achieving clinical improvement through performance measurement have been discussed in the international literature and are presented below.

#### **1.3.1 Accreditation and recertification**

Performance measurement may be used as a requirement in accreditation of provider institutions and recertification of individual providers. In the United States, the NCQA and Joint Commission are two major accreditors who systematically collect performance data from both health plans and hospitals [18]. Since 1999, the International Society for Quality in Health Care also has an international programme that provides accreditation in over 40 countries for regulatory bodies and for organizations that perform accreditation and certification [19]. Owing to the wide variation in accreditation and certification and the complex context in which it is performed, there is a relative scarcity of evidence of its impact. There are, however, studies indicating that accreditation and certification does have a positive impact on clinical practices in several cases [20].

#### **1.3.2 Public reporting**

Public reporting of performance data is increasingly being used to drive health care improvement, with certain evidence that transparency around hospital performance to the

general public can improve quality. One commonly mentioned example of this is widely available risk-adjusted mortality rates following coronary artery bypass grafting (CABG) in the United States [21].

Berwick et al. [22] have investigated the link between reporting of health care performance data and quality improvement and discuss two possible mechanisms: Improvement through selection and improvement through changes in care. Improvement through selection works by patient flows reacting to the reported performance data. There are several different potential ways for patient flows to be affected by reporting of data, including individual patient choice, selective contracting by health care payers, or selective referrals by providers. Hence, this pathway can improve quality by getting the best out of the current distribution of performance. The second pathway, on the other hand, works by impacting the actual underlying distribution of performance. In the changes in care pathway, improved performance is achieved by altering behaviour and processes among those responsible for health care provision. While there is limited evidence for the selection pathway, there is a fair amount of evidence supporting the changes in care pathway [23].

### **1.3.3 Pay for performance**

Tying financial incentives to performance has also been widely employed to improve health care delivery. Conrad [24] describes several different dimensions along which financial incentives may vary, such as:

- The nature of incentive (reward or penalty)
- The target entity (group or individual)
- The type of incentive (general or specific)
- The magnitude of the incentive
- The certainty around incentive levels
- Frequency and duration of incentives (short-term or long-term)

There is an abundance of literature on the topic of linking financial incentives to performance. While “[t]here is no doubt that clinicians and other actors in the health system generally respond as expected to financial incentives” ([11], p 675), the evidence has been mixed on the exact effect of different financial incentives that have been put into practice. This should not come as a surprise given the enormous number of possible permutations of different design aspects described above. In fact, the topic of payment for performance has received such widespread interest in the literature that systematic reviews of systematic reviews have been conducted [25].

### **1.3.4 Clinical audit, feedback and quality improvement**

The practices of accreditation and recertification, public reporting, and pay for performance described above can be described as mechanisms through which governmental actors and payers impact providers through regulation, provider selection, and financial incentives. However, performance measurement can also be used to stimulate efforts for continuous



improvement and development. Many believe that measurement is better used for learning rather than for selection, reward, or punishment [13]: “[T]here will be no predictable and systematic progress in improving quality unless these professionals become engaged in collecting and using performance data to effect change” ([18], p 613). Providers may be more willing to change their practice if they receive feedback showing that they perform below some established benchmark or the norm. The practices of audit and feedback investigated in the literature have to a large extent been in the form of feedback to individual health care professionals on their performance relative to other professionals. The results of interventions involving this type of audit and feedback have been mixed and many different factors may impact their effect, including the baseline performance, the format of feedback (e.g. verbal or written), the source of feedback (e.g. supervisor or colleague, employer), the frequency of feedback, the profession of the recipient, and also whether feedback is accompanied by instructions for improvement and action plans [26].

Beyond feedback to individual professionals, measurement and feedback of performance is also an integral part of quality improvement (QI). QI has been defined as the combined and unceasing efforts of everyone to make changes that lead to better patient outcomes, better system performance and better professional development [27]. Most health care organizations claim to consistently engage in QI, but few QI efforts manage to consistently improve and sustain improvements over time [28].

Data is a fundamental component of QI, both in terms of informing where there is potential for improvement, offer a motivation for change, and to provide measurements that allow for evaluating the impact of implemented changes to care practice [27]. However, there are many characteristics of performance data that impact their usefulness for driving QI. One aspect that has been identified is timeliness of feedback. In a systematic review of how medical registries provide feedback to providers, van der Veer et al. found that timeliness of feedback from registers varied substantially, with time lags up to three years [29]. A systematic review of the application of the PDSA cycle for QI observed that a large majority of studies did not have access to data at monthly or more frequent intervals. As a consequence, it was not possible to continuously evaluate change [30]. Another critical aspect is data quality, with several studies observing that trust by users in the quality of the data used for feedback is of utmost importance [29, 31]. As discussed in section 1.4.2, adjustment for case mix has also been identified as an important factor for providing effective feedback [29, 31].

### **1.3.5 Potential risks associated with performance measurement**

There are several potential pitfalls related to the use of performance measurement. While some of these are clearly dependent on how the measurement is used to improve performance, there are some general themes in the literature. Potential risks have most frequently been discussed in relation to public reporting of performance data and when linking financial incentives to performance measures.

One potential risk with performance measurement that has been raised is that actors in the systems may be inclined to game data to receive financial gains or good results in public report cards. This gaming may include miscoding of diagnoses or excessive coding of comorbid conditions [23, 24].

Another factor that should be considered is the fact that focus on a certain number of indicators can lead to excessive emphasis on the specific indicators being measured and consequently that indicators not being measured could get ignored. There is certain evidence that the incentivization of specific measures may lead to a deterioration in un-incentivized measures. However, there is also some evidence that incentivization of certain measures can lead to positive spillover effects onto other measures [25].

Another issue that has been evoked in the literature is that performance measurement could lead to adverse selection of patients caused by providers having incentives to treat patients with better pre-requisites. This topic has been investigated extensively, not least in relation to the public reporting of mortality rates following CABG discussed above. While the conclusions from those investigations partially conflict, there have been reports that this public reporting led physicians to selecting patients with lower risk and that the access to care for severely ill patients was decreased [18, 23].

A potential risk associated with tying financial incentives to performance is that extrinsic incentives, such as financial ones, may crowd out intrinsic incentives (individuals' inner motivation to perform well). For that reason, the financial incentives should optimally be as aligned as possible with the inner motivation of the health care professionals [24, 25].

## 1.4 METHODOLOGICAL ASPECTS OF PERFORMANCE MEASUREMENT

### 1.4.1 Selection and classification of indicators

Identification of relevant indicators of performance is a fundamental component for meaningful measurement. Indicators are “succinct measures that aim to describe as much about a system as possible in as few points as possible” ([32], p 5). Indicators of performance measurement are often divided into three broad categories, following the Donabedian model [14, 33]:

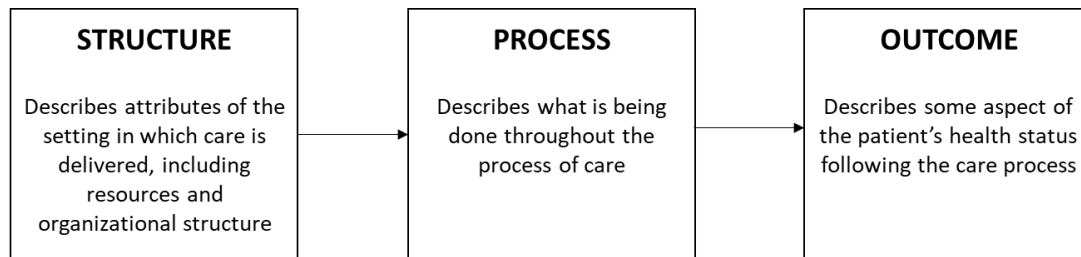


Figure 2 Description of the Donabedian model for measuring quality of care

Traditionally, performance measurement has focused on process indicators to a larger extent than outcomes that matter to patients [34]. Indicators of the care process are very relevant to measure adherence to guidelines and best practice. These indicators are also more directly impacted by the provider than patient health outcomes [35]. However, care process indicators do not have the same intrinsic value to patients as do indicators of health outcome and they also don't reflect the fundamental goal of the health system. Historically, outcomes measurement has largely focused on objective clinical endpoints, but the last decades have seen an increasing interest in the use of patient-reported outcome measures (PROMs) [34, 36]. Initiatives are also being taken to standardize outcome measures internationally for increased transparency and possibilities of benchmarking [34].

To ensure relevance of indicators in terms of capturing health care quality, the following criteria for selection of performance measures have been proposed by the Institute of Medicine [33]:

- Scientifically sound: The measure should be reliable and valid
- Feasible: The measure should be possible to collect without too high burden of measurement
- Important: The measure should be important in terms of health or resource use
- Alignment: The measure should optimally be aligned with existing measures to reduce redundancy and burden of reporting
- Comprehensiveness: The selected measures should jointly address the way the care is delivered and the quality

As described above, adjustment for patient characteristics is also important to ensure reliability and relevance performance measurement.

## **1.4.2 Adjustment for differences in case mix**

One obstacle in many previous initiatives for analysis of variations in health care has been their failure to deal with the differences in the patient population across regions and hospitals. This has led to discussions around whether variations are actually warranted and caused by variations in underlying patient characteristics. To overcome this, adjustment for case mix (sometimes also called risk adjustment) has been proposed. Case mix adjustment is a statistical method which allows data to be modified to control for variations in patient populations, making it possible to take these differences into account when resource use and health outcomes are compared, thereby allowing for “apples-to-apples” comparisons [33]. In cases where financial incentives are linked to performance measures, case mix adjustment can remove the incentives of selecting “easier” patients for treatment (cream skimming). When using performance measures are used for benchmarking, failure to adjust for differences in patient populations can hamper attempts to engage providers in meaningful dialogue about the potential for improving performance. In the absence of case mix adjustment, discussions with providers tend to centre around the unfairness of unadjusted data and the inability of data to reflect the underlying characteristics of their patients. This phenomenon has been dubbed the “my patients are sicker” syndrome [29]. Case mix adjustment aims at distinguishing factors that clinicians can control from factors that are outside their control, such as underlying patient characteristics, to understand actual performance and to stimulate discussions around improvement strategies [37]. A study performed in maternity units in England has concluded that use of benchmarking data for clinical improvement is contingent on data being reliable and adequately adjusted for differences in case mix [31].

Despite case mix adjustment being widely accepted and increasingly used, methodological advancements are still needed. However intuitive case mix adjustment may sound, “[t]his straightforward purpose belies the complexity of devising clinically credible and widely accepted risk adjustment methods, especially when resulting performance measures might be reported publicly or used to determine payments” ([37], p 251).

## **1.5 CHALLENGES TO THE USE OF TECHNOLOGY FOR PERFORMANCE MEASUREMENT**

The cover page of this thesis shows how the Akademiska University Hospital in Uppsala manually followed up their birth care in 1978 using a ballpoint pen and a notebook. One important driver for increased use of performance measurement is the technological transformation of our ability to capture, process, analyse and present data. The amount of data generated in all sectors of society is growing exponentially, with health care being one of the fastest-growing segments. A report from 2014 estimated that the total amount of data in the health care sector would increase from 153 exabytes in year 2013 to 2314 exabytes in year 2020 [38]. From a situation where health care documentation was completely paper-based, it is now possible to gather huge volumes of patient level data from different sources such as digital clinical records, other routine data sources and surveys. In addition to increasing

amounts of data being available, the capacity of processing data has advanced dramatically and techniques for making sense out of heterogeneous and complex data sets for measurement of performance have also improved [12].

However, despite the incredible technological advancements made over the last decades, healthcare has struggled to fully adopt technical innovations and make them an integral part of health care delivery. This problem may arise from a variety of different reasons, such as unwillingness of patients and providers to use the technology or difficulties in integrating with pre-existing technological solutions and established routines. Even when new technology is adopted, there are challenges associated with abandonment by individuals, or failure to scale up, spread and sustain innovations [39]. An explanation for this problem is that health care is becoming increasingly complex, with a multitude of different agents whose acts and interactions within a changing context are not predictable [40]. The inherent complexity of implementing technology in this health care setting is often underestimated. To enable a systematic assessment of barriers to adoption of technology in terms of complexity, Greenhalgh et al. developed the NASSS (Non-adoption or Abandonment of technology by individuals and difficulties achieving Scale-up, Spread and Sustainability) framework to help predict and evaluate the success of a technology-supported health care programme [39, 41].

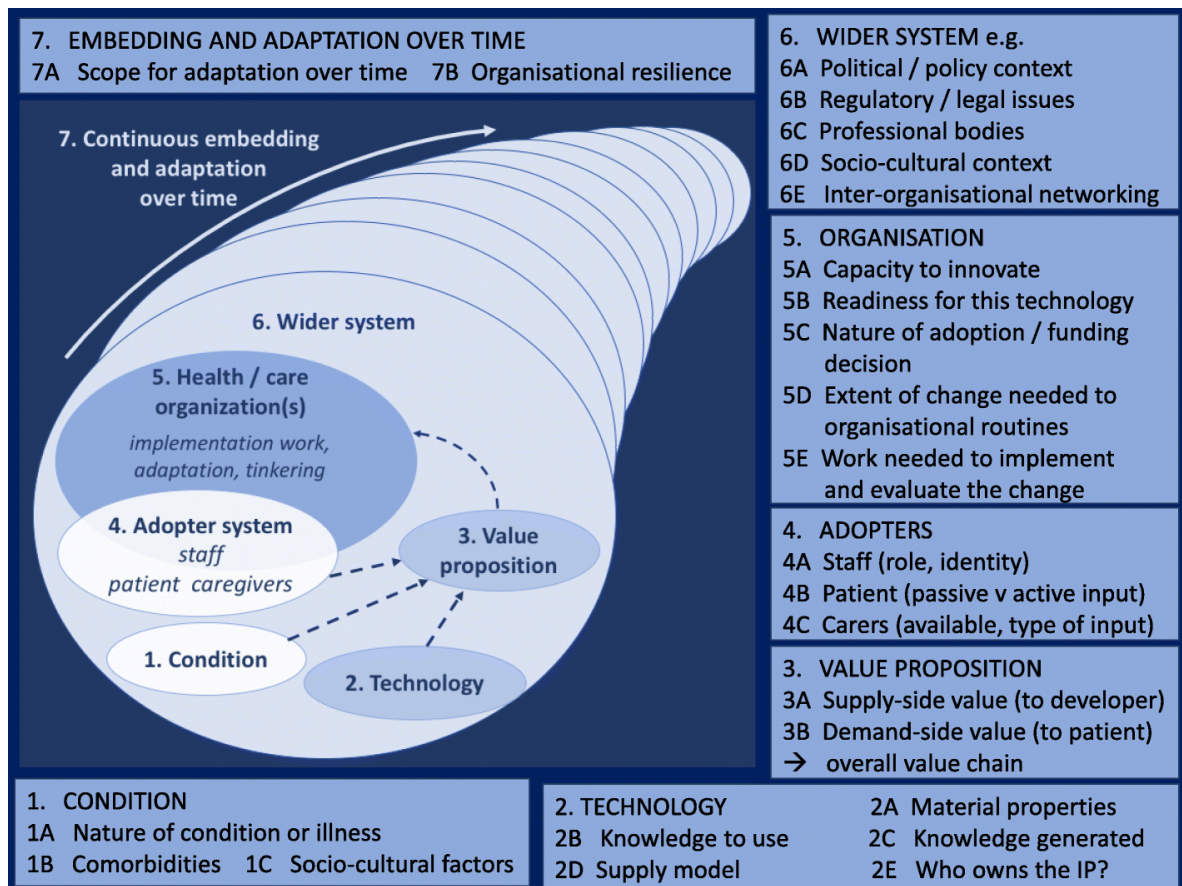


Figure 3 The NASSS framework for considering influences on the adoption, non-adoption, abandonment, spread, scale-up and sustainability of health and care technologies. Image reproduced with permission (<https://creativecommons.org/licenses/by/4.0/>) from Greenhalgh et al. BMC Medicine (2018) 16:66.

As illustrated in Figure 3, the framework consists of seven domains, each with several subdomains. Each domain can be classified as simple (“straightforward, predictable, few components”), complicated (“multiple interacting components or issues”) or complex (“dynamic, unpredictable, not easily disaggregated into constituent components”).

Technology-supported innovations face particular challenges in situations where there is a high degree of complexity in a large number of these domains:

1. **Condition:** Concerns the nature or characteristics of the illness that the technological innovation addresses.
2. **Technology:** Refers to the material and technical complexity of the innovation used, such as its design, the data it generates, ease of use, and also questions of how it fits with pre-existing technology and intellectual property.
3. **Value proposition:** Relates to the value of the innovation both from the supply-side (revenue for the developer) and the demand-side (strong evidence that the technology is desirable and adds value).
4. **Adopter system:** Concerns the need for intended adopters to learn new skills and take on new roles. The domain also includes patients’ or their caregivers’ knowledge and identity in innovation adoption.
5. **Organization:** Relates to the organization’s readiness to adopt new technology, how the decision to implement the technology into the organization was made and how that decision was motivated. Disruptions to established work routines and the amount of work required to adopt the new technology can also affect organizational response.
6. **Wider system:** Refers to the broader context in terms of policy, finance, the regulatory setting, and whether the legal status of new technology is unclear. This domain also includes the involvement of professional bodies as well as networking and knowledge sharing between organizations.
7. **Embedding and adaptation over time:** Refers to the possibility to adapt and coevolve technology and organizations using the technology to changing external conditions over time.

Use of technology for performance measurement and data-driven QI may involve complexity in all of the domains above and the framework provides a useful structure to understand under what circumstances technology-enabled performance measurement is more likely to succeed.

## 1.6 PERFORMANCE MEASUREMENT AND VARIATIONS IN CHILDBIRTH CARE

Pregnancy and childbirth is one of the most common causes for hospitalization in Europe [42] and giving birth constitutes an important event in most people’s lives. The large volumes of childbirth care and the associated costs, coupled with observed variations in practices and outcomes, nationally and internationally, make childbirth care an extremely relevant area for performance measurement.

### 1.6.1 Selection of indicators

In international comparisons of health system performance, maternal and neonatal mortality constitute the most widely used quality indicators of maternity care. However, in developed countries rates of maternal and neonatal mortality are low. In Sweden, the rate of neonatal mortality was 1.6 cases per 1000 live births in 2015, while the rate of maternal mortality was as low as 4 cases per 100 000 live births [43]. While these are still fundamental indicators to improve further, they need to be complemented with additional indicators of quality and performance to reflect the level of maternal care provided. In the words of the WHO Assistant Director-General for Family, Women, Children and Adolescents in 2018: “To meet Sustainable Development Goal 3 of ensuring healthy lives and promoting well-being for all at all ages, we cannot keep our focus solely on survival. High quality care for all pregnant women and their newborns, throughout pregnancy, childbirth and the postnatal periods, is essential to ensure that mothers and children both survive and thrive.”[44]

Numerous international and national organizations have developed recommendations for indicators to be used for quality assessment of maternity care. Escuriet et al. [45] provide an overview of proposed performance indicators throughout maternity care. The 10 most frequently measured events they identified in the literature are summarized below:

1. Caesarean section
2. Vaginal delivery with instrument (type of instrument)
3. Maternal postnatal complications
4. Perineal tears
5. Method of infant feeding
6. Induction and augmentation of labour
7. Vaginal delivery without instruments (may include normal birth)
8. Apgar score
9. Other neonatal complications
10. Mode of labour onset

The European Board and College of Obstetrics and Gynaecology (EBCOG) [46] have proposed the following set of indicators for assessment of intrapartum care:

1. Percentage of induction of labour
2. Mode of delivery by age groups and parity (spontaneous, operative vaginal, elective and emergency CS)
3. Percentage of women by parity receiving augmentation of labour with oxytocin infusion
4. Percentage of women by parity having episiotomy
5. Percentage of women by parity having grade three or grade four perineal tears
6. Rate of intrapartum stillbirths
7. Percentage of newborns with Apgar score below 7 at 5 minutes
8. Percentage of babies born with weight less than 1500 grams

9. Percentage of neonatal deaths attributable to congenital anomalies
10. Percentage of women with postpartum haemorrhage of 1000 ml or more and/or requiring transfusion
11. Percentage of women allowed trial of vaginal birth following previous caesarean section
12. Rate of women having emergency caesarean hysterectomy for severe postpartum haemorrhage
13. Rate of women requiring intensive care unit admission following delivery
14. Rate of babies with neonatal birth injury, neonatal encephalopathy and post-delivery transfer to neonatal intensive care unit

Given that maternity care covers different types of care over a long period of time (antenatal care, intrapartum care, postnatal care and neonatal care) it is not surprising that different initiatives for assessing quality have slightly different focus and thereby have slightly diverging proposals for exact indicators for assessment. However, in terms of intrapartum care, there is a relatively large consistency around the most relevant quality indicators, including mode of labour onset, mode of delivery and rates of maternal and neonatal complications.

As described above, the use of PROMs to capture information on patients' health and wellbeing has been widespread in many therapeutic areas. In maternity care, PROMs could be used to capture outcomes important to women with a holistic rather than strictly clinical assessment of their health throughout pregnancy, labour, and the postpartum period [47]. However, to date little work has been done in terms of systematic use of PROMs in maternity care even though initiatives are currently ongoing at international (through ICHOM [48]) and national levels (through the Pregnancy register). There is also increasing interest in measurement of Patient reported experience measures (PREMs). Women's experience of childbirth has significant short-term and long-term effects on well-being and health, and has also been shown to impact future reproduction [49]. There is a variety of different scales available for assessment of PREMs in relation to childbirth care, but systematic collection of such data is limited [50].

### **1.6.2 Analyses of case mix impact and variations in caesarean section rate**

The most frequently investigated aspect of childbirth care is the use of caesarean section (CS). Rates of CS has been increasing globally without obvious positive effect on health and with substantial variations between countries and between hospitals [51]. CS rate is a well-established indicator of maternity care quality [46, 52, 53] and in light of the maternal and neonatal complications and higher resource use that are associated with CS [54-57], it is an important indicator to understand performance.

Monitoring of CS rate at hospital level has often been performed using the Robson classification. The objective of the Robson classification is to provide a tool for clinical improvement, by allowing for comparisons between hospitals and increasing the



understanding of contributors to CS rates [58]. However, the Robson classification does not fully account for differences in case mix between hospitals and the need for using risk-adjusted CS rates in interhospital comparison has frequently been pointed out [46, 59, 60].

There is an abundance of literature available regarding both the impact of case mix on the risk of CS, as well as estimates of case mix adjusted differences between hospitals. One of the first and most frequently cited studies is the one by Aron et al. [61] from 1998. They included a broad set of clinical predictors and found significant effects of many obstetrical conditions and complications during pregnancy on the risk of CS. They also demonstrated the importance of incorporating pre-existing comorbid conditions in risk adjustment. Their study also highlighted important unwarranted variation in CS rates and showed that ranking hospitals after unadjusted and adjusted CS rates gave different results. Another early study from the United States was performed by Glantz et al. [62], who found factors such as nonvertex presentation, previous CS, parity, and a number of different complications during pregnancy were found to be associated with higher risk of CS. In addition to these early studies, several additional investigations of case mix adjusted CS rates in the United States have been published [63-65].

Numerous studies have also been performed in a European setting. Paranjothy et al. [66] performed a study of 216 maternity units in England and Wales and estimated that 34% of the difference in CS rates across units was attributable to differences in case mix. In another UK study, Bragg et al. [67] studied case mix adjusted variations in CS rates between NHS trusts. They observed factors such breech position, placenta praevia or placental abruption, parity and previous CS to be strong predictors of CS and concluded that "...comparing unadjusted rates of caesarean section should be avoided". An Italian study by Maso et al. [68] also reported strong effects of case mix factors such as maternal age, BMI, gestational age and parity on CS rate. Their analysis showed that risk stratifying based on the Robson classification provided a reasonably good adjustment but that the adjustment was improved by including additional maternal characteristics in the risk adjustment. Several studies of variations in CS rates have been performed in different Italian settings [60, 69, 70]. Elsewhere in Europe, variations in CS rates with case mix adjustment have been presented for Denmark [71], Finland [72], and Ireland [73] while data has also been reported for Australia [74].

### **1.6.3 Analyses of case mix impact and variations in other performance indicators**

As described above, CS rate is a universally accepted indicator of quality in maternity care. However, the extremely large dominance of CS as a birth care performance indicator is likely to a large extent driven by the relative ease with which it can be captured: It's a widely available indicator which can be identified through diagnosis codes and procedure codes available in essentially any patient administrative data and there is no ambiguity around its definition. However, while it is linked to outcomes and resource use [54-57], it is formally only a measure of the care process and despite being by far the most studied quality indicator

in childbirth care, it is by no means the only relevant indicator. To adequately assess and improve performance in childbirth care, a broad spectrum of indicators of care process, resource use, health outcomes, and experiences of women giving birth are needed. As discussed in section 1.6.1, several different indicators of relevance for studying childbirth care performance have been proposed. A number of these have been investigated in the literature.

In terms of health outcomes, a study from the United States investigated variations in both maternal adverse outcomes (postpartum haemorrhage, peripartum infection, perineal lacerations) and a composite measure of neonatal adverse outcomes [75]. That study observed age, BMI and obstetric history to be strong risk factors of virtually all adverse outcomes. The Royal College of Obstetricians and Gynaecologists in the UK has presented case mix adjusted data for obstetric anal sphincter injuries (OASIS) and emergency readmission [76]. Risk adjustment models for improved benchmarking have also been investigated for severe perineal tears in Australia [77]. Factors such as maternal age, nulliparity and higher birth weight were all found to be risk factors of severe perineal tears.

Induction of labour has been proposed as quality indicator because it has an impact on care process and outcomes [52]. A study from the United States demonstrated large variations in labour induction rates and observed factors such as preeclampsia, oligohydramnios, polyhydramnios and post-term pregnancy to be risk factors for induction of labour [78].

Length of hospital stay is an important indicator to understand resource use in relation to childbirth care. However, the impact of demographic and clinical risk factors on length of stay, and variations in length of stay between hospitals, has received limited attention in the literature. This could be due to a perception that length of stay is of lower clinical relevance compared to other indicators of care process and outcomes. Nevertheless, a Canadian study [79] demonstrated large interprovincial variation in length of stay, albeit with limited adjustment for case mix. A study from the United States identified significant variations in length of stay. That study thoroughly investigated the impact of case mix and found that most obstetric complications and pre-existing medical conditions were linked to a longer length of stay, demonstrating that adjustment for maternal characteristics are important for the relevance of such interhospital comparisons [79].

#### **1.6.4 Performance measurement and improvement in childbirth care**

There are few documented examples in scientific literature of targeted improvement programmes in childbirth care involving performance monitoring. One example, however, is ProvenCare Perinatal, a programme intended to reduce unwarranted variation and increase adherence to evidence-based care, where timely data feedback is an integral component. Early evidence indicated positive results from the programme, including a large decrease in CS rate at one of two studied medical centres [80]. Another example from the United States is the use of a report-generating computer system at California Pacific Medical Center. Intensive feedback on CS rates and outcomes was provided to the medical and nursing staff

at one intervention hospital and was compared to a control hospital where the system was introduced at a later stage. The primary endpoint investigated was CS rate, where the system was observed to contribute to reductions in CS rates from 24-25% to 20-21% [81]. In Sweden, Region Östergötland is the region with the lowest current rate of CS (around 12%). They had a CS rate of almost 19% in 2006 [82]. The largest unit in that region obtained a sharp decrease in CS rate through a structured programme including continuous monitoring, organizational changes and training [83].

## **1.7 SVEUS**

Sveus is a Swedish national programme which is central to this thesis and therefore warrants a brief introduction. Sveus was initiated in 2013 by the Ministry of Health and Social Affairs to develop value-based governance tools for the healthcare system. The background to the initiative was an ambition to collaborate cross-regionally to face common challenges such as increasing costs and variations in results. Among reasons for the cross-regional collaboration were to enable better possibilities for benchmarking and comparisons, to have more data available for development of case mix adjustment algorithms, and to share development costs in the project. Seven regions participated in Sveus (Jämtland Härjedalen, Östergötland, Dalarna, Uppsala, Skåne, Stockholm and Västra Götaland) covering around two thirds of the Swedish population. Sveus took inspiration from the framework value-based health care, which is a framework for health care management that has gained attention both internationally and in Sweden [84]. Within this framework, value is defined as patient relevant health outcomes achieved in relation to the costs of achieving those outcomes. Among proposed solutions for achieving higher value are improved transparency and reimbursement for entire cycles of care [85]. Initially, both reimbursement and transparency were core parts of Sveus, but over time reimbursement was deemed of lower relevance for a national collaboration. Therefore, the main focus of Sveus has been on development of methodology and systems to enable advanced analytics and continuous monitoring of case mix adjusted health care performance, with instant feedback to providers.

In 2016, a technical platform for continuous analysis and benchmarking of regions and hospitals was launched as part of Sveus. The platform is continuously fed with new data from administrative systems and quality registers and captures all health care contacts (primary care, outpatient specialized care and inpatient care) for around two thirds of the Swedish population. Both payers and providers have access to a web-based interface and can continuously monitor results in relation to resource, adjusted for differences in case mix.

Sveus as a project ended 31 December 2017 and the collaboration has now transitioned into a continuous phase. Throughout the five years, a total of around 150 participants and 50 different organizations were involved in Sveus. Childbirth care constituted one of eight initial patient groups in Sveus. A cross-professional expert group comprising representatives from professional organizations, payers, providers and quality registers were invited to participate in Sveus childbirth care, with the objective of designing a system for monitoring of childbirth care. A total of 17 meetings were held with the expert group from August 2013 through 2016.

As part of the work, the scope of the episode of care was defined and relevant indicators of health outcomes, care process, resource use, and patient characteristics were identified through a combination of previous literature, exploratory quantitative analysis, and expert opinion.

As part of the development of logic and methods for monitoring of childbirth care, extensive data analysis of historical data was carried out based on a comprehensive research database including data from regional Patient Administrative Systems (PAS), the Medical Birth Register (MBR), the Pregnancy Quality Register, the Neonatal Quality Register and sociodemographic data from Statistics Sweden.

## **2 AIM**

The overall aim of this thesis is to analyse unwarranted variations in health care and to assess how such measurement of performance can enable quality improvement. The specific research questions are:

1. What is the impact of patient characteristics on different indicators of birth care performance?
2. How much does the rate of caesarean section vary between Swedish hospitals after adjustment for differences in case mix?
3. How much do different indicators of health outcomes vary between Swedish hospitals after adjustment for patient characteristics and do hospitals that perform well on one indicator also perform well on others?
4. What factors can facilitate or hinder the adoption of technology-supported quality improvement as perceived by managers and staff at an obstetric unit?



## 3 METHODS

### 3.1 DATA AND STUDY POPULATION

Study I-III were all based on the Sveus childbirth research database. While Study I was focused on investigating the impact of case mix factors on different indicators of performance in childbirth care, Study II and III aimed at investigating interhospital variations in rate of caesarean section (II) and in health outcomes (III). The quantitative analyses for these studies were all based on the same underlying dataset. The database used for Study I-III was based on regional and national databases from 2009 to 2012. Women giving birth during 2011 and 2012 were identified using ICD-10 codes O80-O84 in PAS from the seven Swedish regions which participated in Sveus. For each woman, information on diagnoses and procedures in inpatient care and outpatient specialist care were extracted from PAS from two years before until 12 weeks after the date of admission for delivery. To capture maternal factors not available in PAS, such as parity, previous CS and BMI, as well as Apgar score for the newborn, information for these women was also extracted from MBR [86]. In addition, data from Statistics Sweden were used to capture information on maternal country of birth [87].

Extremely and very preterm deliveries (around 1% of all deliveries) and women who gave birth in a region different from the one they lived in at the time of delivery (around 3% of all deliveries) were excluded from analysis. A total of 140 296 deliveries during the study period were identified in PAS. For 99.6% of these a match in the MBR could be identified, resulting in 139 756 deliveries used for analysis.

### 3.2 DEFINITIONS OF INDICATORS

#### 3.2.1 Baseline characteristics (case mix factors)

Based on previous literature and clinical expertise, a number of baseline characteristics deemed relevant for case mix adjustment were identified. The objective was to determine the impact of patient characteristics outside of the birth clinic's control on indicators of care process, resource use and health outcomes. Hence, only factors that were deemed not to be the result of the delivery ward's care process were included as case mix factors.

*Table 1 Patient characteristics used for case mix adjustment*

Variable	Data source	Criteria
<i>Sociodemographic factors</i>		
Age	PAS	Age at admission for delivery
Born outside EU	Statistics Sweden	
<i>Obstetrical characteristics</i>		
First birth	MBR	
Previous caesarean section	MBR/PAS	Information in MBR or presence of codes O342 or O757 in PAS
Non-cephalic presentation	PAS	O321, O322, O641, O644, O801, O830, O831
Multiple birth	MBR/PAS	Information in MBR or presence of codes O84 in PAS

Premature delivery (w33+0-w36+6)	PAS	O601C eller O603C
BMI	MBR	BMI at first prenatal appointment
<u>Complications during pregnancy</u>		
Cervical insufficiency	PAS	O343
Infection of amniotic sac	PAS	O411
Pre-eclampsia	PAS	O14
Post-term pregnancy	PAS	O489
Gestational diabetes	PAS	O244
Polyhydramnios	PAS	O409
Oligohydramnios	PAS	O410
Placenta praevia	PAS	O44
Premature rupture of membranes	PAS	O42
Bleeding during pregnancy	PAS	O46
Placental abruption	PAS	O45
Herpes	PAS	N77
Intrauterine growth restriction	PAS	O365
Hepatitis	PAS	O266
<u>Comorbidities</u>		
Blood diseases	PAS	D500, D501, D508, D53, D55, D57, D58, D59, D6, D7, D8
Substance abuse	PAS	Z714, Z721, Z715, Z722, F10-F19
Endocrine and metabolic diseases	PAS	E00-E07, E1-E5, E60, E61, E63, E64, E67, E68, E8
Gynaecological diseases	PAS	N882, N883, D25, N99, N31-N37, Q51
Heart and vessel diseases	PAS	I80, I81, I82, I85, I86, I87, I0-I7, I9
Infectious diseases	PAS	A5-A9, B1, B2
Liver diseases	PAS	K70-K77
Lung diseases	PAS	J40-J47
Neurological diseases	PAS	G621, G50-G55, G57, G58, G59, G0-G4, G7, G8, G9
Renal diseases	PAS	N0, N1, N2, I120, I131, N250, Z490, Z491, Z492, Z940, Z992, N18, N19
Mental disorders	PAS	F204, F302, F312-F315, F341, F412, F432, F20, F22-F25, F28, F29, F32, F33
Musculoskeletal diseases	PAS	M05, M06, M3, M6
Bowel diseases	PAS	K50, K51, K52, K55-K59, K6
Tumour diseases	PAS	C1-C9

### 3.2.2 Performance indicators

To understand performance several different indicators were used in Study I-III. The objective was to adequately describe key measures of health outcomes, the resource use and care process. The performance indicators were selected based on a combination of review of previous research and clinical expertise. The indicators are listed below, along with data source and definition.



### Indicators of resource use and care process

- Length of hospital stay (PAS; calculated as date of discharge – date of admission)
- CS (PAS; ICD-10 O82, O84.2 or procedure codes MCA00,10,20,30,33,96)
- Induction of labour (PAS; ICD-10 O61 or procedure codes MAC10, DM002, DT027, DT036)

### Indicators of health outcomes

- Perineal tears of degree 3 and 4 (PAS; ICD-10 O70.2-3) in vaginal deliveries
- Haemorrhage >1000 ml (PAS; ICD-10 O67.8, O72) up to two weeks post-partum
- Post-partum infections up to 12 weeks following admission for delivery, including
  - cystitis (PAS; ICD-10 N30, O86.2)
  - endometritis (PAS; ICD-10 N71, O85.9)
  - other delivery-related infections (PAS; ICD-10 O86.0,3,4,8, Y95.9).
- Apgar <4 at 5 minutes (MBR)

## **3.3 STATISTICAL ANALYSIS**

In Study I, regression analysis was employed to assess the impact of case mix factors on the indicators of interest. Logistic regression was performed for all dichotomous outcomes, while the length of stay in days was analysed using ordinary least squares (OLS) regression. The full set of case mix factors was used in all regression models. To evaluate model fit for the logistic regression models, the c-statistic, calculated as the area under the receiver operating characteristic (ROC) curve, was used. Model fit for the regression model for length of stay was evaluated using R-square.

In Study II and III, each hospital's deviation from the mean rate of each studied indicator was estimated using a fixed effects logistic regression analysis with a dummy variable for each hospital. Effect coding was used for the hospital dummy variables so that each hospital's estimate used the overall rate (the population mean) as a reference value. To study potential correlations between different indicators, the Spearman correlation coefficient was used.

All statistical analysis was carried out using STATA 13.1 (STATA Corporation, College Station, TX).

## **3.4 METHODS USED IN STUDY IV**

Study IV was a qualitative study, conducted at the obstetric unit at a Swedish university hospital. The unit was selected because of its unique and innovative technology-supported QI programme. The unit used an analytics platform to support QI through continuous performance measurement and data feedback. The platform included data on observed and predicted (based on case mix) levels of indicators from the Sveus platform and additional data from local medical records to enable continuous tracking of local performance on a broad spectrum of indicators. Various dashboards for performance measurement were made available to the unit through web-interfaces and were all updated weekly to ensure timely

feedback. The rationale for choosing a qualitative design was to provide an in-depth understanding of how managers and staff at the unit perceived the adoption of technology for supporting QI.

Data was collected through three semi-structured focus group interviews conducted with informants (n=16) during September and October 2018. The focus group interview was used to enable the informants to collectively discuss their experiences concerning the topic and to stimulate reflections that might not have appeared in an individual interview [88]. The interviews were led by one facilitator and two researchers participated as observers. Interviews lasted 75 to 90 minutes and were audio-recorded and transcribed verbatim. The interview guide used is presented in the Appendix.

Purposive sampling was used to include 1) managers (n=4; 2 midwives, 2 physicians), 2) staff that were actively engaged in the QI programme (n=6; 2 midwives, 2 physicians, 2 assistant nurses) and 3) staff that were not actively involved in the QI programme (n=6; 2 midwives, 2 physicians, 2 assistant nurses). The sampling and the division of participants into separate groups were chosen to capture the dissemination of the programme, the perception of it at different organizational levels, and to avoid power imbalance that could hinder the participants' ability to speak freely [88].

Directed content analysis was used, which is a deductive methodology where existing theory or prior research about a phenomenon is used [89]. In this study the data collection and analysis were guided by the Nonadoption, Abandonment, Scale-up, Spread, and Sustainability (NASSS) framework [41]. The analysis was performed in five different steps: First, the entire interviews were read by two researchers to get an understanding of the material. Second, these two researchers independently condensed the first transcribed interview into coding units (reducing the number of words while staying close to the text) and compared results to ensure consistency in level of condensation. Then the two researchers each condensed one of interviews two and three. Third, three researchers sorted the coding units into the NASSS domains and subcategories were created. Fourth, the content of each of the seven domains was discussed and synthesized descriptions of the empirical data were developed by the research team. Fifth, based on these descriptions, each member of the research team participated in categorising each domain into "simple", "complicated", or "complex" in a process where each researcher presented their view, and discrepancies were discussed and resolved.

The Consolidated Criteria for Reporting Qualitative Research (COREQ), a 32-item checklist for interviews and focus groups, was used to ensure adequate reporting of the methodology and analysis [90].

### **3.5 ETHICAL CONSIDERATIONS**

Study I-III were all based on register-data. For these a study protocol summarizing background, data sources, variables, and proposed methods was submitted to and approved by a regional ethical committee. Subsequently, each data holder approved use of data for the study. In the scientific communication only aggregated data, which was not possible to link to any specific individual, was presented. The Stockholm Regional Ethical Review Board approved the study protocol (Dnr 2013/447-31/5, 2013/1686-32) for Study I-III. Study IV was a qualitative study based on focus group discussions. Each respondent signed an informed consent form prior to the interview. The Stockholm Regional Ethical Review Board approved the study protocol (Dnr 2018/1644-31/5).

This research has been carried out as part of my employment in Ivbar, which is a company that develops analytics products for the health care sector. While this constitutes a potential conflict of interest, I also believe that combining research with work in the area of performance measurement has helped me gain a better understanding of the practical aspects of performance measurement and QI. Getting hands-on experience of how both payers and providers work with data day-to-day and make decisions, has helped me understand the possible applications of the sometimes more theoretical research activities.

Study I-III are relatively general in nature and are not directly related to any Ivbar product. Study IV, however, studied a technology-supported QI programme where a technical solution developed by Ivbar was used. My personal involvement in that particular QI programme was limited, but I did meet the clinic at several occasions early in the process to discuss the scope and design of the work. To ensure that my personal involvement in the work and my role at Ivbar did not influence the discussions and interactions among the interviewees, I did not participate in any of the interviews.

## 4 RESULTS

### 4.1 DESCRIPTIVE STATISTICS (STUDY I-III)

Table 2 presents the number of deliveries per hospital in the seven included regions.

Table 2 Regions and hospitals included in the analysis

Region	Population*	Hospital	Abbreviation	Number of deliveries
Jämtland Härjedalen	126 147	Östersund Hospital	RJH ÖSD	2276
Östergötland	433 462	University Hospital Linköping and Vrinnevi Hospital Norrköping	RÖ LN	8956
Dalarna	276 379	Falun Hospital	LTD FL	4728
Uppsala	341 465	Akademiska University Hospital	LUL AS	7010
Skåne	1 262 068	Helsingborg Hospital	RS HBG	5652
		Central Hospital Kristianstad	RS CSK	3258
		Skåne University Hospital Lund	RS SUSL	6321
		Skåne University Hospital Malmö	RS SUSM	8757
		Ystad Hospital	RS YST	2171
Stockholm	2 123 337	BB Stockholm	SLL BBSth	7325
		Danderyd Hospital	SLL DS	12 148
		Karolinska University Hospital Huddinge	SLL K Hudd	8370
		Karolinska University Hospital Solna	SLL K Solna	7475
		Södertälje Hospital	SLL STS	2880
		Stockholm South General Hospital	SLL SÖS	13 616
Västra Götaland	1 598 700	Södra Älvsborg Hospital	VGR SÄS	4810
		NU Hospital group	VGR NU	5345
		Skaraborg Hospital	VGR SkaS	3897
		Sahlgrenska University Hospital Mölndal	VGR SU/M	7036
		Sahlgrenska University Hospital Östra	VGR SU/Ö	10 832

\*2012 Population numbers from Statistics Sweden [91]

Table 3 below presents descriptive statistics for the patient population used for analysis in Study I-III. Mean age at admission for delivery was 30.7 years, while 21% of women were born outside the EU. Slightly below half of women were nulliparous and 10% had a previous caesarean section. Mean BMI at time of first visit to maternity care was 24.5 kg/m<sup>2</sup>. This case mix factor was missing for 5% of the population. The most common complicating factor during pregnancy was post-term pregnancy, followed by intrauterine growth restriction and pre-eclampsia.

In terms of care process, the overall rate of CS and induction of labour was 17% and 15%, respectively, while mean length of stay was 2.6 days. Among vaginal deliveries, 3.6% resulted in a degree 3 or 4 perineal tear. Haemorrhage over 1000 ml occurred in 7.8% of deliveries. The rate of post-partum infection was 4.2%. Among post-partum infections, endometritis was most common (2.2%), followed by other infections and cystitis (1.5% and 0.7%, respectively). An Apgar score below 4 at 5 minutes was observed in 0.32% of deliveries.

Table 3 Descriptive statistics of patient characteristics and performance indicators

Category of indicator	Sub-category of indicator	Indicator	Population average	Number of observations
Patient characteristics	Sociodemographic factors	Age (mean;sd)	30.7;5.2	139756
		Born outside EU	21.3%	139756
	Obstetrical factors	BMI (mean;sd)	24.5;4.7	132917
		First birth	44.9%	139701
		Previous caesarean section	10.0%	139756
		Other position than head first	3.5%	139756
		Multiple birth	1.3%	139756
		Premature	4.2%	139756
		Complications during pregnancy	Cervical insufficiency	0.3%
	Infection of amniotic sac		0.2%	139756
	Pre-eclampsia		3.4%	139756
	Post-term pregnancy		5.0%	139756
	Gestational diabetes		2.0%	139756
	Polyhydramnios		0.5%	139756
	Oligohydramnios		1.9%	139756
	Placenta praevia		0.7%	139756
	Premature rupture of membranes		1.9%	139756
	Bleeding during pregnancy		3.0%	139756
	Herpes		0.6%	139756
	Intrauterine growth restriction		3.7%	139756
Hepatitis	0.8%		139756	
Placental abruption	0.3%	139756		
Comorbidities	Number of comorbidities (mean;sd)	0.23;0.52	139756	
Care process	Care process	CS	16.9%	139756
		Labour induction	15.0%	139756
Resource use	Resource use	Length of stay (mean;sd)	2.60;1.98	139756
Health outcomes	Maternal outcomes	Perineal tear grade 3 or 4	3.6%	114638
		Haemorrhage>1000 ml	7.8%	137940
		Postpartum infection	4.2%	126387
	Neonatal outcomes	Apgar<4 at 5 minutes	0.32%	139198

## 4.2 IMPACT OF CASE MIX FACTORS ON IMPORTANT PERFORMANCE INDICATORS (STUDY I)

Six different indicators of performance were assessed. Measures of health outcomes included perineal tears of degree 3 and 4 in vaginal deliveries, haemorrhage >1000 ml, as well as post-partum infections up to 12 weeks following admission for delivery. The two care process indicators analysed were CS and induction of labour, while length of hospital stay was used as resource use indicator.

A large number of baseline characteristics deemed relevant for predicting outcomes and resource use were identified based on previous literature and based on clinical expertise (see exact list and definitions in Table 1).

The importance of different case mix factors varied across the six performance indicators. Higher maternal age was a consistent risk factor and was observed to increase both length of stay and rates of labour induction, CS, perineal tears, and haemorrhage. Being born outside of the EU was also a risk factor for labour induction and CS, as well as longer length of stay and adverse obstetric events. Obstetrical characteristics such as nulliparity and previous CS were strongly associated with higher risk of CS, longer hospital stay and with higher risk of severe perineal tears, haemorrhage and infections. Multiple birth had a strong impact on risk of haemorrhage. However, premature delivery was associated with significantly lower risk of both perineal tears and haemorrhage.

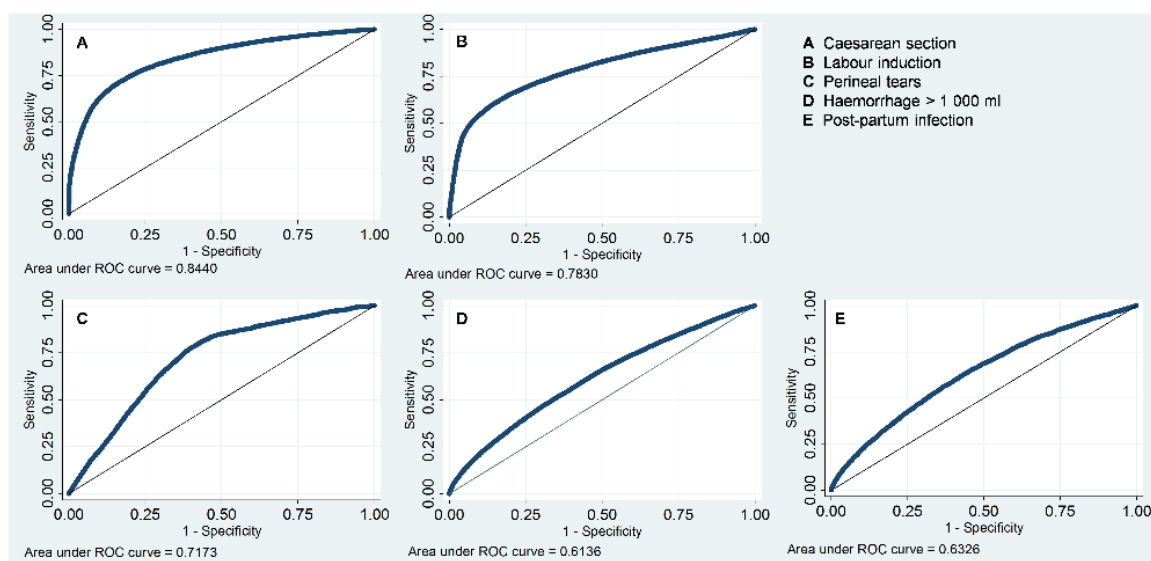


Figure 4 Receiver operating characteristic curves of the five models with dichotomous outcomes

Generally, the impact of case mix was largest for indicators of care process. Both for caesarean section, labour induction and length of stay, many of the case mix factors showed statistically significant effects. The predictive ability was lower for health outcomes. Figure 4 illustrates the ability of the models to predict which patients received CS, labour induction or experienced the different adverse outcomes. The predictive ability was highest for CS (c-statistic 0.84), followed by labour induction (c-statistic 0.78), indicating that the model could reasonably well predict these indicators. For perineal tears the c-statistic was 0.72,

while the predictive ability was lower for haemorrhage and post-partum infections (c-statistics of 0.61 and 0.63, respectively), indicating that the maternal characteristics included had a smaller impact on these two indicators. For length of stay, the regression model was able to explain 28% of the variation.

### 4.3 CASE MIX ADJUSTED VARIATIONS IN CAESAREAN SECTION RATE (STUDY II)

The second study was based on same research database that was used for Study I. The 23 different sociodemographic and clinical characteristics investigated in Study I were used for case mix adjustment. Analyses were performed for the entire study population as well as for two different subgroups: nulliparous, cephalic, full-term, singletons (Robson groups 1 and 2) and multiparous, cephalic, full-term, singletons, with no previous CS (Robson groups 3 and 4). The potential relationship between CS rate and neonatal outcome was investigated using Apgar score<4 at 5 minutes.

The mean CS rate was 16.9% and as Figure 5 shows the observed hospital-level CS rate varied from 12.1% to 22.6%. Differences in case mix caused large variations in expected rate of CS between hospitals, from 13.7% at Skaraborg hospital to 19.9% at Karolinska university hospital Solna. Among the 20 hospitals, 7 had significantly lower case mix adjusted CS rate compared to the population mean, while 6 had significantly higher rate and 7 hospitals did not have a statistically significant deviation from the population mean.

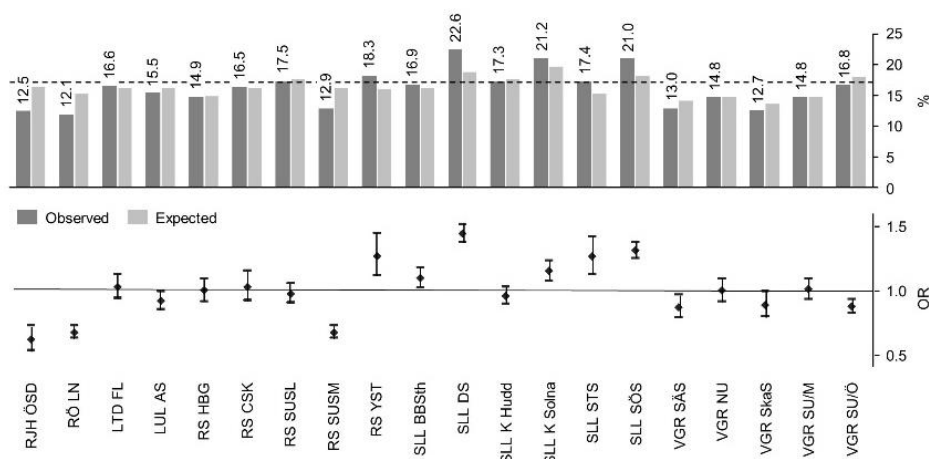


Figure 5 Observed caesarean section rate, expected caesarean section rate and case mix adjusted deviation from mean caesarean section rate (all deliveries). (Note: The odds ratio (OR) represents each hospital's deviation from the weighted average of all hospitals)

In general, the patterns were relatively similar in the two subgroups investigated. Hence, hospitals with a higher than expected CS rate in one subgroup tended also to have higher than expected CS rate in the other subgroup (correlation coefficient 0.44; p=0.052).

If all hospitals had performed as many CS as the 20% of hospitals with lowest case mix adjusted rate, it would have caused a reduction of 2200 CS annually in the regions included in the study, which would have entailed cost savings of around €7M annually during the initial admission for delivery alone.

Figure 7 shows the relationship between case mix adjusted CS rate on one hand and case mix adjusted rate of low Apgar score on the other. As the figure shows, there were variations between hospitals both in the rate of CS and the proportion newborn with low Apgar, but there was no evidence of relationship between the two (correlation coefficient 0.020; p=0.93).

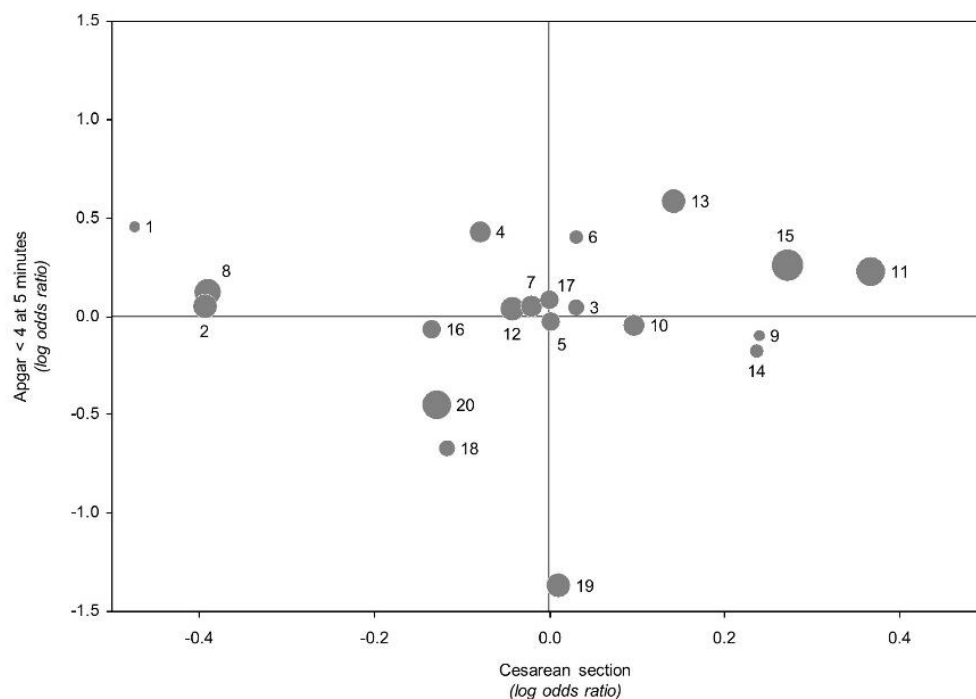


Figure 6 Association on hospital level between CS rate and proportion newborn with Apgar<4 at 5 minutes, both adjusted for case mix (Note: The log OR represents each hospital's deviation from the weighted average of all hospitals. Each bubble represents a hospital. The area of the bubble corresponds to the number of deliveries during the time period.)

#### 4.4 INTERHOSPITAL VARIATIONS IN HEALTH OUTCOMES IN CHILDBIRTH CARE (STUDY III)

This study used the same research database as Study I and II and employed the same case mix factors and statistical methodology as Study II to study variations in performance. Four indicators of health outcomes, all of which have been deemed relevant indicators of quality of maternity care services [45, 46], were selected for analysis.

As presented in Figure 7, statistically significant interhospital variations were observed for all studied indicators. There was no hospital which had a statistically significant higher or lower rate on all four outcomes indicators. The top 5 hospitals for each indicator were compared to the remaining 16 hospitals. If these 16 hospitals had performed as the case mix adjusted



average of the 5 hospitals that performed best for each indicator, a total of 890 OASIS (-23%), 2700 haemorrhages (-26%), 1500 infections (-30%) and 180 newborns with Apgar<4 at 5 minutes (-42%) would have been avoided over the two years studied.

There was a slight positive correlation between performances across different indicators of health outcomes, even though this was not statistically significant in most cases.

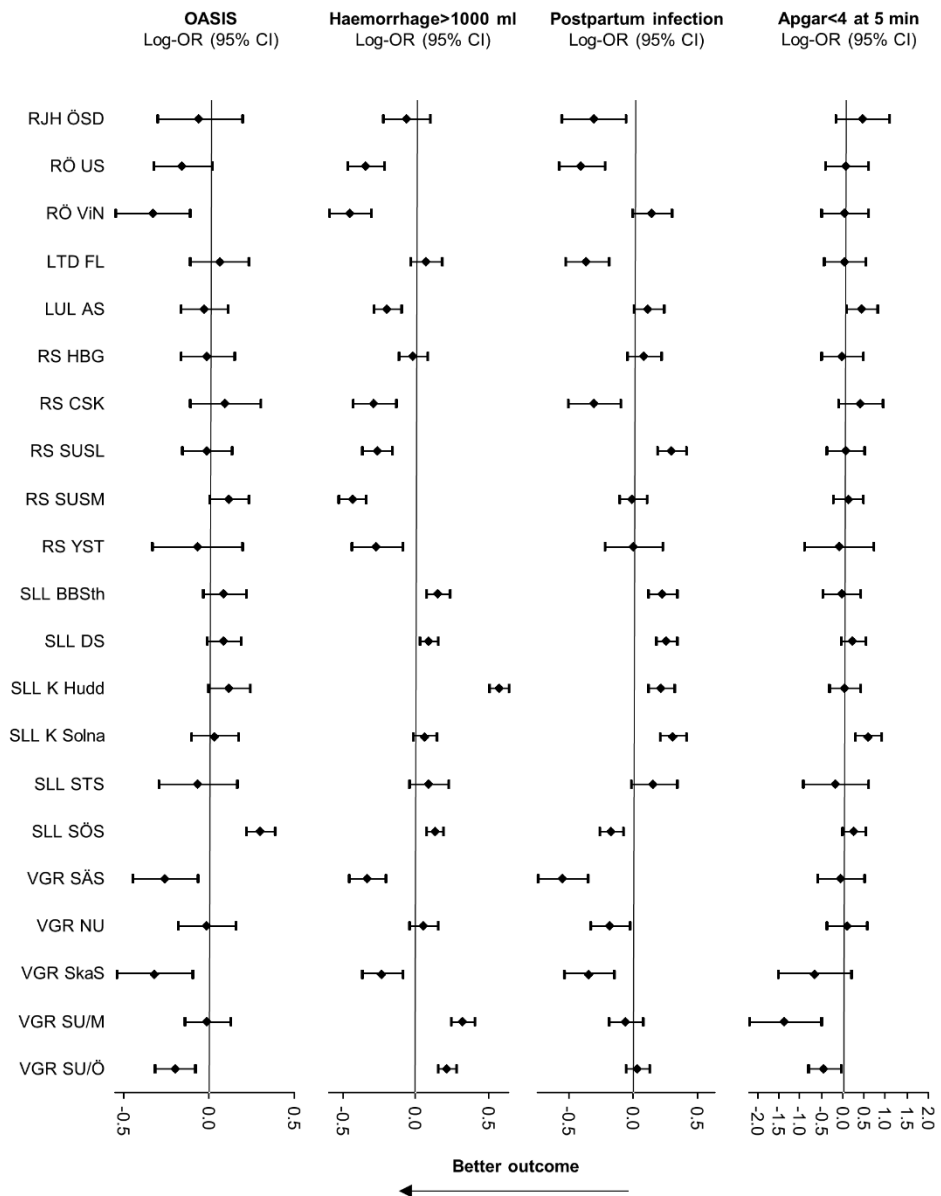


Figure 7 Case mix adjusted deviation from population mean for each health outcome indicator

## 4.5 HOW TECHNOLOGY CAN SUPPORT QUALITY IMPROVEMENT IN BIRTH CARE (STUDY IV)

In Study IV the NASSS framework was used to identify factors that can facilitate or hinder the adoption of technology-supported QI, based on interviews with managers and staff in an obstetric unit.

In terms of the domain **Condition**, the analytics platform was capable to handle the challenge of measuring quality in obstetric care at the department, which ranged from treating low-risk patients in normal labour to high-risk patients with comorbidities and complications.

As for the **Technology** domain, several factors facilitated the adoption of the platform. The interviewees described that data was provided in a timely manner and that case mix adjustment made data more relevant compared to unadjusted data, both of which contributed to its relevance for QI. The presentation of the data and statistics in the dashboards was deemed to be easy to understand and user-friendly. Moreover, interviewees reported that the relationship with the supplier was such that it was possible to customize the analytics platform to the local needs and conditions, particularly in terms of communicating with the hospital IT department and adapting to the local data. One barrier to the adoption was that staff groups described the need for guidance and support to identify, select, extract, and present relevant data. Prior understanding of statistics was also perceived to be helpful to fully understand the presented data.

Related to the **Value Proposition**, the platform was perceived as desirable which supported its adoption. Even before the launch of the technology-supported programme, data was perceived as an important tool for QI. However, historically the data entered into medical records had been difficult to access and use and was not sufficiently up-to-date to support QI at the unit.

As for the **Adopter system**, managers and staff involved in QI teams were the main users of the platform and its adoption did not imply substantial changes to individual roles and identities which supported the adoption. Managers and staff involved in QI teams used the data to identify performance deviations and needs for improvement grounded in valid information. In addition to using the platform, other data sources were used as a complement. For example, surveys, focus groups, and individual interviews with patients were used to gain the patient perspective on quality of care. Based on data from the platform and informed by available evidence and experience from other settings, changes were made to clinical protocols. The data available in the platform could be used to support quick feedback on improvement changes and observed results. A barrier to the adoption of the technology, was that interviewees identified some risks associated with the use of performance data for QI. For example, staff involved in QI teams described a concern that improvement efforts would solely focus on areas that could be measured, i.e. areas that could not be measured quantitatively were at risk of not being prioritized. Some challenges related to spread of the platform were also raised. The ambition was that all employees could have access and use the

analytics platform. In practice, however, staff not directly involved in QI teams stated that they had limited or no knowledge or experience of working with the platform.

In terms of **Organization**, several factors supported the adoption of the platform. The department reported that they had a history of working with QI, for instance with lean, and several QI initiatives were already ongoing before the platform was introduced. Moreover, the QI teams were multidisciplinary, including physicians, midwives, and assistant nurses. This created unity around both the conclusions and the proposed changes among multiple professional groups. The head of unit had an important role in facilitating the adoption of the platform, by engaging staff through showing the possibilities that the platform offered. This bottom-up strategy was believed to ensure sustainable adoption of the platform and decrease risk of abandonment. Managers used data to motivate staff to QI and the staff expressed that they experienced the visualization of completed and ongoing QI efforts as a motivating factor to continue improving. The increased focus on measurement and benchmarking led to an increased and shared understanding of the importance of data reliability. One hinder in the adoption process was that certain variables were found to be of insufficient quality to be useful for performance measurement. Thus, several efforts were made to improve routines regarding data recording.

The adoption was facilitated by the fact that the **Wider system** was characterized by a widespread societal and political debate on the need to improve birth care. Moreover, the hospital implemented a general initiative to adopt value-based health care (VBHC) as an improvement strategy. Managers reported that the unit's initiative linked to the hospital-wide initiative and this gave the unit some support in the improvement effort.

Because the data was collected around one year after implementation of the technology, the study focused mainly on early adoption of the technology and limited information was gained regarding **Embedding and adaptation over time**. During the early period, the results show that the platform supplier and its adopters worked to co-develop the platform in a way that was suitable for local data availability and needs. Moreover, the head of unit had a clear strategy for securing motivation of staff and mitigating risk of abandonment. However, the results do allow for assessing the ability of the department to embed and adapt the technology over longer time periods.

## 5 DISCUSSION

This thesis presents approaches for a thorough measurement of health care performance, using routinely collected register-based data. It demonstrates the importance of case mix adjustment and reveals large interhospital variations in performance after adjustment for case mix. Finally, it presents facilitators and barriers for adoption of technology-supported quality improvement in a hospital department.

### 5.1 INTERPRETATION OF FINDINGS

The importance of case mix adjustment is well established in the literature and this thesis confirms that adjustment for case mix is a prerequisite for valid assessment of performance. The analysis of CS rates showed that the ranking of hospitals differed significantly after case mix adjustment compared to before case mix adjustment, which highlights the importance of accounting for differences in patient characteristics. This thesis also shows that adjustment for case mix is motivational for health care professionals. Interviewees from the department working with data for QI stated that they had previously rationalized poor results by focusing on characteristic of their patient population rather than on the potential for clinical improvement.

In the current era of “big data”, there are unprecedented opportunities for development and continuous refinement of models for adequate case mix adjustment. However, the choice of factors to use for case mix adjustment is not straight-forward. As the name reveals, the objective is to adjust for available information about the treated “cases” (i.e. patients). In this thesis, 23 different baseline characteristics have been used, which together give a thorough understanding of risk factors associated with performance indicators such as CS, length of stay, and adverse obstetric outcomes. Nevertheless, it is likely possible to identify additional factors that have an impact on performance indicators. Case mix adjustment can never be said to be perfect or complete.

Given that the objective is to use the adjustment to reveal differences in performance between birth clinics, only factors outside of the clinics’ control should be adjusted for. The unit of comparison in this thesis is birth clinics and the patients’ clinical status at admission for delivery has been regarded as factors outside of the clinics’ control. However, it could be debated if all the factors used for case mix adjustment are outside of the clinics’ control. For example, both pre-term (delivery before week 36+6 days) and post-term birth (delivery after week 42+0 days) have been adjusted for in the analyses. However, one could argue that birth clinics’ management of labour and the choice of when to induce labour does impact if the woman ends up having these “risk factors” or not. Both pre-term and post-term delivery are very frequently used for case mix adjustment in birth care and have been included because they are largely driven by factors outside of the clinics’ control, but they serve as examples of the potentially difficult choice in the selection of what factors to adjust for. There are also factors that are very clearly within the control of the providers that have been used for case mix adjustment in previous literature. As an example, Leung et al. analysed case mix adjusted

variations in length of stay and included factors such as hospital size and mode of delivery in the model [92]. While it may be interesting to analyse whether hospital size and mode of delivery impact length of stay, they are factors that are driven by care organization more than actual patient characteristics and are consequently not an adequate reflection of case mix.

This thesis also underlines the importance of using a broad set of indicators for measurement of performance. As discussed, prior literature has focused almost exclusively on CS rate. While this is a highly relevant indicator it clearly does not adequately capture the performance of a birth clinic. A comprehensive understanding of performance requires information about which patients are treated, what care process they go through, which resources are consumed, and what outcomes are achieved. This type of information cannot be condensed into one or a couple of different performance indicators. Instead, to assess the ability of a hospital to deliver good health outcomes at a reasonable cost, a wide range of indicators is necessary. This thesis shows that different hospitals performed well on different indicators and all hospitals in this study had room for improvement on one or several indicators, which shows the importance of measuring an array of different indicators to understand where there is potential for improvement. As study II showed, there was no apparent link between CS rate and proportion of newborn with low Apgar score. When comparing findings from study II and study III, a tendency may be observed that hospitals with a higher rate of CS had a higher rate of post-partum infection and haemorrhage. This is logical given that surgery is associated with a risk of infection and bleeding. However, there was a large variation in both these and other outcomes that did not strictly follow the variation in CS rate, which shows that CS rate alone is not an indicator that adequately captures performance.

It is worth pointing out that there is no gold standard method for deriving the possible potential for improvement based on the performance distribution of the units investigated. Performing better than average on a specific indicator does not necessarily mean that the performance is good or that the result is satisfactory. However, to use the hospital with the best value for each specific indicator may also not represent a reasonable benchmark. Firstly, due to natural fluctuations and a certain degree of random variation over time, the hospital with the best performance on a specific indicator could be regarded as an outlier. Such random variations present a particularly important problem when the number of events of interest is low, due to short time interval for analysis, low number of observations per unit, or rarity of the event of interest. Secondly, given the well-known issue of differences in coding between providers there is also a possibility that deviations in coding practice at some clinics influence the results. To reduce the impact of these issues, the performance of the top 20 per cent was used as the benchmark in analyses of improvement potential. An alternative approach to deriving the benchmark from the distribution of current performance is to decide on a fixed target level that represents good performance, e.g. based on guidelines or expert consensus.

In addition to observed differences between hospitals within regions, certain significant differences can also be observed between regions. For example, the regional CS rate varied between 11.6% in Region Östergötland and 21.6% in Stockholm in 2014 [93]. This suggests that there are regional differences in ways to organize and provide care that impacts performance. These regional differences are most likely more accentuated when crossing country borders. Given differences in care practice and culture between countries, international benchmarking should therefore present additional opportunities to understand potential for improvement. There is an abundance of literature around global rates of CS which show that Sweden has a low CS rate in an international perspective [94]. There are scarce international data around other birth care performance indicators investigated in this thesis. However, as part of the Euro-Peristat project, a study investigated rates of severe perineal tears in 20 European countries. This study found Sweden to be amongst countries with highest rates of perineal tears but also concluded that there are likely to be significant variations between countries in the assessment and reporting of severe perineal tears [95].

It is worth emphasizing that not all variation between hospitals, regions and countries is bad. As discussed in the Introduction section, the term “unwarranted variation” is used to describe variations that are not explained by differences in patient characteristics or patient preferences [9]. The objective of the extensive case mix adjustment performed in this thesis is to remove the effect of differences in patient characteristics on observed variations. However, the impact of possible differences in patient preferences is more difficult to ascertain. This is not an issue for indicators of health outcomes and experience measures, where all patients can safely be assumed to prefer better outcomes and experiences of health care. Nevertheless, it is possible that women at different hospitals and regions have diverging preferences in terms of delivering through caesarean section as opposed to vaginal delivery. Also variations in process and resource use indicators such as labour induction and length of stay may be impacted by patient preferences. However, the preferences of patients in relation to treatment decisions are obviously impacted by the information received from the health care system so patient preferences should not be regarded as a completely external factor that is cannot be impacted by health care providers. Innovation, which is by definition deviation from the norm, is another potential source of variation which should not be regarded as negative [96]. However, this type of variation should be seen during relatively limited time period when early adopters introduce new technologies or routines and therefore differ from the norm. If the health care system manages to spread effective innovations and diffuse best practice, these differences should disappear over time. This underlines the need for systematically measuring performance continuously to enable quick feedback to providers regarding the impact of changes to clinical practise.

In order to use performance measurement for motivating health care professionals and enable learning and improvement, these professionals need to be involved in the selection of indicators and take ownership over them [18, 97]. One dimension of complexity in health care is the multitude of different professionals involved in care delivery, who in some cases have diverging views on what constitutes quality of care [98]. This thesis shows that access to

performance data with a variety of different indicators combined with multi-professional teams engaging in QI may bridge such differences and facilitate collaboration by creating a common view of current performance and a unity around the targets set. Moreover, while timely feedback of relevant data can support QI, it will obviously not by itself improve health care. Because “healthcare will not realise its full potential unless change making becomes an intrinsic part of everyone’s job, every day, in all parts of the system” [27], effective ways to engage and motivate those who should perform the improvement efforts continuously are required. This thesis finds that leadership to motivate staff and to create a culture of striving towards improving care processes is a fundamental component of continuous improvement.

The complexity of introducing technology in health care implies that organizations need to be resilient and determined when introducing new innovations and it is important not to expect that a technology-supported programme will be perfect immediately [39]. Instead, a collaborative approach between the supplier and the organization, and an openness for adaptations to the local context is required. An organization using technology for feedback of performance data needs to take measures to integrate this into the local context and be open for gradual improvements over time. This includes both actual improvements of the technology itself and the data used, but also improvements in how the technology is being adopted and used in the daily work.

## **5.2 STRENGTHS AND LIMITATIONS**

An important strength of the quantitative analyses in this thesis is that they are based on data covering around two thirds of all deliveries in Sweden during the time period investigated. Almost 140 000 deliveries were included in the analyses, which is a large sample of patients to robustly analyse the relationship between different case mix factors and performance indicators, and to estimate the amount of unwarranted interhospital variation. Moreover, Sweden’s long history of systematic data collection using personal identification numbers for record-linkage across different data sources provides excellent opportunities for register-based studies using routinely collected data. In this thesis data from Patient Administrative Systems, the Medical Birth Register, and sociodemographic data from Statistics Sweden allowed for a comprehensive analysis of patient characteristics, health outcomes, care process and resource use in Swedish birth care.

One obvious and fundamental challenge to measurement of performance is its dependency on valid data. The main limitation of the analyses of unwarranted variation in this thesis is the possibility of differences in coding between regions and hospitals. These differences may be due to differences in care process, screening practices, or clinical criteria for certain conditions, but can also simply be due to differences in the reporting of diagnosis and procedure codes into medical charts and into the administrative data. The research database included both administrative data and data from MBR, which is based on information from medical records. For some variables, information was available in both sources which enabled validation of the data. Performance indicators such as mode of delivery, labour induction and severe perineal tears were available in both sources of data. For these variables

the coding quality of diagnosis and procedure codes in administrative data was very high and in large agreement with the data in the MBR. One example of a problematic variable is gestational diabetes, where the large observed differences in the frequency of the diagnosis is likely to a large extent driven by differences in screening regimes [99]. The magnitude of differences in coding practices is difficult to ascertain. In study II, the provider with the highest and lowest observed value for each case mix factor was presented and the large differences between the two indicates variability beyond what can be explained by underlying patient population. Given the limited use and feedback of the data that is currently being reported by health care personnel, it is not surprising that the data quality is not perfect. Increasing use of reported data and transparency between hospitals and regions around the frequency of different conditions would likely be a strong driver for harmonization of the assessment and reporting of different conditions relevant for studying performance. One of the findings of study IV was that the unit working with technology-supported QI identified variables with insufficient quality and implemented new routines to standardize reporting of information in order to increase data quality.

One limitation is the absence of any PROMs in the assessment of performance. The collection of PROMs for measurement of outcomes of care has historically been limited but is becoming increasingly used. In Sweden, systematic collection of PROMs and PREMs has recently been initiated through the Pregnancy register which will give excellent future possibilities for studying variations in the experiences and patient-reported outcomes of women giving birth in Sweden. However, no such data was available during the time period investigated in this thesis. Consequently, the assessment of maternal health outcomes was restricted to those identified in administrative data. In terms of neonatal outcomes, the analysis was restricted to the recorded Apgar score for the newborn. In the research database there was information on all diagnoses before and after delivery for the women but there was no information available on diagnoses for their newborns in neonatal and paediatric care. This limits the possibilities of drawing conclusions about actual long-term outcomes for the newborn.

In addition to providing detailed analyses of performance and demonstrating unwarranted variations between birth clinics, a strength of this thesis is that it also includes the perspective of how performance measurement can be used in clinical practice to support improvement efforts. There is wide-spread belief that technology has the potential to support QI by providing timely and adequate feedback of performance data. While previous research has investigated methodological considerations related to performance measurement there is limited information in the literature about the hinders for engaging in data-driven QI supported by technology. In addition, there is limited knowledge about what is needed at the clinical level in terms of organization and adoption to really make use of performance measurement in the clinical improvement work. By conducting group interviews in a hospital department that has worked intensively with technology-supported QI, success factors and barriers have been better understood.



A strength of the qualitative study was the separate group interviews with different levels of staff. Both managers, staff involved in QI teams, and staff not involved in QI teams were included and were interviewed separately. The purpose with this approach was to reduce risk for power imbalances which could have hindered the interviewees to speak freely. It also allowed for a better understanding of adoption and spread throughout the organization and not only among managers and staff involved in QI. One methodological limitation is the fact that participants in the group interviews were recruited by the head of unit, who led the QI-programme. It is possible that this led to certain selection bias, for example if staff who were more positive towards QI or new ways of working were recruited to a larger extent. The reason that the head of unit was responsible for recruitment was that she had the best understanding of who belonged to which group and she also had an overview of the staff's schedule and was able to plan attendance in interviews. We deemed it to be difficult to recruit in another way but do acknowledge that this approach might have influenced selection of respondents.

The research team involved in the analysis of the material was interdisciplinary and consisted of five researchers with various background and with experience from different areas such as medicine, obstetrics, medical management, organizational change, statistics, and health economics. The presence of these different areas of expertise ensured that diverse perspectives on the data collected were presented and different views were evaluated to achieve a nuanced interpretation of the data.

### **5.3 SUGGESTIONS FOR FURTHER RESEARCH**

The analyses of variations in birth care presented here included information on all maternal diagnoses during and after delivery but did not include information on diagnoses for the newborn in neonatal and paediatric care, which constitutes an important part of the outcome in obstetric care. Consequently, neonatal morbidity could not be captured and used as an indicator of health outcomes. Transfer to neonatal care units may be used as an indicator of neonatal health, but certain local differences in routines for transferring newborns to neonatal care units limit the possibilities for using it as a direct indicator of health outcomes. Record-linkage of mothers and their newborns to capture actual neonatal morbidity would be an interesting topic for future studies of variation in childbirth care performance.

Another interesting topic for future research would be to investigate women's preferences regarding variations in rates for adverse maternal and neonatal outcomes across providers, to understand how differences in risk for these adverse outcomes would be weighed against each other by patients. One approach to conduct such studies is "stated preference" where preferences are elicited through hypothetical scenarios. This could be done by asking women which of the different hospitals (with varying risks of adverse events) that they would prefer. Another potential research approach is "revealed preference", where actual individual behaviour is used to elicit preferences. In a setting where pregnant women had access to publicly reported outcomes data and were able to freely choose their preferred hospital, the importance of these outcomes as perceived by the women could be investigated.

This thesis studies case mix adjusted variations in Sweden and the results show potential for improvement. However, performing similar analysis of risk adjusted variations including several countries could potentially reveal differences that are not seen within countries. Given that large cross-country differences in factors such as education, culture, and organization, variations would be expected to be larger than within-country variations. Moreover, for a large university hospital, benchmarking with university hospitals in other countries may provide more insights than benchmarking with a small hospital in the same region. The challenge would be to identify and define indicators that can be validly compared between countries, knowing that the countries can have different coding systems both for diagnosis codes and for medical procedures. As discussed in this thesis there are differences in coding practices and definitions of conditions between different hospitals and regions in Sweden, and such differences are likely more accentuated between countries. Nevertheless, research should be performed around the possibilities of international benchmarking and best practice sharing across countries.

One area of research that could be significantly developed is the incorporation of PROMs for measurement of performance in the area of birth care. There is evidence that health care professionals value PROMs in the clinical decision-making process. However, there are barriers to increased use of PROMs, such as a lack technology for capturing this information efficiently as well as challenges associated with the presentation and interpretation of the data [100]. Nevertheless, recent evidence that use of PROMs is associated with improved survival in cancer care further underlines the importance of systematically incorporating the patient perspective into measurement of performance [101].

Beyond capturing additional information from patients, another interesting area of future research would be how patients can be involved in performance measurement to a larger extent. This could give better information on what the most relevant aspects of measurement are from a patient-perspective. More research around how performance data can be reported back to patients and not only to health care professionals would also be of great value. Given that patients have more limited information about clinical decision making and less experience of interpretation of health care performance indicators compared to professionals, different ways of presenting data need to be explored and tested with patients.

## **6 CONCLUSION**

### **6.1 STUDY CONCLUSIONS**

This thesis shows that a broad range of baseline characteristics of women giving birth have a large impact on performance indicators in childbirth care. Consequently, sociodemographic and clinical maternal characteristics should be accounted for when conducting comparisons of performance between hospitals.

There are significant differences in CS rate between Swedish hospitals. Adjustment for case mix has an impact on the ranking of different hospitals but does not remove the large variations between hospitals. This indicates that there is a potential for fewer CS in Swedish childbirth care and consequently lower resource use. Health outcomes also vary between hospitals after adjustment for differences in case mix. The performance of hospitals was not the same for different outcome indicators and all hospitals have potential for improvement in certain aspects of labour management.

Another conclusion from this thesis is that technology has the potential to enable systematic QI through motivating professionals by providing timely and adequate feedback of performance. However, the adoption of such technology is complex and requires openness for gradual learning and improvement, and a leadership to engage professionals and other stakeholders in solving challenges that arise throughout the adoption process.

### **6.2 POLICY RECOMMENDATIONS FOR PERFORMANCE MEASUREMENT**

Given the extensive literature available and the findings of the studies in this thesis, there are several recommendations that should be considered in relation to performance measurement:

- Involve health care professionals and other stakeholders in the selection and definition of performance measures to ensure that they feel ownership over indicators.
- Enable health care professionals by giving them access to frequently updated and reliable information about their performance for use in continuous quality improvement.
- Ensure measurement of several types of indicators to ensure relevance for all stakeholders in the health care system: both patient relevant outcome measures and measures of the care process.
- Harmonize selection and measurement of indicators across hospitals, regions, and, if possible, nations to enable benchmarking and identification of best practice.
- When explicit financial incentives are introduced, make sure that they align with overall goals of the health care systems and that they do not conflict with the inner motivation of health care professionals. Also put processes in place to monitor development over time, to ensure that gaming does not occur and that unintended consequences do not appear.
- Put effort into statistical analysis to ensure reliability and validity of performance measures and employ case mix adjustment in interhospital comparisons. This is

important for acceptance among professionals and for understanding true improvement potential.

- Ensure accessible presentation of statistical data so that all stakeholders can draw appropriate conclusions.
- Support leaders in health care to create conditions for systematic quality improvement and to help them understand how they can motivate staff through use of data.
- Increase the understanding in health care organizations of the complexity associated with using technology to support quality improvement so that they have a strategy for dealing with challenges that arise throughout the adoption process.

There are also several additional recommendations specifically for birth care that are relevant to consider:

- Perform measurement of CS rate because it is linked to several indicators of health outcomes and resource use.
- Also ensure systematic measurement of other indicators than CS to get a good understanding of entire care process and the health outcomes.
- Collect information about PROMs and PREMs in relation to childbirth as these are fundamental factors to understand both short- and long-term outcomes.
- Use a broad spectrum of indicators of case mix (e.g. demographic information, reproductive history, other obstetric factors, and comorbidities) to adjust analyses of performance.

## 7 ACKNOWLEDGEMENTS

**Peter Lindgren**, my main supervisor, for great coaching and friendship over the years. Your trust and laissez-faire leadership together with the ability to really support when needed has been great for me and your fabulous sense of humour has made many of my days.

**Isis Amer-Wåhlin**, my co-supervisor, for strong support, hard work and an eagerness to always push forward. Your passion and drive have been inspirational throughout this journey.

**Mats Brommels**, for welcoming me to the Medical Management Centre, for supporting the research activities within the Sveus project and for stepping in and helping me when required.

**Eva Wallin**, for great mentorship, for giving me early insights into the art of scientific rigour at STAD, and most of all for very enjoyable after works and lunches together with **Johanna Gripenberg** throughout the last almost 20 years.

Thanks to **Linus Jönsson** for together with **Peter** taking me onboard the health economics journey in 2007 at European Health Economics and for mentoring me through my very first scientific articles.

**Anders Gustavsson, Anna Ringborg, Fredrik Borgström, Ingrid Lekander and Oskar Ström** for doing most of the health economics journey together with me and making the journey such an enormous pleasure. A special thanks to **Anders, Fredrik and Oskar** for taking a giant leap into the unknown together with me. I'm extremely proud of what we have accomplished together. Thanks also to **Åsa By** for great collaboration and to all colleagues at **Quantify Research** for lots of work and lots of fun.

Thanks also to **Jonas Wohlin** for helping me switch direction slightly and for a fantastic venture into performance measurement. All fantastic colleagues at **Ivbar**: it's a true pleasure to go to work every day and get the chance to collaborate with such a diverse and incredibly talented group of people. A particular thanks to **Niclas Graham**, who with great determination crunched most of the data that this thesis is based on.

Fellow researchers in the clinical management group, **Pamela Mazzocato and Carl Savage**, for sharing your knowledge about theories, frameworks and qualitative research. A particular thanks to **Pamela** for great collaboration on study IV and for sharing your expertise in quality improvement. Thanks also to **Sara Tolf** for tremendous work on study IV and for excellently guiding me through the do's and don'ts of qualitative research. Thanks also to **Daniel Söderberg** for hard work with data collection and analysis in study IV.

A special thanks to **Sara Korlén** for support and for happily ventilating the occasional burdens of being a doctoral student with me.

I wish to extend my gratitude to the entire **Sveus steering group** and to all participating regions for persevering with such a large and complex research project as Sveus, spanning over 5 years. A particular thanks to the **Sveus childbirth care expert group** who have put a

lot of work into discussions around indicators, definitions, case mix adjustment and interpretation of results. A special thanks to **Lars Ladfors** for truly supporting me all along.

I'm grateful for financial support for the study by the Forte research programme 'Value and Choice' (grant nr 2012/1688).

Thanks to all fantastic **friends**, you know who you are. Dinners, game nights, travels and laughs with all of you is what gives me energy! A special mention to the old **Farsta crew** for spending life's most defining years together with me, for all the fantastic adventures around the world and for still being around.

**Maria, Martin, Andreas** and **Katarina** for being such fabulous extended family members and for making vacations at the Riviera so enjoyable.

**Ingrid** and **Gunnar** for plenty of support over the years. Without your backing and my frequent escapes to the library in Kungshamn, this thesis would not be.

**Gabbe**, thanks for growing up together with me, putting up with me and for giving me different perspectives. Thanks also for bringing **Sara's** energy and smile into the family.

**Pappa**, thanks for inspiration, for great times together, and for together with **Karin** hosting yearly summer trips to France. Thanks also for providing great feedback on this work.

**Mamma**, thanks for your absolutely relentless support through thick and thin. It's a great comfort to know that I have you and **William** a short boat trip from home.

**Johanna**, my love, thank you for EVERYTHING. You are the light that always shines and that guides me through life.

## 8 REFERENCES

1. *The world health report 1999. Making a difference.* 1999, WHO.
2. Przywara, B., *Projecting future health care expenditure at European level: drivers, methodology and results*, in *Economic Papers 417, European Economy.* 2010, European Commission.
3. Peiro, S. and A. Maynard, *Variations in health care delivery within the European Union.* *Eur J Public Health*, 2015. **25 Suppl 1**: p. 1-2.
4. Brownlee, S., et al., *Evidence for overuse of medical services around the world.* *Lancet*, 2017. **390**(10090): p. 156-168.
5. Gibbons, L., et al., *The Global Numbers and Costs of Additionally Needed and Unnecessary Caesarean Sections Performed per Year: Overuse as a Barrier to Universal Coverage.* 2010, World Health Report, Background paper 30, WHO.
6. Glasziou, P., et al., *Evidence for underuse of effective medical services around the world.* *Lancet*, 2017. **390**(10090): p. 169-177.
7. Balas, E.A. and S.A. Boren, *Managing Clinical Knowledge for Health Care Improvement.* *Yearb Med Inform*, 2000(1): p. 65-70.
8. Eisenberg, M.J., et al., *Use of evidence-based medical therapy in patients undergoing percutaneous coronary revascularization in the United States, Europe, and Canada. Coronary Angioplasty Versus Excisional Atherectomy Trial (CAVEAT-I) and Canadian Coronary Atherectomy Trial (CCAT) investigators.* *Am J Cardiol*, 1997. **79**(7): p. 867-72.
9. Wennberg, J.E., *Tracking medicine: a researcher's quest to understand health care.* 2010, New York: Oxford University Press.
10. *The NHS Atlas of Variation in Healthcare. Reducing unwarranted variation to increase value and improve quality.* 2016, RightCare, NHS.
11. Smith, P., et al., eds. *Performance Measurement for Health System Improvement: Experiences, Challenges and Prospects.* European Observatory on Health Systems and Policies. 2009, Cambridge University Press.
12. Walshe, K., *Managing Performance*, in *Healthcare Management*, K. Walshe and J. Smith, Editors. 2011, McGraw-Hill Education.
13. Berwick, D.M., *A primer on leading the improvement of systems.* *BMJ*, 1996. **312**(7031): p. 619-22.
14. Donabedian, A., *Evaluating the quality of medical care.* 1966. *Milbank Q*, 2005. **83**(4): p. 691-729.
15. Loeb, J.M., *The current state of performance measurement in health care.* *Int J Qual Health Care*, 2004. **16 Suppl 1**: p. i5-9.
16. <http://www.dartmouthatlas.org/>.
17. <https://skl.se/tjanster/merfranski/oppnajakforelser/halsoochsjukvard.1563.html>.
18. Epstein, A., *Performance measurement and professional improvement*, in *Performance Measurement for Health System Improvement: Experiences, Challenges and Prospects*, P. Smith, et al., Editors. 2009, Cambridge University Press.

19. International Society for Quality in Health Care. Available from: <https://isqua.org/accreditation-iap/what-is-the-iap>.
20. Shaw, C.D., et al., *The effect of certification and accreditation on quality management in 4 clinical services in 73 European hospitals*. Int J Qual Health Care, 2014. **26 Suppl 1**: p. 100-7.
21. Shekelle, P., et al., *Does public release of performance results improve quality of care? A systematic review*, in *Quest for Quality and Improved Performance (QQUIP)*, RAND Corporation, Editor. 2008.
22. Berwick, D.M., B. James, and M.J. Coye, *Connections between quality measurement and improvement*. Med Care, 2003. **41**(1 Suppl): p. I30-8.
23. Shekelle, P., *Public performance reporting on quality information*, in *Performance Measurement for Health System Improvement: Experiences, Challenges and Prospects*, P. Smith, et al., Editors. 2009, Cambridge University Press.
24. Conrad, D., *Incentives for health-care performance improvement*, in *Performance Measurement for Health System Improvement: Experiences, Challenges and Prospects*, P. Smith, et al., Editors. 2009, Cambridge University Press.
25. Eijkenaar, F., et al., *Effects of pay for performance in health care: a systematic review of systematic reviews*. Health Policy, 2013. **110**(2-3): p. 115-30.
26. Ivers, N., et al., *Audit and feedback: effects on professional practice and healthcare outcomes*. Cochrane Database Syst Rev, 2012(6): p. CD000259.
27. Batalden, P.B. and F. Davidoff, *What is "quality improvement" and how can it transform healthcare?* Qual Saf Health Care, 2007. **16**(1): p. 2-3.
28. Strome, T.L., *Healthcare analytics for quality and performance improvement*. 2013, Hoboken, New Jersey: John Wiley and Sons, Inc. xiii, 265 pages.
29. van der Veer, S.N., et al., *Improving quality of care. A systematic review on how medical registries provide information feedback to health care providers*. Int J Med Inform, 2010. **79**(5): p. 305-23.
30. Taylor, M.J., et al., *Systematic review of the application of the plan-do-study-act method to improve quality in healthcare*. BMJ Qual Saf, 2014. **23**(4): p. 290-8.
31. Simms, R.A., et al., *Using data and quality monitoring to enhance maternity outcomes: a qualitative study of risk managers' perspectives*. BMJ Qual Saf, 2014. **23**(6): p. 457-64.
32. Pencheon, D., *The good indicators guide: understanding how to use and choose indicators*. 2008, NHS Institute for Innovation and Improvement.
33. Institute of Medicine, *Performance Measurement: Accelerating Improvement*. 2006: The National Academies Press.
34. Porter, M.E., S. Larsson, and T.H. Lee, *Standardizing Patient Outcomes Measurement*. N Engl J Med, 2016. **374**(6): p. 504-6.
35. Mant, J., *Process versus outcome indicators in the assessment of quality of health care*. Int J Qual Health Care, 2001. **13**(6): p. 475-80.
36. Nelson, E.C., et al., *Patient reported outcome measures in practice*. BMJ, 2015. **350**: p. g7818.



37. Iezzoni, L., *Risk adjustment for performance measurement*, in *Performance Measurement for Health System Improvement: Experiences, Challenges and Prospects*, P. Smith, et al., Editors. 2009, Cambridge University Press.
38. EMC Digital Universe. *Vertical Industry Brief: The Digital Universe. Driving Data Growth in Healthcare*. 2014 05 Nov, 2018]; Available from: <https://www.emc.com/analyst-report/digital-universe-healthcare-vertical-report-ar.pdf>.
39. Greenhalgh, T., et al., *Analysing the role of complexity in explaining the fortunes of technology programmes: empirical application of the NASSS framework*. BMC Med, 2018. **16**(1): p. 66.
40. Plsek, P.E. and T. Greenhalgh, *Complexity science: The challenge of complexity in health care*. BMJ, 2001. **323**(7313): p. 625-8.
41. Greenhalgh, T., et al., *Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies*. J Med Internet Res, 2017. **19**(11): p. e367.
42. Eurostat. *Hospital discharges and length of stay statistics*. 2016 13 Jun, 2017]; Available from: [http://ec.europa.eu/eurostat/statistics-explained/index.php/Hospital\\_discharges\\_and\\_length\\_of\\_stay\\_statistics](http://ec.europa.eu/eurostat/statistics-explained/index.php/Hospital_discharges_and_length_of_stay_statistics).
43. *Global Health Observatory country views*. 2017, WHO.
44. World Health Organization. *A "good birth" goes beyond having a healthy baby*. 2018; Available from: <http://www.who.int/mediacentre/commentaries/2018/having-a-healthy-baby/en/>.
45. Escuriet, R., et al., *Assessing the performance of maternity care in Europe: a critical exploration of tools and indicators*. BMC Health Serv Res, 2015. **15**: p. 491.
46. *Standards of Care for Women's Health in Europe: Obstetric and Neonatal Services*. 2014, European Board and College of Obstetrics and Gynaecology.
47. Mahmud, A., et al., *Developing core patient-reported outcomes in maternity: PRO-Maternity*. BJOG, 2014. **121 Suppl 4**: p. 15-9.
48. Nijagal, M.A., et al., *Standardized outcome measures for pregnancy and childbirth, an ICHOM proposal*. BMC Health Serv Res, 2018. **18**(1): p. 953.
49. Gottvall, K. and U. Waldenstrom, *Does a traumatic birth experience have an impact on future reproduction?* BJOG, 2002. **109**(3): p. 254-60.
50. Nilver, H., C. Begley, and M. Berg, *Measuring women's childbirth experiences: a systematic review for identification and analysis of validated instruments*. BMC Pregnancy Childbirth, 2017. **17**(1): p. 203.
51. Betran, A.P., et al., *Rates of caesarean section: analysis of global, regional and national estimates*. Paediatr Perinat Epidemiol, 2007. **21**(2): p. 98-113.
52. Sibanda, T., et al., *Intrapartum care quality indicators: a systematic approach for achieving consensus*. Eur J Obstet Gynecol Reprod Biol, 2013. **166**(1): p. 23-9.
53. Bailit, J.L., *Measuring the quality of inpatient obstetrical care*. Obstet Gynecol Surv, 2007. **62**(3): p. 207-13.

54. Shorten, A., *Maternal and neonatal effects of caesarean section*. BMJ, 2007. **335**(7628): p. 1003-4.
55. Ecker, J.L. and F.D. Frigoletto, Jr., *Cesarean delivery and the risk-benefit calculus*. N Engl J Med, 2007. **356**(9): p. 885-8.
56. Lydon-Rochelle, M., et al., *Association between method of delivery and maternal rehospitalization*. JAMA, 2000. **283**(18): p. 2411-6.
57. Allen, V.M., C.M. O'Connell, and T.F. Baskett, *Cumulative economic implications of initial method of delivery*. Obstet Gynecol, 2006. **108**(3 Pt 1): p. 549-55.
58. Robson, M.S., *Can we reduce the caesarean section rate?* Best Pract Res Clin Obstet Gynaecol, 2001. **15**(1): p. 179-94.
59. Betran, A.P., et al., *A systematic review of the Robson classification for caesarean section: what works, doesn't work and how to improve it*. PLoS One, 2014. **9**(6): p. e97769.
60. Colais, P., et al., *Risk adjustment models for interhospital comparison of CS rates using Robson's ten group classification system and other socio-demographic and clinical variables*. BMC Pregnancy Childbirth, 2012. **12**: p. 54.
61. Aron, D.C., et al., *Impact of risk-adjusting cesarean delivery rates when reporting hospital performance*. JAMA, 1998. **279**(24): p. 1968-72.
62. Glantz, J.C., *Cesarean delivery risk adjustment for regional interhospital comparisons*. Am J Obstet Gynecol, 1999. **181**(6): p. 1425-31.
63. Kozhimannil, K.B., M.C. Arcaya, and S.V. Subramanian, *Maternal clinical diagnoses and hospital variation in the risk of cesarean delivery: analyses of a National US Hospital Discharge Database*. PLoS Med, 2014. **11**(10): p. e1001745.
64. Lieberman, E., et al., *Assessing the role of case mix in cesarean delivery rates*. Obstet Gynecol, 1998. **92**(1): p. 1-7.
65. Linton, A., M.R. Peterson, and T.V. Williams, *Clinical case mix adjustment of cesarean delivery rates in U.S. military hospitals, 2002*. Obstet Gynecol, 2005. **105**(3): p. 598-606.
66. Paranjothy, S., C. Frost, and J. Thomas, *How much variation in CS rates can be explained by case mix differences?* BJOG, 2005. **112**(5): p. 658-66.
67. Bragg, F., et al., *Variation in rates of caesarean section among English NHS trusts after accounting for maternal and clinical risk: cross sectional study*. BMJ, 2010. **341**: p. c5065.
68. Maso, G., et al., *The application of the Ten Group classification system (TGCS) in caesarean delivery case mix adjustment. A multicenter prospective study*. PLoS One, 2013. **8**(6): p. e62364.
69. Stivanello, E., et al., *Risk adjustment for cesarean delivery rates: how many variables do we need? An observational study using administrative databases*. BMC Health Serv Res, 2013. **13**: p. 13.
70. Fantini, M.P., et al., *Risk adjustment for inter-hospital comparison of primary cesarean section rates: need, validity and parsimony*. BMC Health Serv Res, 2006. **6**: p. 100.

71. Wehberg, S., et al., *Risk factors and between-hospital variation of caesarean section in Denmark: a cohort study*. *BMJ Open*, 2018. **8**(2): p. e019120.
72. Pallasmaa, N., et al., *Variation in cesarean section rates is not related to maternal and neonatal outcomes*. *Acta Obstet Gynecol Scand*, 2013. **92**(10): p. 1168-74.
73. Lutomski, J.E., J.J. Morrison, and M.T. Lydon-Rochelle, *Regional variation in obstetrical intervention for hospital birth in the Republic of Ireland, 2005-2009*. *BMC Pregnancy Childbirth*, 2012. **12**: p. 123.
74. Nippita, T.A., et al., *Variation in hospital caesarean section rates and obstetric outcomes among nulliparae at term: a population-based cohort study*. *BJOG*, 2015. **122**(5): p. 702-11.
75. Bailit, J.L., et al., *Risk-adjusted models for adverse obstetric outcomes and variation in risk-adjusted outcomes across hospitals*. *Am J Obstet Gynecol*, 2013. **209**(5): p. 446 e1-446 e30.
76. *Patterns of Maternity Care in English NHS Hospitals 2011/2012*. 2013, The Royal College of Obstetricians and Gynaecologists.
77. Baghurst, P.A. and G. Antoniou, *Risk models for benchmarking severe perineal tears during vaginal childbirth: a cross-sectional study of public hospitals in South Australia, 2002-08*. *Paediatr Perinat Epidemiol*, 2012. **26**(5): p. 430-7.
78. Glantz, J.C. and D.S. Guzick, *Can differences in labor induction rates be explained by case mix?* *J Reprod Med*, 2004. **49**(3): p. 175-81.
79. Wen, S.W., et al., *Trends and variations in length of hospital stay for childbirth in Canada*. *CMAJ*, 1998. **158**(7): p. 875-80.
80. Berry, S.A., et al., *ProvenCare perinatal: a model for delivering evidence/ guideline-based care for perinatal populations*. *Jt Comm J Qual Patient Saf*, 2011. **37**(5): p. 229-39.
81. Main, E.K., *Reducing cesarean birth rates with data-driven quality improvement activities*. *Pediatrics*, 1999. **103**(1 Suppl E): p. 374-83.
82. *Förlossningsrapport 2013 [Childbirth report] (in Swedish.)* 2014, Sydöstra sjukvårdsregionen  
([http://plus.rjl.se/info\\_files/infosida41185/kval\\_grupp\\_obstetrik\\_arsrapport\\_2013.pdf](http://plus.rjl.se/info_files/infosida41185/kval_grupp_obstetrik_arsrapport_2013.pdf))
83. Blomberg, M., *Avoiding the first cesarean section--results of structured organizational and cultural changes*. *Acta Obstet Gynecol Scand*, 2016.
84. Porter, M.E., *What is value in health care?* *N Engl J Med*, 2010. **363**(26): p. 2477-81.
85. Porter, M.E., *A strategy for health care reform--toward a value-based system*. *N Engl J Med*, 2009. **361**(2): p. 109-12.
86. *The Swedish Medical Birth Register - A summary of content and quality*. 2003, Centre for epidemiology. The national board of health and welfare.
87. *Historic population register*, in *Population and Welfare Statistics* 2006, Statistics Sweden Örebro.
88. Kitzinger, J., *Qualitative research. Introducing focus groups*. *BMJ*, 1995. **311**(7000): p. 299-302.

89. Hsieh, H.F. and S.E. Shannon, *Three approaches to qualitative content analysis*. Qual Health Res, 2005. **15**(9): p. 1277-88.
90. Tong, A., P. Sainsbury, and J. Craig, *Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups*. Int J Qual Health Care, 2007. **19**(6): p. 349-57.
91. Statistics Sweden. *Population 1 November by region, age and sex. Year 2002 - 2016*. 2016 1 Aug, 2017]; Available from: [http://www.statistikdatabasen.scb.se/pxweb/en/ssd/START\\_BE\\_BE0101\\_BE0101A/FolkmandNov/?rxid=e67d909b-1ada-4703-8d17-6d613980114b](http://www.statistikdatabasen.scb.se/pxweb/en/ssd/START_BE_BE0101_BE0101A/FolkmandNov/?rxid=e67d909b-1ada-4703-8d17-6d613980114b).
92. Leung, K.M., et al., *Hospital- and patient-related characteristics determining maternity length of stay: a hierarchical linear model approach*. Am J Public Health, 1998. **88**(3): p. 377-81.
93. *Graviditeter, förlossningar och nyfödda barn. Medicinska födelseregistret 1973-2014. [Pregnancies, deliveries and newborn children. The Medical Birth Register 1973-2014] (in Swedish.)*, in *Sveriges Officiella Statistik*. 2015, The National Board of Health and Welfare: Stockholm.
94. OECD, *Health at a Glance 2017: OECD Indicators*, OECD Publishing, Editor. 2017: Paris.
95. Blondel, B., et al., *Variations in rates of severe perineal tears and episiotomies in 20 European countries: a study based on routine national data in Euro-Peristat Project*. Acta Obstet Gynecol Scand, 2016. **95**(7): p. 746-54.
96. Plsek, P.E. and T. Wilson, *Complexity, leadership, and management in healthcare organisations*. BMJ, 2001. **323**(7315): p. 746-9.
97. Damschroder, L.J., et al., *Effects of performance measure implementation on clinical manager and provider motivation*. J Gen Intern Med, 2014. **29 Suppl 4**: p. 877-84.
98. Storkholm, M.H., et al., *Money's (not) on my mind: a qualitative study of how staff and managers understand health care's triple Aim*. BMC Health Serv Res, 2017. **17**(1): p. 98.
99. Lindqvist, M., et al., *No consensus on gestational diabetes mellitus screening regimes in Sweden: pregnancy outcomes in relation to different screening regimes 2011 to 2012, a cross-sectional study*. BMC Pregnancy Childbirth, 2014. **14**: p. 185.
100. Boyce, M., J. Browne, and J. Greenhalgh, *The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research*. BMJ Quality & Safety, 2014. **23**: p. 508-518.
101. Basch, E., et al., *Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment*. JAMA, 2017. **318**(2): p. 197-198.

## 9 APPENDIX

Intervjuguide

# Fokusgruppintervjuer med deltagare i datadrivet förbättringsarbete inom förlossningsvården

## Deltagare

Personal på kvinnokliniken på Akademiska sjukhuset i Uppsala.

## Inledningstext

Syftet med detta projekt är att öka förståelsen om hur hälso- och sjukvårdsorganisationer kan utveckla sin förmåga att öka kvaliteten med hjälp av datadrivet förbättringsarbete.

Forskningsprojektet ämnar studera pågående datadrivna förbättringsarbeten i Uppsala och Stockholm. Vi avser att kunna få en djupare förståelse för hur dessa förbättringsansatser genomförs i praktiken samt att identifiera hindrande och underlättande faktorer i förbättringsansatserna utifrån ert perspektiv.

Avsikten med projektet är att utveckla konkreta förslag till utveckling av effektiva sätt att förbättra hälso- och sjukvården.

Intervjun kommer att ta max 90 minuter. Intervjufrågorna handlar om vad datadrivet förbättringsarbete innebär, hur ni arbetar med datadriven förbättring, hur ni arbetade tidigare med förbättringsarbete, vilken typ av data ni använder, vilka utmaningar ni ser med att arbeta på det här sättet och vilka fördelar ni ser.

Vi har valt en gruppintervju av två anledningar: 1) för att ta mindre tid från ert patientarbete och 2) vi är intresserad av hur ni arbetar på klinisk nivå med förbättringsarbete och inte på individnivå. Därför kommer vi att analysera på grupp nivå, personal, chefer.

Frågorna är öppna för att ni ska kunna beskriva era tankar fritt. Deltagandet är frivilligt. Ni kan när som helst välja att avbryta intervjun. All data kommer att behandlas konfidentiellt och inspelningarna kommer att raderas efter transkribering. Deltagandet i intervjun är anonymt på det sättet att ni inte kommer nämnas vid namn i något producerat dokument. Endast vår forskargrupp på MMC kommer att ha tillgång till intervjumaterialet. Vi kommer att sammanfatta resultaten i en vetenskaplig artikel.

För att underlätta analysarbetet kommer intervjun att spelas in. Är det ok för er?

Då sätter jag igång inspelningen. Och då upprepar jag frågan, är det ok att vi spelar in?

## Frågor

### Inledningsfråga

- 1. Skulle ni kunna kort berätta om vilka ni är (profession, roll i organisationen, hur länge du har jobbat här)?**

### Kvalitetsarbete generellt

- 2. Skulle ni kunna berätta hur ni jobbar för att förbättra vården för era patienter?**
  - a. Beskriv konkret? Vad innebär det?
  - b. Finns det någon särskild metod ni använder er av?
  - c. Har ert sätt att arbeta med förbättring förändrats under de senare åren?
  - d. Finns det någon särskild patientgrupp som ni fokuserar på (**condition**)?
- 3. Hur mäter ni och följer upp den vård som ni bedriver?**
  - a. Vilken data?

### Technology (data & plattformen)

Som vi förstår så har det pågått ett arbete för att följa upp kvalitetsarbete på ett systematiskt och digitaliserat sätt på er klinik, med hjälp av TV-vyn?

- 4. Hur använder ni plattformen?**
  - a. Vilken data använder ni?
  - b. Hur använder ni data?
- 5. Vilken kunskap/information får ni via data/plattformen? (OBS! kanske överflödigt)**
- 6. Vilken kunskap behövs det för att använda data/plattformen?**
- 7. Vilket stöd behövs det för att använda data/plattformen? (det som finns och saknas)**

### Value proposition and value chain (programteori)

Om vi blickar tillbaka då er klinik började arbeta med plattformen.

- 8. Varför började kliniken använda data och plattformen?**
  - a. Men dina egna ord, vilket problem försökte ni att lösa med dessa?
  - b. Vilket behov fanns det? (commitment)
  - c. Vilka resultat förväntade ni er?
- 9. Värdebaserad vård syftar att utveckla bättre metoder för att mäta och följa upp värde i vården med ambitionen att åstadkomma bästa tänkbara hälsoutfall för patienter. Hur anser ni att denna ambition har förverkligats i praktiken?**

### Adopter system

- 10. Har data och plattformen förändrat ert sätt att med jobba med förbättringsarbete?**
  - a. Har ni använt någon vedertagen modell för projektet?
  - b. Hur har chefernas arbete förändrats?
  - c. Hur har medarbetarnas arbete förändrats?
  - d. Vilka resultat har ni observerat för patienter?
    - i. Är det någon särskild patientgrupp/område där ni har fått mest nytta av data/plattformen?
  - e. Hur har samarbete med andra aktörer förändrats? T.ex. andra kliniker på sjukhuset, andra vårdgivare, landsting?
- 11. Vilka fördelar har ni upplevt mot tidigare sätt att arbeta med förbättring?**
- 12. Vilka utmaningar har ni upplevt mot tidigare sätt att arbeta med förbättring?**

## Organization

**13. Hur gick införandet av plattformen till?**

- a. Vilka var nyckelpersoner?
- b. Hur upplevde ni införandet av plattformen? (commitment)

**14. Hur såg förutsättningarna ut för att ni skulle kunna arbeta med den här plattformen (t.ex. resurser, stöd, tid, kompetens)? (efficacy)**

## Wider system

**15. Hur påverkar yttre faktorer, till ex. policy, krav från beställaren, hur ni arbetar med data/plattformen?**

## Embedding and adoption over time

**16. Vad har ni lärt er av arbetet och användning av data/plattformen?**

**17. Hur har arbetet och användningen av plattformen utvecklas över tid? Har ni behövt göra några förändringar?**