External inspections of healthcare organizations

A study of organizational and clinical effects of inspections on the management of sepsis in Norwegian hospitals

Gunnar Husabø

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2021

UNIVERSITY OF BERGEN

External inspections of healthcare organizations

A study of organizational and clinical effects of inspections on the management of sepsis in Norwegian hospitals

Gunnar Husabø

Thesis for the degree of Philosophiae Doctor (PhD)

at the University of Bergen

Date of defense: 11.05.2021

Year: 2021

Title: External inspections of healthcare organizations

Name: Gunnar Husabø

Print: Skipnes Kommunikasjon / University of Bergen

Scientific environment

My research has been funded by the Western Norway University of Applied Sciences (HVL), where I have worked at the Department of Social Sciences. From January 2017, I have been a PhD student at the Department of Global Public Health and Primary Care at the University of Bergen (UiB). I was enrolled in the Norwegian Research School in General Practice (NAFALM) from 2017 to 2019.

My main supervisor has been Einar Hovlid, Associate Professor at HVL and UiB and researcher at the Norwegian Board of Health Supervision. My co-supervisors have been Professor Gunnar Tschudi Bondevik (UiB) and Professor Jan Frich (University of Oslo).

Acknowledgements

In *The Face of Battle*, a book about military history and famous battles, John Keegan starts off with a caveat of sorts: "I have not been in a battle; not near one, nor heard one from afar, nor seen the aftermath" (Keegan, 1976, p. 15). Similarly, I have to confess: I have never provided care to someone with sepsis; nor have I observed sepsis treatment; I have not even – to the best of my knowledge – seen a case of sepsis.

Having admitted ignorance, I must thank those who made it possible for me to write a thesis about inspections of sepsis care. First, I want to thank the Norwegian Board of Health Supervision for launching the research project that this thesis is a part of, and for their continuing support of the project. I also want to thank the staff at the Norwegian Board of Health Supervision who contributed to planning and carrying out the project.

I am grateful to the research communities I have been included in as a PhD student at the University of Bergen. I have benefitted enormously from the way the Section for General Practice, the Group for health service research, and the National research school in general practice generously have received a wayward non-healthcare worker in their midst. I also thank my former colleagues in Deloitte Risk Advisory, who may not have the time or inclination to read this thesis but who nonetheless gave me innumerous practical lessons in doing external assessments throughout the ten years we worked together.

I would be remiss if I did not thank the Department of Social Science and Western Norway University of Applied Sciences for funding my work. I also want to thank my colleagues at the department for their words of encouragement during my time working on this research project. Especially, I want to acknowledge the crucial contribution made by Geir Sverre Braut. In his capacity working for the Norwegian Board of Health Supervision, he played an essential role in the data collection. He was also an important co-author of two of the articles included in the thesis, and he graciously took the time to read and provide suggestions on the penultimate version of the synopsis.

I was also greatly helped by my other co-authors. I am grateful to Erik Solligård, Stig Harthug, and Hans Kristian Flaatten for helping me understand the care processes involved in the diagnosis and treatment of sepsis; Kieran Walshe for his thoughtful comments and his unparalleled insight into regulation theory; Jon Helgeland for lending a fresh pair of eyes to the quantitative analyses; Inger Lise Teig for experienced help with both doing interviews and developing the qualitative analyses; and Roy Miodini Nilsen for his invaluable help in designing the studies and reviewing the analyses, and for always taking the time to respond to my stupid statistics questions.

Equally I am grateful to my supervisors. Over the last four years co-supervisors Jan Frich and Gunnar Bondevik have expertly provided instructions on the ins and outs of life as a PhD student and fledgling researcher. And I could not have asked for a better main supervisor than Einar Hovlid. He supervised far beyond the call of duty, and I owe a bottomless debt of gratitude for the time he has devoted to providing support, advise, and encouragement throughout this project.

I am grateful to the healthcare professionals, managers, and inspectors who gave us their time in the interviews. I also want to thank the unsung heroes: those who work tirelessly in the emergency departments in order to save peoples' lives, and those from the County Governors' who do inspections that contribute to making the services better and safer.

Amalie, Selma, and Hedvig have my thanks for being such charming, rambunctious, and warmhearted daughters. Lastly, I thank Trude for proofreading the synopsis, for which I am grateful, and for her patience and kindness, of which I am in awe.

Abstract

External inspections of healthcare organizations serve different purposes, one of which is to bring about quality improvements in the care offered to patients. Research has offered many different perspectives on how well inspections and other external assessment approaches fare in this endeavor. There is, however, no general agreement about the effects of inspections or the mechanisms by which inspections influence quality improvement work.

This thesis is part of a research project, headed by the Norwegian Board of Health Supervision, into the effects of inspections. The thesis examines the effects of inspections of care for patients with sepsis in the emergency departments of 24 hospitals in Norway during the period 2016 to 2018. The research aim has been to find out if inspections of healthcare organizations can lead to improved work processes and health outcomes for patients, and to explore the mechanisms that link external inspections and internal improvement efforts. The results have been reported in three published studies.

The two first studies relied on data from electronic health records and the Norwegian Patient Registry. **Study 1** is an observational study of the care processes and patient outcomes for patients with sepsis presenting to emergency departments. We included data from 1559 patients presenting to hospital emergency departments with sepsis. Assessing the timeliness of diagnostic procedures for recognizing sepsis, we found that 72.9% (95% confidence interval [CI]: 70.7-75.1) had documented triage within 15 minutes of presentation to the emergency departments, 44.9% (95% CI: 42.4-47.4) were examined by a physician in accordance with the triage priority, 44.4% (95% CI: 41.4-46.9) were adequately observed through continual monitoring of signs while in the emergency department, and 25.4% (95% CI: 23.2-27.7) received antibiotics within 1 hour. Next, we estimated associations between diagnostic procedures and time to antibiotic treatment, and between time to antibiotic treatment and mortality. We found that delay or non-completion of key diagnostic procedures predicted a delay of more than 2.5 hours to antibiotic treatment. Patients who received antibiotics within 1

hour had an observed 30-day all-cause mortality of 13.6% (95% CI: 10.1-17.1), in the timespan 2 to 3 hours after admission 5.9% (95% CI: 2.8-9.1), and 4 hours or later after admission 10.5% (5.7-15.3).

Study 2 is a stepped wedge study of the effects of the inspection on care processes and patient outcomes. The study included 7407 patients presenting to hospital emergency departments with sepsis. We first studied the effects of the inspection on process measures for sepsis diagnostics and treatment. We found significant improvements in the proportions of patients examined by a physician within the time frame set in triage (odds ratio [OR] = 1.28, 95% CI: 1.07-1.53), undergoing a complete set of vital measurements within 1 hour (OR = 1.78, 95% CI: 1.10-2.87), having lactate measured within 1 hour (OR = 2.75, 95% CI: 1.83-4.15), having an adequate observation regimen (OR = 2.20, 95% CI: 1.51-3.20), and receiving antibiotics within 1 hour (OR = 2.16, 95% CI: 1.83-2.55). We then studied the effects of the inspection on length of hospital stay and 30-day all-cause mortality. We found a significant reduction in mortality and length of stay, but these findings were no longer significant when adjusting the analyses for year of admission.

Study 3 used focus group interviews with inspection teams, hospital management, and hospital staff to explore how the inspections affected quality improvement work in inspected hospitals. The data of this study were twelve focus groups interviews, with a total of 47 participants. We identified three themes that were central for understanding how the inspection could contribute to clinical improvement in the emergency departments: 1) increasing awareness about the need to improve the quality of care by providing data on clinical performance, 2) building acceptance for improvement through professional credibility and focus on clinical practice, and 3) fostering leadership commitment.

In conclusion, the three studies suggest that the inspections brought to light deficiencies in the emergency departments' work with recognizing sepsis and treating patients presenting with sepsis. After the inspections, the process measures of diagnosis and treatment had improved, and our analyses showed that these changes were associated

with the inspections. This indicates that the inspection had a positive effect on the emergency departments' efforts to improve the management of patients with sepsis. The focus group interviews offer some insight into how the inspection may have contributed to improving care: The inspection could help hospitals to identify and understand weaknesses in the clinical care processes and bolster the organizational commitment to systemic quality improvement. This shows the potential for the inspection to become an instrument of improvement. It is, however, also important to recognize that such change processes are context-dependent, and that regulatory agencies should reflect on how to design inspections in a way that contributes to the goal of improved care for patients.

Contents

List of publications

Study 1: Husabø, G., Nilsen, R. M., Flaatten, H., Solligård, E., Frich, J. C., Bondevik, G. T., Braut, G. S., Walshe, K., Harthug, S., & Hovlid, E. (2020). Early diagnosis of sepsis in emergency departments, time to treatment, and association with mortality: An observational study. *PloS One*, 15(1), e0227652.

Study 2: Husabø, G., Nilsen, R. M., Solligård, E., Flaatten, H., Walshe, K., Frich, J. C., Bondevik, G. T., Braut, G. S., Helgeland, J., Harthug, S., Hovlid, E. (2020). Effects of external inspections on sepsis detection and treatment: a stepped-wedge study with cluster-level randomisation. *BMJ Open*, 10(10), e037715.

Study 3: Husabø, G., Teig, I. L., Frich, J. C., Bondevik, G. T., & Hovlid, E. (2020). Promoting leadership and quality improvement through external inspections of management of sepsis in Norwegian hospitals: a focus group study. *BMJ Open*, 10(11), e041997.

Lists of tables and figures

Tables

Figures

List of abbreviations

CDC: Centers for Disease Control and Prevention

CI: Confidence interval

CQC: Care Quality Commission

CRT: Cluster randomized trial

ED: Emergency department

EHR: Electronic health record

GEE: Generalized estimating equations

ICD-10: International Classification of Diseases, Tenth Revision

ISO: International Organization for Standardization

LNA: Large-N analysis

MAR: Missing at random

ME: (Multilevel) mixed effect

MICE: Multiple imputation for chained equation

NBHS: Norwegian Board of Health Supervision

NPR: National Patient Registry

OECD: The Organisation for Economic Co-operation and Development

OR: Odds ratio

REK nord: Regional Ethics Committee of Norway North

SIRS: Systematic inflammatory response sign

SNA: Small-N analysis

SW-CRT: Stepped wedge cluster randomized trial

1. Background

"The County Medical officer threatens with inspection if the municipality does not bring the staffing crisis in the health department under control," read the lead paragraph of a December 4, 2020 article from the Norwegian Broadcasting Company. The news story was related to the challenges that the City of Bergen had experienced after most of the public health physicians had quit their jobs amidst the COVID-19 pandemic.[1] Its portrayal of the role of inspections is a familiar one: inspection as a tool of uncovering wrongdoings, maybe even a form of punishment. But then there is the other face of inspections: inspections as a tool for improvement. We trust the inspections to contribute to making the healthcare services better and safer. This thesis tries to answer if and how inspections can achieve these latter goals.

Reviews of the existing research suggest that we do not have an adequate understanding of how inspections mediate change in care delivery in healthcare organizations and what the effects of inspections are (Flodgren, Goncalves-Bradley, & Pomey, 2016; Hovlid, Braut, et al., 2020). Yet, inspections and other external assessment schemes are widely used as strategies for improving healthcare quality (Shaw, Groene, & Berger, 2019). In order to maximize the value of inspections as a tool for improvement, we need studies of how and to what extent inspections impact the internal improvements in inspected organizations. The thesis addresses these questions, reporting the results of research into a nation-wide inspection by the Norwegian Board of Health Supervision (NBHS) of sepsis diagnosis and treatment in emergency departments in Norwegian hospitals from 2016 to 2018.

This introductory chapter is intended to provide a general overview of NBHS and the sepsis inspection. It first describes regulatory governance of healthcare in Norway and NBHS's approach to inspections. It then briefly describes the pathology of sepsis and

[1] Sommerfeldt, Rognsvåg, and Anthun (2020). To clarify: This is a journalistic framing of the story, and nothing suggests that the County Medical Officer actually had presented the possibility of launching an inspection as a "threat".

explains why NBHS chose to conduct inspection of sepsis care in emergency departments. It also sketches out how the sepsis inspection was carried out. Finally, it describes the organization of the thesis and the outline of the remaining chapters.

1.1 Health regulation in Norway

As one of several semiautonomous subordinate agencies of the Norwegian Ministry of Health, NBHS is responsible for supervising public and private providers of health care, child welfare, and social services (Ministry of Health and Care Services, 2019).

NBHS is primarily involved with the supervision and control of service providers and the conduct of health personnel. Many of the tasks related to monitoring health care and public health, as well as most tasks related to standard-setting and crisis management, are the responsibilities of other agencies of the Ministry, such as The Directorate of Health and The Institute of Public health (see Figure 1). NBHS's supervisory role was strengthened and "purified" in the early 2000s (Ministry of Health, 2002). This reform was part of a broader international trend where governments sought to move from integrated models, where regulation was one of several tasks delegated to the agencies, towards more specialized "single purpose" agencies for supervisory tasks (Lægreid, Roness, & Rubecksen, 2008).

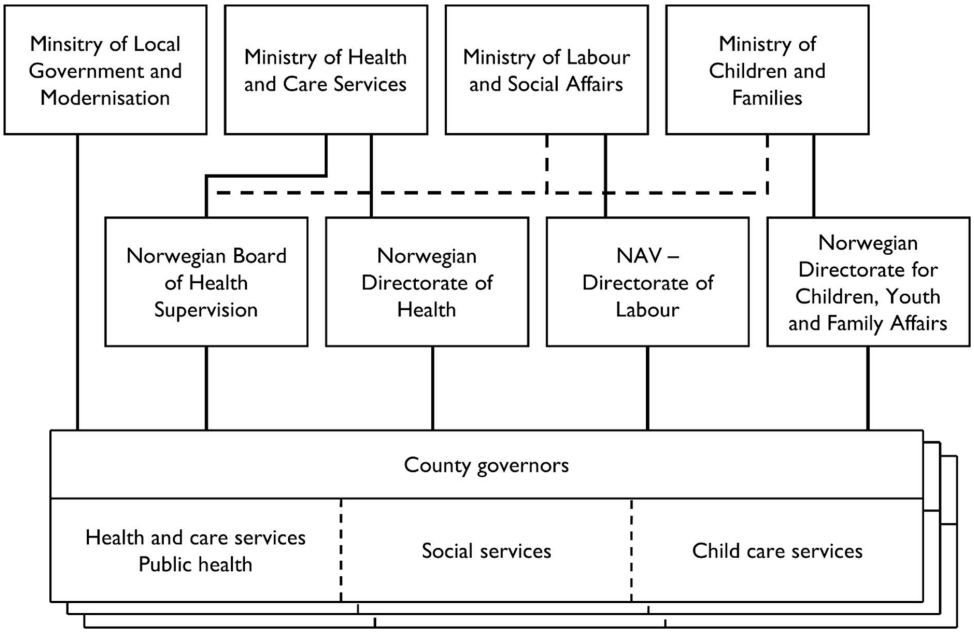


Figure 1 Ministries and state agencies involved in health and welfare policy and their relationship to the County governors. From NBHS (2020a).

In Norway, health care, social services, and child welfare services are mandated by law to be professionally sound and safe for patients and users. Hospitals and other healthcare organizations are also required to establish internal control and quality systems.[2] The supervisory role of NBHS therefore involves determining whether the services meet requirements laid down in legislation and whether they have adequate systems for internal control. Additionally, quality improvement is a *prima facie* goal for NBHS: To "contribute to strengthening the safety and quality of the services" is mentioned in the first line of the national budget's chapter concerning NBHS's mission (Ministry of Health and Care Services, 2019, p. 176).

[2] As an example of the legal requirements of safety and quality in services, see The Health Personnel Act (1999). The requirements for internal control systems are found in Regulation on management and quality improvement (2016). For an introduction to how requirements of internal controls have been introduced into Norwegian healthcare regulation, see Øyri, Braut, Macrae, and Wiig (2020).

The main supervision methods employed by NBHS are "area surveillance", incident-related inspection, and proactive inspections. Area surveillance involves collecting and analyzing information in order to assess the quality of services and whether they meet the legal requirements. Incident-related inspections are investigations of possible deficiencies in services, based on information from users, relatives, other government agencies or the media. From 3000 to 4000 such inspections are instigated each year. In addition, specialist healthcare services are required to report serious adverse events to NBHS, and all reported adverse events are investigated by NBHS. Proactive inspections are performed using a system audit approach and rely on information from documentation, interviews, and sample testing. Each year between 200 and 400 proactive inspections are performed. Each inspection results in a report describing the inspection process and important findings, including any deficiencies that have been uncovered. The reports are made available for the public through the Internet (Norwegian Board of Health Supervision [NBHS], 2019).

NBHS's work with inspections is coordinated with the County Governors, who are tasked with carrying out the inspections. The County Governors are agencies of the central government, one for every county in Norway (at the time of the sepsis inspection there were 18 counties), and each office includes a position as Chief County Medical Officer. The two agencies – NBHS and County Governors – are collectively called "the supervision authorities" (Ministry of Health and Care Services, 2019).[3]

Both NBHS and the County Governors can initiate inspections. Each year, NBHS prioritizes two to three areas for nation-wide inspections, based on risk assessments. NBHS coordinates and plans nation-wide inspections, develops inspection methodology, and aggregates and analyzes the findings from inspection reports. During

[3] The formal relationship between the County Governors and NBHS has changed several times over the past decades. In 2003 NBHS was described as a central professional body for the Chief County Medical Officers (Ministry of Government Administration and Labour, 2003). An organization of "NBHS within the County" was introduced, whereby the work with inspections was performed independently of the County Governors. This arrangement lasted until 2012 (Braut, 2019).

the planning stages of a nation-wide inspection, NBHS produces common guidelines for the inspection in cooperation with the County Governors (NBHS, 2019).

While it is not within the scope of this thesis to give a comprehensive overview of different national regimes for regulatory governance, it is worth noticing some important aspects of the Norwegian system as compared to that of other European countries. In Norway, certification and accreditation programs have not had a strong foothold. It is one of relatively few countries in Europe that do not have a national accreditation program (Shaw et al., 2019), and though certification and accreditation schemes have been adopted by some hospitals and healthcare services, this has been done on a voluntary basis and following local initiatives (Johannesen & Wiig, 2017). Thus, NBHS is the only body performing external assessments of patient care in hospitals on a nation-wide basis in Norway. NBHS does not, however, have registration, authorization, or licensing of hospitals as one of their tasks, as is common for many supervision authorities (Sutherland & Leatherman, 2006). Authorization of new hospitals is done by the Ministry of Health.

1.2 The sepsis inspection

The sepsis inspection was a nation-wide inspection into the emergency departments of 24 hospitals throughout Norway.[4] The theme of the inspection was based on a risk assessment. Through earlier nation-wide and incidence-based inspections, NBHS had identified that patients in hospital emergency departments did not always receive adequate care. They found prolonged times to diagnosis and treatment that increased the risk of detrimental patient outcomes, and patients with sepsis were considered to be one of the groups that were most at risk (NBHS, 2018).

[4] Conventionally, NBHS uses the singular form “inspection” when describing a nation-wide inspection, even though it consists of a series of inspections at different sites, each reaching its conclusion in a separate inspection report.

1.2.1 Sepsis

The current definition of sepsis is a *life-threatening organ dysfunction caused by a dysregulated host response to infection* (Singer, Deutschman, Seymour, & et al., 2016), or, in laymen's terms: "a life threatening condition that arises when the body's response to an infection injures its own tissues and organs" (Czura, 2011, p. 2). This definition quite recently superseded the definition that was in widespread use when the present research project was being planned: that of a *systemic response to infection*, marked by the presence of both infection and a systematic inflammatory response sign (SIRS) (Bone et al., 1992; Levy et al., 2003). The clinical criteria for sepsis commonly used in Norwegian hospitals at the time the present study started was suspected or documented infection along with two or more of the following conditions: (1) fever or hypothermia, (2) elevated heart rate, (3) tachypnea (elevated respiratory rate), or (4) alteration of white blood cell count.

NBHS chose sepsis treatment as a subject of these inspections because the condition is deemed critical, judged by criterions of severity and incidence. Estimated at 48.9 million yearly incident cases and 11 million sepsis-related deaths globally, sepsis is one of the leading causes of death world-wide (Rudd et al., 2020).[5] The annual population incidence of hospitalized sepsis in Norway is estimated to be 140 out of 100 000 inhabitants. Around one out of four patients died while hospitalized for sepsis (Knoop, Skrede, Langeland, & Flaatten, 2017).

The Surviving Sepsis Campaign issued the "Barcelona declaration" at the annual congress of European Society of Intensive Care Medicine in 2002. This declaration was intended as a call to action for increasing awareness about sepsis, and for improving and disseminating clinical guidelines for diagnosis and treatment of the condition (Slade, Tamber, & Vincent, 2003). The campaign published the first set of

[5] To these high mortality figures, we should add a qualifier: Patients with sepsis are often frail and dealing with comorbidity. Most sepsis-related deaths are not attributable to sepsis and not possible to prevent (Singer, Inada-Kim, & Shankar-Hari, 2019).

guidelines the following year (Dellinger et al., 2004).[6] The guidelines recommended early recognition, continuous monitoring of vital signs, initiation of empiric antibiotics as soon as possible, and early goal directed treatment for patients with severe sepsis. Some of these care processes have long been mainstays of recommended sepsis care, such as continuous monitoring of vital signs, while others, such as goal directed treatment, were based on then-recent research (Nguyen et al., 2006; Noah, 2014).

In emergency settings, diagnosing sepsis can be challenging. Previous research has found that sepsis often is not recognized early enough (Morr, Lukasz, Rübig, Pavenstädt, & Kümpers, 2017). In emergency departments, different groups of medical professionals cooperate under sometimes stressful conditions, resulting in increased risk of loss of information. As patients entering the emergency department often are transferred to other departments in the hospital, it is also difficult to conduct a systematic follow-up of patients presenting to the emergency department with sepsis. Hence, an adverse outcome will not necessarily come to the attention of the personnel who cared for the patient in the emergency department.

1.2.2 The inspection process

Six regional inspection teams from the County Governors performed the inspections. Each team had three to four inspectors from the County Governors' health and welfare departments and one additional team member who had the role of an external medical specialist.

Each inspection team was tasked with choosing four hospitals within their region for inspection. The hospitals were notified of the inspections from two to five months ahead of the site visit. Like all proactive inspections, the sepsis inspection was carried out as system audits, following the ISO standard for quality audits (International Organization for Standardization, 2012). The data used for assessing the quality of care

[6] The guidelines for adult patients have since been updated in 2008, 2012, and 2016 (Dellinger et al., 2008; Dellinger et al., 2013; Rhodes et al., 2017).

were collected through document reviews, interviews with management and healthcare personnel, and review of the electronic health records of 66 patients presenting to the emergency department with sepsis. This data collection was usually completed during the site visit over a period of two days for each hospital. All inspection teams used a common written guide for their inspection. The guide included a set of indicators for sepsis care, information about how data should be collected for these indicators, and the criteria against which the indicators should be assessed. The site visit concluded with a meeting where the inspection team presented the main findings to the hospital management and staff.[7]

After the site visit, the inspection team wrote a report for each inspected hospital. The report described the nonconformities identified during the inspection. After a draft version of the report had been sent to the hospital for comments, a finalized version was published on the Internet. The inspection team returned to the hospital twice, at eight and fourteen months after the site visit, for new reviews of health records. Each follow-up review included the last 33 patients presenting with sepsis, and a report describing the findings was sent to the hospital. After the initial site visit and the follow-up reviews, the hospitals were required to develop and implement quality improvement measures in order to close out the identified nonconformities.

1.3 Chapter outline

This is an article-based thesis, which means that it is divided into two parts. The three published studies with appendices are included in the second part of the thesis. This first part is the synopsis: the contextualization, presentation, and discussion of the studies.

The first four chapters of the synopsis cover the background and research design of this project. Following the overview of the regulatory role of NBHS and the sepsis inspection in the present chapter, chapter 2 gives an overview of theory and previous

[7] The processes of selecting hospitals for review and for the reviews of electronic health records are described in more detail in section 4.3.2.

research about healthcare regulation and inspections. Chapter 3 presents the research aim and research questions that have been guiding the work with this thesis. It also gives an overview of the three studies included in the thesis. Chapter 4 describes the methods used to collect and analyze data.

The results of the research are presented and discussed in the last three chapters of the synopsis. In chapter 5, I present the main findings of the three studies. I go on to discuss these findings in chapter 6, which starts out with a methodological discussion before I discuss the findings in light of the research questions posed in chapter 2. I conclude the synopsis in chapter 7 with some thoughts about the thesis's contributions to inspection practice and future research.

2. Theory and previous research

In this project theory plays two important roles. First, it serves as a guide to designing the research and understanding the phenomenon that is being studied. Second, it serves as an aid in an analytical generalization of our findings (Yin, 2018). In the previous chapter I described NBHS as a regulatory agency. This chapter starts out with a brief overview of theory about regulation. I next explain the role of regulation in health care, and I also take the opportunity to explain why I have chosen to use the translation "inspection" for the Norwegian term "tilsyn". The remainder of the chapter describes how inspections can influence quality improvement in healthcare organizations, both from a theoretical point of view and from the view of what previous research has found regarding the effects of inspections on quality of care.

2.1 Regulation

The term "regulation" covers a vast area, almost to the point of including all kinds of activities where someone intentionally, or even unintentionally, does something to influence others (Baldwin, Cave, & Lodge, 2011; Koop & Lodge, 2017; Majone, 1994; Selznick, 1985). Such a definition would in my opinion render the whole concept useless. For the sake of this thesis, it seems sensible to follow Walshe (2003), who argues that regulation should refer to those cases where an agency, in the public interest, has been granted centralized powers to act as a third party to transactions or inter-organizational relationships. Regulation is achieved through processes of control that are closely involved with the regulated activities (Selznick, 1985). Regulation does not have to be undertaken by the government, but it usually has at least governmental support or authority.

Western economies and societies saw an increase in regulation from the 1980s. Some authors have interpreted this increase as a result of the reduction of direct state control (Majone, 1994) and the rise of New Public Management (Power, 1997, 2000). Severing the role as a direct service provider, the state turned to instruments like audit and

inspection as ways of controlling service provision. Seen from this perspective, the rise of regulation is the result of a long-term institutional development of political and bureaucratic arrangements. Traditional notions of political accountability have been seen as less and less effective as the complexities of government have increased, thus requiring the introduction of new regulatory strategies to hold the organizations accountable (Walshe, 2003).

Literature on regulation draws a distinction between three control components or methods that form the basis of any control system: information-gathering (a detector), standard-setting (a director), and behavior-modification (an effector) (Hood, Rothstein, & Baldwin, 2001; Hood, Scott, James, Jones, & Travers, 1999). Regulators gather information through a mixture of approaches: They sometimes actively collect information, and they sometimes rely on other parties to provide information. There is also a diversity of approaches to standard-setting. Some standards are based on technical and research-based knowledge, often imported from other domains or countries, while others are arrived at through a more *ad hoc* approach.

When it comes to behavior modification, the division between strategies of *deterrence* and *compliance* has been central to the theory of regulation. While strategies of deterrence aim to detect violations of law, determine who is responsible, and penalize violators, strategies of compliance aim to prevent violations without carrying out punishments (Reiss, 1984). Implied in the compliance strategy is a tendency to view the regulated organizations as trustworthy or at worst well-intended but organizationally inept. Regulators choosing deterrence strategies, on the other hand, often see organizations as "amoral calculators" whose non-compliance stems from selfish economic calculations (Kagan & Scholz, 1984). Seen through the lens of game theory, the choice of whether to pursue a hard line poses a dilemma for the regulator.

A hard line might put an unnecessary burden on the regulated organization, while a soft line might lead to regulatory capture (Scholz, 1991). [8]

From the 1990s on, regulation theory moved beyond the dichotomy of deterrence and compliance into more flexible notions of regulation (Walshe, 2003). The idea of responsive regulation has been an important contribution. It is founded on the view that regulators will be more successful if they are "benign big guns" (Ayres & Braithwaite, 1992, p. 19) with access to a wide array of sanctions that enables them to pursue strategies of both persuasion and punishment. Responsive regulation presupposes the development of internal control systems within the organizations, and it highlights risk assessment as a central element in the regulatory strategy (Power, 2007). The idea is illustrated through the image of the pyramid of enforcement strategies. One iteration of this pyramid has "self-regulation" at its base and "command regulation with nondiscretionary punishment" at its apex. Regulatory action is thought to start at the base of the pyramid, and escalates up through the pyramid with ever more insistent enforcement strategies only if the lighter options available at the base of the pyramid do not secure compliance (Ayres & Braithwaite, 1992).

The theory of responsive regulation has later been expanded into other strategies and labels. One of these strategies is *smart regulation*, which includes a broader range of instruments and regulatory actors, all the while keeping with the idea of the enforcement pyramid (Gunningham, 2010). Another strategy is *really responsive regulation*. Really responsive regulation emphasizes that the regulator must be sensitive to the context and outcome of regulation, including the organizational cultures of regulated organizations, the institutional environments in which the regulators act, and the performance of the regulatory regime (Baldwin & Black, 2008). It is important

[8] "Regulatory capture" has different meanings and connotations, such as identification or sympathy with the regulated industry, a wish to be employed in the industry in the future, or simply just as "absence of toughness" (Makkai & Braithwaite, 1992). Here, I use the term in the latter, broader sense, as describing a situation where the regulated organization exploits the absence of regulatory toughness to subvert the regulatory goals or steer the regulation in a way that is beneficial to themselves (Baldwin et al., 2011)

to note these are regulatory design *strategies* (Baldwin et al., 2011), and that real-life regulatory regimes combine different approaches (Gilad, 2010).

2.2 Regulation of healthcare organizations

Healy (2011) points to three arguments for regulating the healthcare sector: the goal of protecting the public, the desire to improve performance, and pressures for greater accountability.

Regulation of health care can take many forms. Sutherland and Leatherman (2006) have introduced a taxonomy organizing regulatory interventions used in health care according to the focus of the regulatory activity: profession, market, and institution. The professional focus consists of arrangements like licensing, registering, and credentialing healthcare workers. The market focus refers to the government intervening in the markets' supply or demand sides. This can be done in several ways, including by way of antitrust legislation, requirements of compulsory reporting, and gag rules for providing information about specific healthcare services. The institutional focus consists of directive approaches, i.e. setting targets and standards against which organizations can be held to account, and approaches relying on external oversight and assessment.

The subject of this thesis, inspection, is an external assessment scheme. External assessment could span activities like certification, accreditation, audit, peer-review, and ad hoc public investigations. Such activities are not always anchored in legislation or public policy. They could be voluntary activities that are initiated neither to meet government requirements nor to escape the potential ire of government in the future but rather to gain the trust of consumers. In that case it is doubtful that we could call this a form of regulation. External assessment is, however, part and parcel of modern government. In sociological terms, external assessments represent acts of inscription, rendering the reality manageable and diagnosable (Latour & Woolgar, 1986; Rose & Miller, 1992).

Healthcare regulation draws in unequal measures on strategies of deterrence, compliance, and responsiveness. The thrust of regulation activities is generally thought to have shifted from compliance towards responsive and reflexive regulation, though the extent of this shift has varied from sector to sector within health care (Leistikow, 2018; Leistikow & Bal, 2020). Inspection always retains an element of enforcement, as it, by definition, is a tool governments use in order to control whether legal requirements are fulfilled (Åsprang, Frich, & Braut, 2015).

Much theorizing around regulation has to do with private institutions being regulated by the government with the stated intention to make sure that those institutions are behaving in line with requirements. Or put in economic terms: The government intervenes in the market using regulation with the stated intention to correct market failures. When discussing NBHS's inspection of sepsis care in hospitals in light of regulation theory, it is important to note that this is a form of regulation *inside* government (Downs, 1967; Hood et al., 1999), where a governmental body, the NBHS, regulates government owned and run bodies, the hospitals. This form of governmental self-regulation is fundamentally different from situations where the government regulates privately held companies. This means that we must look beyond traditional economic theories, which rely on assumptions of profit maximization and the opportunity of passing on compliance cost to the consumers (Konisky & Teodoro, 2016).

2.2.1 Translating "tilsyn"

Before moving on, we need a brief note on terminology. The subject of this thesis is an external assessment scheme that in Norwegian is called "tilsyn", which is a term also used in neighboring Scandinavian countries Sweden and Denmark. I have chosen to translate this term into "inspection". According to Åsprang et al. (2015, p. 228), the word *tilsyn* has no precise English equivalent but aspects of *tilsyn* are captured by a

number of English words, like supervision, inspection, control, and audit.[9] Complicating the attempt to translate *tilsyn* into English, is the Norwegian practice of naming supervisory authorities *tilsyn*, as is the case for the NBHS, whose Norwegian name is *Helsetilsynet*.[10]

To unsnarl this semantic hodgepodge, we must first outline what kind of assessment activity *tilsyn* is. To do this, we will turn to a recent review of external assessment strategies in healthcare improvement: Shaw et al. (2019) have delineated three main approaches: accreditation, certification, and supervision.[11] They point out four common features of these approaches: (1) a focus on healthcare provider organizations rather than individuals, (2) a focus on organizational structures and service delivery processes (rather than resource inputs), (3) assessment against published standards and criteria, and (4) an aim to improve safety and quality of care. *Tilsyn* checks all these boxes.[12] The approaches are differentiated by, among other things, who performs the assessment and the standards used to assess the organizations. In contrast to accreditation and certification, which are conducted by independent bodies in accordance with accreditation programs and international standards, supervision is conducted by

[9] Etymologically speaking, many of these words share a similar background. A literal translation of *tilsyn* into English would be "see to [it]", while inspection is "looking into" (Inspection, 2020) and supervision is "oversee".

[10] Other examples include *Datatilsynet* (The Norwegian Data Protection Authority), *Mattilsynet* (The Norwegian Food Safety Authority), and *Arbeidstilsynet* (The Norwegian Labour Inspection Authority).

[11] Another, related, external assessment scheme not mentioned by Shaw et al., is that of performance audit. In Norway, performance audit is performed on the state level by the General Auditor, and on county and municipal level by public or private auditors. Like NBHS's systemtilsyn, performance audits are systems audits. There are however evident differences between their approaches. Performance auditors are generalist organizations not confined to specific sectors. Thus, performance audit within the healthcare organizations can be characterized as a form of out-group regulation (Hood et al., 1999). Another difference between the two approaches is that performance auditors arrive at audit criteria through a more flexible process than the NBHS does, having the opportunity to choose considerations of efficiency, or political decisions and goals, in addition to criteria derived from law and medical standards. Yet another difference is that inspectors have institutionalized ways of escalation not available to auditors (Power, 1997). Furthermore, while NBHS and other tilsyn organizations in Norway are government agencies, controlled (albeit at arms-length) by the executive branch, auditors are controlled by the legislative bodies, either at state or municipal level.

[12] I refer here to what was described in chapter 1 as proactive inspections. The first two features listed by Shaw et al. will not apply to all incident-related inspections, which may – though by no means always or only – focus on the individual personnel, rather than the organization.

government regulatory agencies that assess performance against standards set by legislation.

Out of the three approaches that Shaw et al. laid out, supervision is the one that most appropriately fits the NBHS's *tilsyn*. As explained in chapter 1, the Norwegian government does not use accreditation to regulate healthcare organizations. Neither is there a requirement of certification of health organizations. In Norway therefore, the differences in governance and legal basis establishes a clear demarcation between *tilsyn*, certification, and accreditation.

I use the term "inspection" instead of "supervision". Even though the NBHS itself at times seems to prefer the term (NBHS, 2019), "supervision" can be thought to signify a continuous "watching over", while I aim to denote a specific assessment process with defined start and end points. Opting to use the term "inspection" brings us in line with terminology used in Flodgren, Goncalves-Bradley, and Pomey's systematic review (2016) and in research papers studying schemes similar to the sepsis inspection.[13] More importantly, they serve different purposes. While "supervision" conveys the prescribed role of the regulator towards the regulated organization, "inspection" refers to the activity of performing an external assessment of a particular area of services. Shaw et al. (2019) use the term inspection in such a way, describing inspection as a method of assessment used within supervision schemes.[14]

A working definition for the purposes of this thesis is then that *[proactive] statutory inspection is a form of external assessment, led by the supervisory authorities, using a system audit methodology*.

[13] See for instance Boyd, Ross, Robertson, Walshe, and Smithson (2018); Castro-Avila, Bloor, and Thompson (2019); Hovlid, Braut, et al. (2020); Hovlid, Høifødt, Smedbråten, and Braut (2015); Hovlid, Teig, Halvorsen, and Frich (2020); Schaefer and Wiig (2017); Toffolutti, McKee, and Stuckler (2017); Walshe (2003)

[14] There is also a close affinity between *tilsyn* and the Norwegian term *inspeksjon* (i.e. the most direct Norwegian translation of the English "inspection"). In fact, in the Norwegian context the usage of *tilsyn* as a term describing inspection has itself replaced *inspeksjon*, which historically was more frequently used (Kringen, Lindøe, & Braut, 2015).

This definition allows us to identify inspection as a specific form of external assessment defined by its loci of control (government) and methodology (system audit).

2.3 Inspections and quality improvement

Along with providing assurance that minimum standards are met and accountability for levels of performance, improvement of quality and performance is a primary goal for healthcare regulation (Sutherland & Leatherman, 2006).[15] This is, as we have seen in chapter 1, also an explicit goal for NBHS. The meaning of "quality improvement" does not only refer to achieving improved services (however such improvements might be measured and assessed). It hinges just as much on the efforts to make improvements. As a standard definition of quality improvement reads:

> the combined and unceasing efforts of everyone—healthcare professionals, patients and their families, researchers, payers, planners and educators—to make the changes that will lead to better patient outcomes (health), better system performance (care) and better professional development (learning) (Batalden & Davidoff, 2007, p. 2)

This definition also highlights several important dimensions of "quality" itself, as the term is commonly used within health care. [16] It explicitly mentions the professions, the system of care, and the outcome. We need to understand the relationship between these elements if we are to understand the quality of care (Donabedian, 1988). We see here that this understanding of quality has a close affinity to the term "performance", which has transformed a predominantly financial term to a more multidimensional concept encompassing elements such as quality and customer satisfaction (Pollitt, 2006).

[15] It should be noted that in practice, the purposes of healthcare regulation are often not clear cut: Regulation may also be a way of distancing the government from a regulated organization or a token political response to complex problems (Walshe & Boyd, 2007).

[16] According to Dahler-Larsen (2019, p. 5), it "may be futile to seek any core meaning inherent in the term [quality]" other than its common denominator: "Quality is positively loaded".

Healthcare quality, as any form of public service quality, has to do with meeting the users' needs within the requirements of good practice and in a cost-efficient manner (Øvretveit, 2009).

Designing and conducting inspections with the goal of quality improvement involves aspects of all the three control components standard-setting, information-gathering, and behavior-modification. Standard-setting, or direction, has to do with communicating expectations to the healthcare organizations. These are usually written statements in the form of standards, rules, or regulation, and they serve both to make the inspections fair and transparent and to be used later for information gathering and behavior modification (Walshe, 2003).

Information-gathering, or detection, is perhaps the most important of the three components when it comes to inspections. In an inspection of quality of care, we can think of the inspection as a quality assessment. Quality assessment can by itself be a legitimate goal of healthcare regulation. It also serves other purposes, for instance defining a baseline care, against which the healthcare organizations and other interested parties can measure progress in the future (Sollecito & Johnson, 2019).

The third component, behavior-modification (or enforcement), is what the inspection authorities do to instigate changes in the inspected organization. According to Walshe (2003, pp. 210-211) it "covers everything from having a friendly word informally with a senior manager about a problem to fining, 'naming and shaming' or even delicensing the organization". "Enforcement" may therefore be a somewhat misleading term for the process of changing behaviors. Indeed, it could be argued that all three control components, direction, detection, and enforcement, can affect quality improvement in the inspected organizations. Accordingly, we now turn to look more specifically to the topic of change mechanisms.

2.3.1 Change mechanisms

Researching how inspections might lead to quality improvement, we need to identify the mechanisms by which they do so – i.e. the recurrent processes that link inspections and improvement and explain how one leads to another (Mayntz, 2004).

Our search for such mechanisms forces us to break open the black box of what is going on during an inspection. Furthermore, it leads us to look at what happens inside the healthcare organizations. We cannot expect to uncover law-like causal mechanisms. As Pawson, Greenhalgh, Harvey, and Walshe (2005) have pointed out, the efficacy of interventions related to regulation and inspections depend on context and implementation and should therefore not be thought of as "magic bullets".

Walshe (2003) points out that even if regulation is an external approach to quality improvement, it relies on internal approaches in order to achieve the desired results. That is, while other external approaches to quality improvement like economic incentives and legislation can fail to consider organizational culture and mechanisms for organizational change, inspections often focus on the particular issues of the inspected organization.

In an evaluation of the approach of England's Care Quality Commission (CQC) to inspections, Smithson et al. (2018) present a framework for describing how regulation can affect provider performance before, during, and after inspection. The framework has a typology of eight mechanisms for inspection impact primarily classified by the type of linkage between the inspection and changes in the services provided.

The first type of effects, *anticipatory impact*, refers to situations where the service providers implement measures to meet the regulator's expectations in advance of inspections. Anticipatory impact may represent genuine improvement efforts that seek to make substantial improvements to patient care, as documented in a study from Norway (Schaefer & Wiig, 2017) where inspectors reported that sometimes the organizations started improvement efforts already when they received the letter informing about upcoming inspections. Anticipatory effects may also represent efforts to "game" the system, where the focus is merely on meeting requirements rather than

improving services (Bevan & Hood, 2006). Strategic behavior is to be expected when any indicator linked to organizational output is presented as a goal or a target. According to what is known as "Goodhart's law", a measure that becomes a target ceases to be a good measure (Strathern, 1997).[17] In other words: Those whose actions are assessed according to the score of the target measure tend to engage actively in ways to influence that measure. An example of gaming comes in a study from England that found an apparent rise in level of cleanliness in hospitals in the months leading up to announced inspections of hospital cleanliness by CQC (Toffolutti et al., 2017). Shortly after the inspections, the levels of cleanliness dropped back to normal. The authors suggest that this pattern implies that the hospitals engage in a form of gaming, where staff members make special efforts when they know they are being inspected.

Smithson et al. identify three impact mechanisms that are related to the direct interaction between the regulator and service provider. *Directive impact* is the result of enforcement actions taken by the regulator, such as requiring the service provider to act in a specific way. *Organizational impact* covers the organizational developments that do not come about from specific directions from the regulator but rather from the providers' reflection and analysis following inspections. Health care is complex, often marked by a high degree of uncertainty and dependent on interaction between different individual actors (Plsek & Greenhalgh, 2001). The types of organizational developments that can come about following inspections include changes in internal team dynamics, leadership approaches, and organizational culture. *Relational impact* has to do with the relationship between the inspection teams and the management and staff at the service providers. This is described as a "social process" where informal or soft actions impact the providers.

The last four impact mechanisms described by Smithson et al. are not related to the direct contact between the regulator and the service provider. *Informational impact*

[17] Writing within the field of monetary policy, Goodhart (1981) argued that once it has been chosen for control purposes, any indicator will tend to break down, meaning that the observed statistical regularities associated with the indicator will cease to exist. Hence, this mechanism or "law" does not necessarily predict deceitful "gaming" of the target, it merely states that the validity of the measure is influenced by its use as a target.

refers to the effect of publicizing information about the performance of the service provider. Such information can for instance be picked up by the media and influence the public's perception of the service provider. A related effect is *stakeholder impact* where other stakeholders take action or interact with the service provider as a result of inspections. *Lateral impact* refers to inter-organizational interactions such as service providers learning from each other through sharing information in networks. The last category, *systemic impact,* happens when aggregated findings from inspections are used to instigate systemic changes across multiple service providers.

Though there are grey areas between the eight mechanisms in this framework, organizational and relational impact are the ones that receive most attention in this thesis. In their narrative review of research on change in healthcare organizations subject to external assessment, Hovlid, Braut, et al. (2020) explore such mechanisms further. They point to guidance from the inspection team on how to follow up inspection findings as a way the inspection team can help facilitate improvement (see for example Benson, Boyd, & Walshe, 2006). However, Hovlid, Braut, et al. argue, the findings must be perceived as valid and reliable by the relevant stakeholders, for instance by focusing on practice-specific issues (Campbell, Chauhan, & Lester, 2010) and using terms and concepts that professionals understand (Åsprang et al., 2015). Ensuring the clinicians' "ownership" of indicators can be crucial (Collopy, 2000).

The emphasis on valid and reliable findings echoes an important theme in quality improvement research and practice. Measurements have become a cornerstone of quality improvement and a rallying point for regulatory approaches of all kinds (Berwick, 1998). The rationale behind setting measurable targets is that they direct attention towards objectives, and that they are thought to lead to greater persistence and foster commitment to reaching the objectives (Sutherland & Leatherman, 2006). In practice, healthcare and public services are flush with surrogate indicators that are only problematically related to the goals that the organizations are meant to achieve (Lipsky, 2010). The search for better indicators involves several steps, including finding the indicators that are the best predictors of risk to patients, exploring impacts of different

scoring systems, and finding ways to link quantitative and qualitative information (Bardsley, 2017).

While inspections formally address the executive chain of command in organizations, the *relational impact* implies that the inspections can impact staff as well as management. One reason why this is the case in inspections of healthcare organizations, may be that these organizations often have strong professional groups that resist managerialism (Walshe, 2003). To overcome challenges such as this, the inspection team should have good communication skills and knowledge of the inspected area (Arianson, Elvbakken, & Malterud, 2008; Walshe & Phipps, 2013).

These arguments point to the important role organizational culture plays in quality improvement (Braithwaite et al., 2010). When a group engages in shared learning and problem solving, beliefs and values come to be taken for granted (Schein & Schein, 2016). Beliefs and values are at the heart of quality improvement, as a key issue in quality improvement is how to convince people who are either recalcitrant or preoccupied with other tasks to join forces and devote their time to the improvement effort. Hovlid, Braut, et al. (2020) maintain that the inspections can contribute to enhancing communication about clinical work and, hence, the understanding that the people within the organization have of the clinical system and its interdependencies. Knowledge of the clinical system and interdependencies can then be the foundation of the planning and implementing of systemic improvement measures.

2.3.2 Measuring the effects of inspection

Pointing to the influx of quality assessment techniques that had taken place in health care in the 1990s, Brennan (1998, p. 710) stated that it was critical to find the answer to the following question: “Is there measurable evidence that regulation is working to improve the quality of health care?” When we look at the state of research into health-care regulation more than 20 years on, this question still seems pertinent.

An important point of reference is a systematic review of external inspection conducted by Flodgren et al. (2016). This is a high-quality review[18] focusing on the effects of inspections in improving care and patient outcomes. Noting the lack of methodically robust studies within this area, Flodgren et al. conclude that it is unclear whether inspections can improve professional practice and health outcomes. This conclusion is supported by Castro (2018). Reviewing the effects of external assessment as reported in 90 individual studies, including observational and quasi-experimental ones, she found that evidence of the effects of inspection and accreditation remains uncertain.

Newer studies have not been able to shake the picture painted in the aforementioned reviews regarding the effect of inspections on quality of care or patient outcomes. Two studies from the Netherlands by Wesselink and colleagues have assessed the effects of inspections using randomized controlled design. Investigating changes in the provision of smoking-cessation counseling by midwifery practices, Wesselink et al. (2014) found that their primary outcome measure, the use of a specific smoking-cessation intervention strategy, improved significantly after the inspection authorities had set a deadline by which all practices should comply with specific professional norms. However, they did not find significant effects on the use of the strategy following assessments by the inspection authority. The second study (Wesselink, Lingsma, Ketelaars, Mackenbach, & Robben, 2015), which was of diabetes care provided in care groups, found that neither care processes nor health outcomes improved following the inspections.

Three recent studies, using data on the effects of CQC inspections in England from the time period 2012/2013 – 2016, did not find significant effects, either. One of these studies used the comparatively robust method of controlled interrupted time series (Castro-Avila et al., 2019). This study found no positive effects of inspections on the rate of pressure ulcers and falls with harm in hospitals. In the second study (Allen, Walshe, Proudlove, & Sutton, 2019), the authors developed indicators of performance

[18] Teetering on the brink of infinite regress, Castro (2018) assesses the quality of systematic reviews assessing the qualities of studies assessing the quality of inspections and accreditations assessing the qualities of healthcare organizations. Flodgren et al. (2016) was the highest scoring out of seven systematic reviews.

for emergency departments. The researchers then compared performance with inspection ratings for acute hospitals, and they studied changes in performance from before to after the inspections. They found no association between performance and hospital rating, and no effect of the inspections on performance. The third study investigated inspections of maternity sites, using a similar method (Allen, Walshe, Proudlove, & Sutton, 2020). They found no clear evidence of any association between performance and inspection ratings, or of improvement in maternity sites that had received a poor rating, leading the authors to call into question the validity and reliability of the inspection ratings.

Though previous research has found little to suggest that inspections have an effect of improving patient care and patient outcomes, we should be careful with ruling out such effects altogether. Inspections are complex interventions, and their effectiveness depend on the context in which they are conducted and how they are planned and carried out (Brubakk, Vist, Bukholm, Barach, & Tjomsland, 2015). Furthermore, studies investigating effects run into the problem of "spillover-effects", where the inspection benefits those service providers who are not inspected as well (Wesselink et al., 2015). This would be an example of what is referred to as *lateral impact* in the framework of Smithson et al. (2018).

3. Research aims

This thesis is a part of a research project that examines the associations and links between external inspections and improvements in healthcare organizations (Hovlid et al., 2017). As described in the preceding chapters, quality improvement is an important goal of regulation and inspection schemes in health care. But though we have theoretical models describing how inspections might bring about improvement, there is still far from conclusive evidence to determine whether these theoretical suppositions bear out in practice.

The aim of this thesis, then, has been to find out if inspections of healthcare organizations can lead to improved work processes and health outcomes for patients, and to explore the mechanisms that link external inspections and internal improvement efforts. This overall research aim has been operationalized into three main research questions.

First, we needed to gain an understanding of the potential problems related to quality in the emergency departments and whether the inspection succeeded in targeting the quality problems appropriately. Having an overview of time to diagnosis and treatment would give a baseline from which we could study the effects of the inspection. Furthermore, assuming that the effectiveness of an inspection is related to the validity of its assessment, we wanted to find out whether the indicators for diagnosis and treatment were inherently associated, i.e., that time to treatment could be expected to improve if time to diagnosis improved. If documented, such associations could improve our understanding of how the inspection worked. Thus, the first research question was formulated:

> **Research question 1**: Are the care processes evaluated in the inspection relevant for assessing the quality of care in the emergency departments, and what can the care processes tell us about the performance of the emergency departments on a systemic level?

Second, given that previous research has come to uncertain conclusions regarding inspection effects, we wanted to find out if the sepsis inspection was able to affect these care processes and patient outcomes in a positive way:

> **Research question 2:** How have the inspection affected the performance indicators for diagnosis and treatment of sepsis, and for patient outcomes?

Finally, we found that it was important to get a better understanding of how the inspection worked. Answering research question 2 would only provide us with an estimate of the strength of the association between inspections and improvements. We additionally sought to explore the mechanisms by which the inspections could have affected improvement processes in the inspected hospitals. This would entail understanding how the inspection was implemented, what the circumstances of that implementation were, and how the involved actors reasoned:

> **Research question 3:** What is the role of inspections in bringing about improvement efforts in the emergency departments?

Each main research question was addressed in a separate study using data collected and generated from the sepsis inspection. For the first two studies, quantitative data and analytical approaches were used, while the third study used a qualitative approach. The research aims for the quantitative studies were operationalized according to the specific performance indicators used in the inspection, while the qualitative study was guided by a more open-ended and explorative research aim, as shown in Table 1.

Table 1 Overview of the three studies

	Study 1	Study 2	Study 3
Title	"Early diagnosis of sepsis in emergency departments, time to treatment, and association with mortality: An observational study."	"Effects of external inspections on sepsis detection and treatment: a stepped-wedge study with cluster-level randomisation."	"Promoting leadership and quality improvement through external inspections: a focus group study"
Aim(s) of the study	• To assess the timeliness of diagnostic procedures for recognizing sepsis in emergency departments • To estimate associations between diagnostic procedures and time to antibiotic treatment, and to estimate associations between time to antibiotic treatment and mortality	• To evaluate the effects of external inspections on hospitals' clinical processes for detecting and treating sepsis • To evaluate the effects of external inspections on length of hospital stay and 30-day mortality.	• To study how external inspections may foster clinical improvement.
Design	Cross-sectional	Stepped wedge	Focus group
Data /sample	1559 patient records from 24 hospitals	7407 patient records from 24 hospitals.	12 focus group interviews with a total of 47 participants from 4 hospitals and 4 inspection teams

Though this thesis is a retrospective assessment of the output and outcome of a specific government intervention, it should not be considered an across-the-board evaluation of the sepsis inspection.[19] Whereas it certainly evaluates aspects of the inspection, we have not operationalized evaluation criteria based on NBHS's stated goals of the sepsis inspection. Nor have we studied the process from the inspection's inception until its end in the manner of a process evaluation.

I consider this thesis to be a form of case study where the goal is to explore, describe, and explain how the phenomenon of inspections is related to quality improvement. The sepsis inspection, then, is a case of statutory inspections (as defined in section 2.2.1, a form of external assessment, led by the supervisory authorities, using a system audit methodology). This makes for a relatively specific type of intervention, which in the Norwegian context follows specific guidelines for conducting and reporting the supervision.

The research project did not influence which theme or area NBHS should prioritize, and the research project was launched at a time when the sepsis inspection seemed as a suitable candidate. Given that the project was intended to assess a nation-wide NBHS inspection, the choice would always be limited to one or two active inspection campaigns. It is nonetheless a fitting case that provides us with the opportunity to operationalize and assess changes in care processes and patient outcomes. (The merits of choosing the sepsis inspection as the theme of this research project is discussed further in chapter 6.)

In the next chapter I will go on to explain the overall research design and the details of the data collection and analyses in the three studies.

[19] Evaluation can be defined as a "careful retrospective assessment of the merit, worth, and value of administration, output, and outcome of government interventions, which is intended to play a role in future, practical action situations". (Vedung, 1997, p. 3)

4. Methods and materials

Built from the ground up, this chapter starts with the philosophical foundations of the thesis. I then present the overall research strategy, describing the overall case study framework of the thesis and presenting the stepped wedge approach that has guided the research. Next, I describe the nuts and bolts of how the data have been collected, before moving on to detail the analytical approach of the three studies.

4.1 Philosophical foundations

I believe methodology should not be used as a synonym for a method or a technique. Rather it refers to the systematic investigation of the reasoning applied to a field of study (Moses & Knutsen, 2012). Therefore, any discussion of methodology can potentially extend into an investigation of the ontological and epistemological underpinnings of scientific claims.

Before diving into further philosophical ruminations, I must explain my point of departure: I regard the sepsis inspection to be an organizational intervention. Though it assesses clinical care, the inspection also touches on organizational issues within the emergency departments. Furthermore, the mechanisms of the inspection – some of which I have described in chapter 2 – are organizational: In order to improve, it is necessary to make organizational changes. Therefore, the thesis is ultimately grounded in social, rather than clinical, science.

In this project, I claim a paradigmatic position somewhere in the middle of positivism and constructivism, in the vicinity of *realism*. Realism holds that the even though the social world includes knowledge itself, and thus cannot exist independently of knowledge, it should not be conflated with our experience of it (Sayer, 2000). That is, I ontologically accept the existence of a real world independent of our experience of it, though I concede that there are many layers to this reality, and that, epistemologically, my access to it is highly complicated (Moses & Knutsen, 2012).

For realists it is permissible to use the same methods as natural scientists. However, social science must operate in a double hermeneutic (Sayer, 2000). This means accepting the constructivist position that in social science, the object is a subject (Flyvbjerg, 2001), and that social interaction is contingent and open-ended (Bourdieu, 1977). Furthermore, it requires appreciating that social science is often preoccupied with interpretive understanding (*verstehen*) rather than explaining (*erklären*), the former being "a linguistic matter of grasping the content of familiar and unfamiliar forms of life" (Giddens, 1995, p. 240).

Subscribing to this realist philosophy, I sympathize with the research agenda represented by the realist slogan "what works for whom in what circumstances … and why" (Pawson, 2013, p. 15). In this thesis, this methodological agenda is corollary to an appreciation of the context-dependency and contingency of inspections and the need for a holistic understanding of the linkages between inspection and improvement.

4.2 Research strategy

As mentioned in chapter 3, I consider it useful to think of the thesis as a case study where the sepsis inspection represents a case of statutory inspections of healthcare providers. The project utilizes a mixed methods approach with a stepped wedge design implementation of the inspections at its core.

4.2.1 Case study

Case studies are sometimes seen as a subset of qualitative research methods (George & Bennett, 2005). Under this definition the present project would not be considered a case study, as it is centered around the collection and analysis of quantitative data. When using the term case study, I instead follow Yin (2018), who argues that case studies can include (and even be restricted to) quantitative data. According to Yin, the case study is "an empirical method that investigates a contemporary phenomenon (the 'case') in depth and within its real-world context" (p. 15), and the inclusion of different forms of data is central to the case study design. We have chosen an approach where

we collected quantitative and qualitative data concurrently, and then performed a connected mixed methods data analysis where the qualitative analyses were built on what we had learned from the quantitative analyses (Creswell & Plano Clark, 2011). This provides both a quantitative vantagepoint for assessing the association between inspections and changes in care processes and patient outcomes, and a qualitative vantagepoint for understanding how the inspections and changes in processes and outcomes are linked.

It is necessary to interject that I describe the three studies along with the synopsis as one study. Against this conceptualization one could argue that what we have is really three separate studies within a research project, and that the synopsis is a synthetization of the studies. This could call into question the usage of both the terms case study and mixed method. Though this is a reasonable objection, I find the usage of both terms warranted: The studies have been guided by a common research aim and the quantitative and qualitative data are collected simultaneously and intended to be complementary to each other.

An important step in describing a case study is defining and bounding the case (Yin, 2018). As explained above, we are dealing with a case of statutory inspections. It is naturally bounded by the theme of the inspection and the timeframe in which it was being conducted (2016 – 2018, with the earliest quantitative data dating back to 2014). Embedded in the case of the national sepsis inspection (the first level) are units of analysis on lower levels. One could think of the work of each of the six inspection teams on a regional level as a second level and of the inspections at each of the 24 hospitals as a third level. Comparing and contrasting each of the within-case subunits thus can give valuable information about how the inspections worked. This analytical approach is used in study 3, though the design stops short of a full-out comparative case study.

4.2.2 Stepped wedge

The stepped wedge cluster randomized trial (SW-CRT) design is a form of crossover cluster randomized trial (CRT) where clusters cross over between control and treatment at different time points (Hussey & Hughes, 2007). First conceived of in the late 1980s (The Gambia Hepatitis Study Group, 1987), and only used a handful of times until the mid-2000s, the stepped wedge design is a fairly recent research design (Brown & Lilford, 2006).

Let us untangle the disparate elements of the stepped wedge approach. "Cluster randomization" refers to the data sampling strategy. It means that randomization occurs at the cluster or group level, not at the level of the individual observations.[20] An example of cluster randomization would be to randomly select schools from a list of all the schools in a district and subsequently include every student from those schools in our sample. In a CRT, the treatment is administered to all observations or individuals within a cluster. If we introduced pancakes for lunch at half of the schools in our imaginary CRT, this would be a parallel CRT where the pancake schools (group A) are treatment clusters, and the other schools (group B) are control clusters. If we were interested in psychological well-being, we could let the students fill out questionnaires, thus measuring the effect of the cluster-level pancake intervention on the student level. The results would have to be obtained from all control and trial clusters at two points in time, before and after the pancake lunches were introduced.

Piggybacking on the previous example, suppose we had second thoughts about serving pancakes to only half of the schools, due to ethics or fear of irate students retaliating against us. We could choose to forgo the parallel design in favor of a crossover design. After a couple of weeks of pancake lunches, we would switch around[21] so that the trial clusters become control clusters and vice versa. Now, students in group B schools are being served pancakes while students in group A schools are back to the plain,

[20] The stepped wedge design has also been used in non-cluster trials where randomization of the treatment sequence is done at the individual level (Brown & Lilford, 2006). Here, I will only consider the cluster variant.

[21] Often, a "washout" period between is introduced treatment periods so that the effect of an intervention does not contaminate the subsequent measurements (Hussey & Hughes, 2007).

nutritionally adequate lunches. This way, all clusters are included in the study as both control and treatment groups.

In stepped wedge designs, like other crossover designs, all clusters are both control and treatment groups. However, whereas the traditional crossover trial has two time periods, one where group A receives the treatment and one where group B receives the treatment, the stepped wedge has multiple treatment periods (Hussey & Hughes, 2007). At the start of the trial, no clusters have received the treatment. Then, the clusters receive treatment one at a time in a randomized sequence. At the end of the trial, all clusters have received the treatment (Hemming, Haines, Chilton, Girling, & Lilford, 2015). Once again using our imaginary pancake CRT as an example, we start out with no one getting pancakes. Then we introduce pancake lunch to one randomly selected school each week until all the schools are serving pancakes.

Lawrie, Carlin, and Forbes (2015) give a formal definition of the stepped wedge design as (1) having at least three periods; and having each cluster (2) starting in control condition, (3) ending in treatment condition, and (4) changing from control to treatment exactly once. Within this definition, however, we can find different varieties of stepped wedge designs, and it has been suggested that the stepped wedge is not so much a specific study design as a family of designs (Haines & Hemming, 2018). One feature that distinguishes different types of stepped wedge models, is whether measurements are taken at each step or not. If we measured the happiness of all students, pancake eating or not, each week throughout the entire trial, we would have a *complete* stepped wedge design. If we on the other hand only measured their happiness the week they first received the pancakes and the week before that, we would have an incomplete design (Hemming, Lilford, & Girling, 2015). Stepped wedge designs also differ by how data on subjects are obtained. If we handed out questionnaires to the same students every week, we would have a cohort stepped wedge design. If we rather recruited a random sample of students for each round of questionnaires, we would have a cross-sectional stepped wedge design (Barker, McElduff, D'Este, & Campbell, 2016).

The stepped wedge is a study design that is especially suited for evaluating the effects of service delivery interventions. Hemming, Haines, et al. (2015, p. 1) describe it as a "a pragmatic study design that reconciles the constraints under which policy makers and service managers operate with the need for rigorous scientific evaluations". Three specific advantages bear mentioning here: First, if it is practically, ethically, or politically untenable to withhold the intervention from parts of the population, the stepped wedge ensures that the intervention reaches all clusters. Second, if it is not possible to treat all clusters simultaneously, it allows for a systematic and, by way of randomization, ethically justifiable principle of administering the intervention. Third, the stepped wedge is a suitable alternative to cluster crossover designs if it is impossible to reverse or cancel the effect of the treatment.

The sepsis inspection fit all three criteria above. First, it was not possible to include hospitals as mere control clusters; all hospitals from where we collected data had to be included in the inspections. The sheer cost of on-site, manual assessment of thousands of health records made it impossible for us to do this work on our own. We therefore relied on the inspection teams to collect and analyze data from the health records of individual patients. The inspection teams collected these data to use as evidence for their inspection and we "inherited" the data for use in our research project.

Collecting data for the research project only would not be possible for the inspection team. The inspection authorities are expected to follow up instances where patients have not received care that meets the legal requirements. This is a specific ethical obligation,[22] the effect of which is that the inspection authority cannot start investigations without the recourse to formal reactions. If the inspection authorities were to collect data merely to produce a control group for the inspection and they ended up addressing findings of nonconformity to the hospital, this would be tantamount to inspecting, i.e. treating, the control group.

[22] The framing of this ethical obligation with its emphasis on legality is due to the importance of rights-based legislation of healthcare in Norway (see 1.1).

Second, it was also impossible to inspect all hospitals simultaneously. National inspections are spaced out in time because of the practical difficulties in scheduling a specific time when all inspections should start. In the sepsis inspection, inspectors in "late" inspections seconded with those teams that performed the inspections earlier. In addition, some inspectors and medical experts participated in more than one inspection team.

Third, the intervention cannot be designed as a two-way crossover trial where half of the hospitals start in a "treated" state and then switches to "untreated" for the post-measurements. It is obviously impossible to reverse the effects of an inspection. The lessons learned and actions taken after an inspection will not be wiped from memory, nor should they. On the basis of these arguments, it was therefore decided that the stepped wedge was the best available study design for assessing the effects of the inspection.

4.3 Collecting and generating data

The plan for performing the inspection and collecting the data was conceived in a collaboration between the research project and NBHS.

4.3.1 Shaping the wedge

Ensuring a wide geographical inclusion in the sepsis inspection was an important consideration for NBHS. Based on counties and health authority areas, they first designated six geographic regions: Northern Norway, Central Norway, Western Norway, Southern Norway, Eastern Norway, and the capital region. The inspections were rolled out to 24 hospitals: four hospitals for each region. All hospitals were either government owned or government funded, and all had responsibility for emergency care within their given geographical area. The County Governors chose which hospitals to inspect, using size as the main inclusion criterion. The inspected hospitals thus included all university hospitals and all regional hospitals, as well as a selection of smaller local hospitals.

For each region, an inspection team was put together, consisting of a team leader and two or three other inspectors from the County Governors' offices. The representatives from the County Governors had formal training and background either from medicine (primarily as physicians or nurses) or from law. Each team also included a senior consultant physician with specialized knowledge in diagnosing and treating patients with sepsis. The senior physicians were handpicked experts whose regular work were in a hospital outside of the region the team operated within. For 15 of the inspections, one or two representatives from other County Governors' offices were seconded to the inspection team as observers.

The period from April 2016 until March 2017 was divided into six time periods of around two months each. These time periods were the steps of the stepped wedge. The regional teams were randomized to be assigned to one step each, within which they were to conduct inspections at four hospitals within the region.

4.3.2 Collection of quantitative data

The sampling was conducted sequentially rather than randomly. For each inspection, samples of health records were collected from lists of the most recent 250 eligible patients up to a specific date, referred to as P0, P1, P2, and P3. P0 included patients up until a common cut-off date, October 1, 2015. The other three lists used cut-off dates relative to the date of the inspection for each hospital. P1 had the inspection start as its cut-off date, while P2 and P3 had the dates of the two follow-up visits, respectively 8 and 14 months after the initial inspection.

To identify eligible patients, the Norwegian Patient Registry was searched for patients discharged with an International Classification of Diseases, Tenth Revision (ICD-10) diagnostic code commonly used to classify sepsis and infections (appendix 1 to study 1). The health records for the corresponding admissions were accessed by the inspection teams. Eligibility was determined using information in the health records about the clinical status of the patients at the time of admission. Following the established definition that was used at the time of the development of the protocol

(Levy et al., 2003),[23] we used as criteria suspected infection together with two out of four systemic inflammatory response syndrome signs (SIRS) (see 1.2.1). We did not include elevated leukocyte counts, because the results of the blood works often would not be ready when the physicians performed their initial diagnose.

The inspection team enrolled patient records into the primary journal review until they reached a goal of 33 patients per sample. Later, NBHS researchers would draw a secondary sample of 50 additional patients from the same list for an extra journal review. The primary sample would be used to assess differences in process measures from before to after inspection, while the primary and secondary samples would be used together to assess patient outcomes. The number of electronic health records to be included was arrived upon using a program developed by Hemming and Girling (2014) for power calculation in stepped wedge designs. According to the calculations, the primary sample of 33 patients per cluster would enable detection of a 15 percentage points difference in completion of a dichotomous process measure (from a starting point of 50%). Combined, the primary and secondary samples would detect a 7 percentage points difference in mortality (from a starting point of 15%).[24]

Within a given time period, the number of patients with sepsis varies between hospitals and regions, primarily because of differences in hospital size. Hence, the uptake time for each date–hospital dyad would vary.[25] Figure 2 shows an overview of the P0 to P4 uptake times for the six regions.

---

[23] Since the early 1990s, there has been three major international conferences, each one spawning a new consensus definition of sepsis. The definitions are reported in Bone et al. (1992), Levy et al. (2003), and Rhodes et al. (2017), and are colloquially referred to as *sepsis 1, sepsis 2, and sepsis 3*, respectively. The present study was planned before the sepsis 3 definition was accepted and disseminated. Changing sepsis definitions is by no means a novelty. The term *sepsis* comes to us from the ancient Greek [σηψις], which referred to some form of dangerous putrefaction. Sepsis was the opposite of but related to *pepsis,* which was a helpful process of maturation or fermentation. See Aristotle's *Generation of Animals* (trans. 1910) and Majno (1991).

[24] The calculations were based on the planned study design pattern with four data collection time points per hospital and specified as an incomplete stepped wedge design with binomial comparison, a significance level of .05, a power of .8, and an estimated intra cluster correlation of .05. Further details are available in the study protocol (Hovlid et al., 2017).

[25] In some lists, records of patients with admission dates extending beyond the cutoff date had been included. If these records had been enrolled into the record review, we tried to reclassify the patient to the appropriate time span.

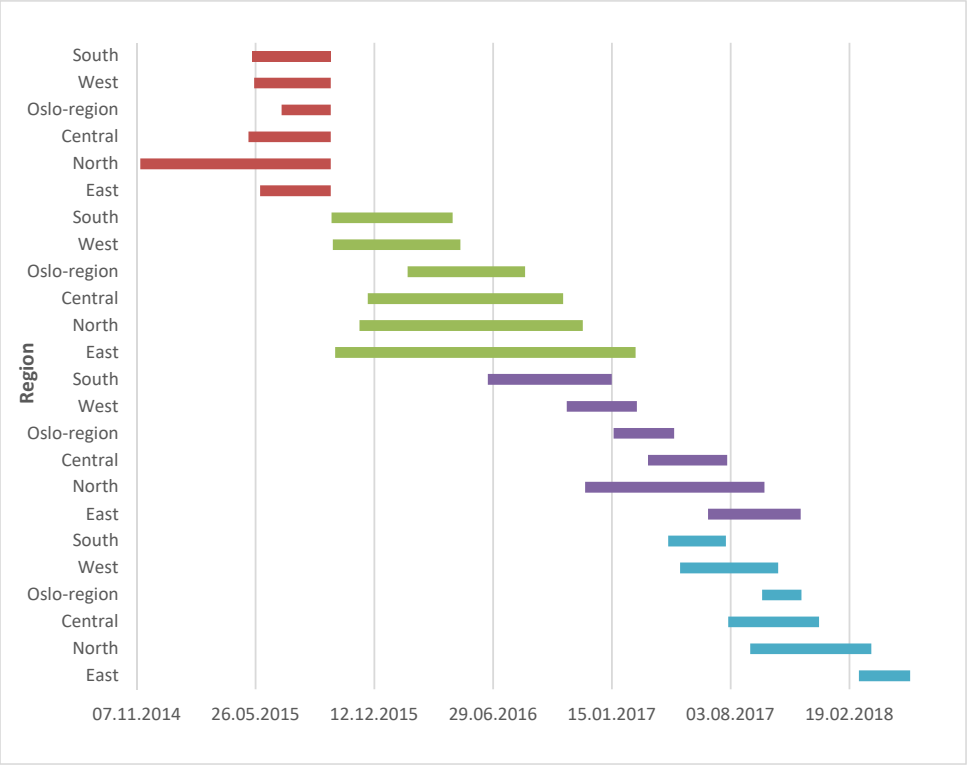


Figure 2 Uptake times by regions.Each bar represents the duration from first patient admission to last patient admission of a date–region dyad. The fill color of the bars indicate the time point of each sample: p0 (red), p1 (green), p2 (purple), and p3 (blue).

Both the primary samples collected by the inspection teams and the secondary samples collected by the NBHS researchers included information about admission and discharge dates, national identification number, and information about presence of organ dysfunction. The inspection teams recorded information on forms that later were digitized via *Epi Info* (CDC). The NBHS researchers recorded the information directly into Epi Info.

Though parts of the data extraction perchance could have been automated, NBHS opted for manual retrieval. The reason for this is that the data were retrieved from 24 different hospitals, each with their own technical implementation of the nationally mandated electronic health record system. Automating the data retrieval would pose technical

difficulties, and the time saved by automatic retrieval could potentially be outweighed by the time needed to design and verify such a retrieval.

Extracting data from the patient records called for great fastidiousness, as the inspection teams would use the sampled records to document any nonconformities. Assessments of individual health records could be contested by the emergency department management or staff. The inspected hospitals had access to the data collected by the inspectors, and they were able to verify the data if they wished to do so. If a nonconformity had been shown to be unsubstantiated, the authority and legitimacy of the inspection teams could be undermined. It was therefore in the best interest of the inspection teams not to be careless in their retrieval of data.

Data on comorbidity and 30-day mortality were provided from the National Patient Registry, based on the patients' national identity number and date of admittance.

I combined data from the EHR reviews and the National Patient Registry using *Microsoft Access*. I subsequently assessed the data quality in *Stata*, ensuring that they were correctly formatted and within valid ranges. Duplicates were deleted, keeping the first admission of patients who had been admitted more than once. I then recoded the data for use in the analyses and created variables for year of admission and seasonality. For the data used in study 3, I developed a Stata code for extracting age and sex using the national identity number, and I calculated length of stay using admission and discharge dates.

4.3.3 Focus-group interviews

When selecting which hospitals to conduct focus group interviews in, we followed a logic of quota sampling (Patton, 2015). We chose one hospital for each of the six regions where the inspections have been performed, so that hospitals across the whole of Norway were represented. For each hospital we ran two focus groups:

- employees working within the emergency department who are involved in the diagnosing and treating of patients with sepsis

- hospital leaders/managers who are responsible for the emergency department or other units involved in treating and diagnosing patients with sepsis

The groups were recruited via a contact person from their hospital – typically the leader in charge of quality management at the hospital level. The contact person would schedule the interview with the leaders who had responsibility either directly for the emergency department or for services involved in the tasks of diagnosing and treating patients with sepsis in the emergency department. Usually, the contact person would then recruit nurses and doctors (including residents) working in the emergency departments for the clinician focus groups on the basis of suggestions from the managers. The number of participants in the clinician and manager focus groups varied from interview to interview. In some cases, some participants had to cancel on short notice in order to tend to other pressing tasks.

Along with the interviews at the hospitals, we also performed interviews with all six inspection teams that have carried out the sepsis inspection. For these interviews, we aimed at convening the whole inspection team, though scheduling conflicts sometimes prevented one or two of the team members from participating.

For each hospital/inspection team, we conducted two focus-group sessions. One interview immediately before the first inspection at the hospital, and one from 8 to 12 months later. We aimed at including those who had participated in the first session to join the second, but due to scheduling and turnover this often proved difficult.

The task of interviewing was delegated to three interviewers, including myself. For 24 of the 36 interviews, two interviewers were present. I participated in 15 of these interviews. For the remaining 12 interviews, I was the only interviewer. We applied a semi-structured and open-ended interview guide, specifically adapted for each of the three groups. In developing the interview guides, we were informed by a theoretical framework of mechanisms of change in organizations subject to external assessment (see chapter 2). The focus-group sessions varied both in terms of the topics that were covered and in terms of the length (from 50 to 120 minutes).

A comment is warranted here about the difference between face-to-face and virtual interviews. Three interviews were conducted through group conference calls. Virtual interviews have their advantages. They provide access to geographically isolated and dispersed participant groups, and they can be cost-effective (Braun & Clarke, 2013). This was the reason why we chose to use group conference calls. For one interview it was not possible to convene the whole focus group to a physical meeting. The two other interviews would require arranging two separate trips, each taking at least two full days, on a short notice. However, teleconferencing raised a new set of challenges to the interview situation. It required more intentive listening in the absence of body language and visual clues. Also, the sound quality wound up to be poor in two of three interviews, mainly due to technical difficulties, occasionally making it difficult to distinguish between the different speakers. This made it harder to nudge the inactive participants to join the conversation. Working over a conference line, you also need to be careful when uttering monosyllabic hums that are intended to show the speaker that you are in fact following along. Such utterances can be construed as an interjection or a question that has been clipped due to signal loss. Consequently, the speaker, to a degree, loses the comfort of confirmation from the interviewer/facilitator. Comparing the transcripts from the teleconference interviews with those of the live interviews, I found the interviewee contribution to the dialogue to be shorter in the teleconferences.

4.4 Quantitative instruments

The quantitative data were primarily analyzed in studies 1 and 2, but we also used the quantitative data to aid in the case selection for study 3. Table 2 presents the variables used in the three studies related to admission information, patient background, and patient health status.

Table 2 Patient background information and health status

		Study		
Variable	**Data source***	**1**	**2**	**3**
Hospital	EHR	x	x	(x)
Admission date (calendar date or year)	EHR	x	x	
Admission pre- or post-inspection	EHR		x	(x)
Age (in years)	EHR/NPR	x	x	
Sex (male/female)	EHR/NPR	x	x	
Presence of organ failure†	EHR	x	x	
Comorbidity‡	NPR	x	x	

* EHR = Electronic health records, NPR = National Patient Registry

† Organ failure was defined as fulfilling one of the following criteria at arrival to the emergency department: oxygen saturation <90% or PaO2/FiO2 <40 kPa, altered mental status, urine output <0.5 mL/kg/hour or increase in serum creatinine >50 micro mol/L, international normalized ratio >1.5 or activated partial thromboplastin time > 60 seconds, platelet count < 100 or 50% reduction in previous three days, serum bilirubin >70 mmol/L, serum lactate >4 mmol/L, blood pressure <90 systolic, mean arterial pressure <60, or fall in mean arterial pressure >40 mm Hg. This definition was based on guidelines from Norwegian Directorate of Health (2018) and international guidelines (e.g. Dellinger et al., 2013).

‡ Charlson comorbidity index (Charlson, Pompei, Ales, & MacKenzie, 1987)

The data on measures of care delivery were collected by the inspection teams (see 4.3.2 above). NBHS had prioritized which care processes to include based on international guidelines (e.g. Dellinger et al., 2013; National Institute for Health and Care Excellence, 2016) and advice from experts. The measures of care delivery were all coded as dichotomous variables, indicating whether the process had been completed (or completed within the prescribed time limit) or not. Table 3 gives an overview of the measures of care delivery used in the studies.

Table 3 Measures of care delivery

Variable	Data source*	Study 1	Study 2	Study 3
Triage within 15 minutes of arrival	EHR	x	x	
Patient assessed by a physician in accordance with the urgency specified in the initial triage	EHR	x	x	
Measurement of vital signs within 1 hour of arrival.	EHR	x	x	
Measurement of blood lactate within 1 hour of arrival.	EHR	x	x	
Blood samples† taken within 1 hour of arrival.	EHR	x	x	
Blood cultures taken before administration of antibiotics.	EHR	x	x	
Adequate supplementary investigations to detect the focus of infection.	EHR	x	x	
Adequately observation‡ while in the emergency department.	EHR	x	x	
Antibiotics administered within 1 hour.	EHR	x	x	(x)
Antibiotics administered within 2 hours.	EHR	x		
Antibiotics administered within 4 hours.	EHR	x		
Intravenous fluids administered within 1 hour.	EHR		x	
Oxygen therapy within 30 minutes§	EHR			

* EHR = Electronic health records

† Leukocyte count, haemoglobin, C-reactive protein, creatinine, electrolytes, platelet count, glucose, bilirubin, blood lactate

‡ 'Adequate' is defined as continual observation and measurement and documentation of vital signs at least every 15 minutes in critically ill patients with sepsis and organ failure, measurement and documentation of vital signs every 15 minutes if a physician has not examined a patient with sepsis but no documented organ failure, and every 30 minutes after first examination in such patients unless the physician decides otherwise.

§ Oxygen therapy was originally planned as one of the treatment processes that would be evaluated in the study (Hovlid et al., 2017). During the inspections, however, this process was for the most part excluded due to conflicting research evidence regarding its treatment effectiveness. It was included only in a few inspection reports, and we chose to leave it out of the studies.

In study 1 we used 30-day mortality as the patient outcome measure. In study 2, we also added length of stay in days (see Table 4).

Table 4 Patient outcome measures

		Paper		
Variable	**Data source***	**1**	**2**	**3**
All-cause mortality within 30 days of admission	NPR	x	x	
Length of stay (in days)	EHR		x	

* EHR = Electronic health records, NPR = National Patient Registry

4.5 Statistical analyses

4.5.1 Study 1

We performed three rounds of statistical analyses in study 1. First, we used descriptive statistics to describe the patient sample and to analyze the proportion of patients having documented receiving care within recommended time limits. Second, we analyzed the association between non-completion or delay of four diagnostic procedures (triage, examination by physician, lactate measurement, and observation regimen) and prolonged time to administration of antibiotics using regression models. Third, we analyzed the association between time to antibiotics and mortality, both by obtaining observed mortality across hourly intervals of time to antibiotics, and by estimating a logistic model regressing 30-day mortality on a polynomial function of time to antibiotics in minutes.

Data were missing from several of the variables in the data set. Based on our knowledge of the procedures for data collection and verification employed by the inspection teams, we were satisfied that the risk of missing values due to negligence or faulty retrieval on the part of the inspection team was minimal. Hence, the cause of missing data would probably be that the information was not recorded in the first place.

For the descriptive statistics, we reported the proportions of patients having documented care that conformed to recommended standards. This proportion therefore included missing values in the denominator. For the regression analyses, missing values posed a challenge. If we had not corrected these missing data, a sizable fraction of the sample would have been excluded from the analyses due to listwise deletion.

And, furthermore, we could have risked introducing bias into the model because the probability of missing data on a particular variable might be related to the value of that variable or to the values of other variables in the data set.

We assumed the data to be *missing at random* (MAR), i.e. that "the conditional probability of missing data on Y, given both Y and X, is equal to the probability of missing data on Y given X alone" (Allison, 2001, p. 4). This assumption implies that, for example, we allow for there to be varying numbers of missing observations of time to antibiotics between those who have organ failure and those who do not have organ failure. Within each of those two categories, however, the probability for missing data on time to antibiotics has to be unrelated to the value of time to triage.

While there is no way to statistically check for the viability of the MAR assumption, one can first of all reason whether the value of a variable is likely to affect the chance that said variable is missing information. Moreover, MAR assumes missingness to be independent on unobserved values after adjusting for study/control variables. Adding more study/control variables will therefore reduce residual dependence of missingness, and hence make the MAR assumption more tenable (Allison, 2009).

Our assumption that the data was MAR was based on a careful consideration of the possible reasons for missingness, along with thorough studies of patterns of missingness in our material. As we had information on time to antibiotics in two different measures, minutes and hours, we knew that the propensity for missing data on time to antibiotics in minutes was dependent on the value of the variable itself. That is – the longer the time to antibiotics (measured in hours), the smaller the proportion of patients who had recorded time to antibiotics in minutes. Furthermore, we knew that missingness on the three variables suggested for imputation were related to missingness on other variables, but this is allowed for (Allison, 2001). What we cannot show statistically, is whether the probability for missing data depended on unobserved quantities. This would be the case for example if the probability of missing data on time to triage varied according to time to triage even after controlling for the other variables included in the model. Granting that we cannot be certain, we believed that

there were no such significant unobserved effects, and we chose to consider our data to be MAR.

When data are MAR, regression analysis with listwise deletion might lead to biased estimates. For example, if time to antibiotics is dependent on time to triage, but our observations systematically misses data on time to triage for those with delayed triage, then regressing time to antibiotics on time to triage will yield biased estimates. Therefore, and because listwise deletion also results in a radically reduced number of observations included in the regression analysis, we decide to use a combination of recoding and multiple imputation in study 1.

We recoded missing values as not completed for those diagnostic variables where data would have been recorded automatically if the measurement had been taken. For the other variables, we used multiple imputation with the "multiple imputation for chained equation" (MICE) approach, in which each variable to be imputed is fitted with a separate regression model (Allison, 2009). This procedure is explained in detail in appendix 2 to study 1. Prior to fitting the multiple imputation models, we assessed residual distribution, multicollinearity, and influential cases according to standard assumptions of regression models (Vittinghoff, Glidden, Shiboski, & McCulloch, 2012), to rule out the possibility of introducing bias into the imputated data sets.

For the analyses of the association between diagnostic procedures and time to treatment, we used time to antibiotics in minutes as an outcome variable, fitting the model using Ordinary Least Square regression within Stata's *mi estimation*, which estimates model parameters from the imputated data sets. We obtained cluster standard errors to account for unchecked within-cluster error correlation and resulting heteroskedasticity across hospitals (Cameron & Miller, 2015). We reported associations between time to antibiotics and each diagnostic variable, first as crude associations, and then adjusted for age, organ failure, comorbidity, and time to admission. Finally, we included all diagnostic variables and adjustment variables into one single model, to obtain a prediction of delay to antibiotic treatment if all four diagnostic procedures were delayed or non-completed.

The final model for the predicted time to antibiotic treatment for patient *i* was thus:

$$y_i = \beta_0 + \beta_1 Triage15m_I + \beta_2 Physic_i + \beta_3 Observ_i + \beta_4 Lactate_i + \beta_5 Organ_i + \beta_6 Charlson_i + \beta_7 Age_i + \beta_8 Date_i + \beta_9 Date_i^2 + \beta_{10} Date_i^3 + \epsilon_i$$

The predictors corresponding to the regression coefficients $\beta_1 - \beta_4$ are dummy coded so that 1 represents an incomplete care process, i.e. failing to complete triage, examination by physician, adequate observation, and measurement of lactate as recommended. The predictors for organ dysfunction and Charlson comorbidity index were also dummy coded, while the predictor for age was coded as a discrete variable counted in years. The date predictors were entered as a third order polynomial expression to account for changes over time. The first order date was number of days from November 13, 2014, the first admission in our material.

We then estimated a model for the association between time to antibiotics and mortality, again with the *mi estimate* command but this time using logistic regression. Time to antibiotics was entered as a third order polynomial expression, and patient's age, year of admission, comorbidity, and presence of organ failure were included as dummy coded adjustment variables. The model estimating 30-day mortality can be written:

$$log\left[\frac{P}{1-P}\right] = \beta_0 + \beta_1 Min + \beta_2 Min^2 + \beta_3 Min^3 + \beta_4 Age + \beta_5 Organ + \beta_6 Charlson + \beta_7 Y2015 + \beta_7 Y2016 + \beta_7 Y2017$$

Using the Stata command *mimrgns*, we obtained predictive probabilities of mortality at 15 minutes intervals. We also obtained predictive logit estimates of confidence intervals at the same intervals, which we re-calculated into confidence intervals of probability before plotting the estimates as a curvilinear function along with observed mortality per hour in a Stata *twoway* graph.

We then performed sensitivity analyses of the estimated models using different subsets of the sample (see appendix 3 to study 1).

4.5.2 Study 2

The analyses in study 2 compared care processes and patient outcomes before and after the inspection, using crude and adjusted regression analyses for analyses of changes in care processes and adjusted analyses of changes in patient outcomes.

As explained in section 4.4, study 2 used many of the same variables as study 1. However, study 2 used a larger dataset. Whereas study 1 was based on the primary samples collected by the inspection teams at the first site visit, with additional information from the Norwegian patient registry, study 2 included additional data from the two record reviews performed 8 and 14 months after the inspection. For the analyses of patient outcomes, we also included the secondary samples collected by NBHS. Unlike in study 1, we did not impute data. As compared to study 1, we had more observations, which lessened the problem of missing data. Additionally, the imputations in study 1 were only done for observations for which we had information on time to antibiotics. Thus, we had a reliable auxiliary variable available for the imputation models for our dependent variable time to antibiotics. In study 2 we included the observations that lacked information on time to antibiotics, and we also estimated models where the other care processes (in addition to time to antibiotics) were dependent variables. Multiple imputation on dependent variables risks adding noise to the estimates in the absence of an auxiliary variable in the imputation model that correlates strongly with the imputation variable (White, Royston, & Wood, 2011).

Both study 1 and study 2 included hierarchical data with repeated measurements within clusters, i.e. two or more record reviews within the same hospital. With such data, outcomes are often correlated across observations. For traditional models, this violates the assumption of independence between observations, and can yield biased parameter estimates and underestimated standard errors (Guo & Zhao, 2000). It is therefore necessary to include correlation structures into the analysis. As explained in 4.5.1, we incorporated this into the analyses in study 1 by using robust cluster standard errors. For the analyses in study 2, we used mixed-effects (ME) models. In two-level analysis, ME splits the residuals into a random component specific for each cluster and a random component for each individual observation (Rabe-Hesketh & Skrondal, 2008).

Originally, we intended to use population averaged generalized estimating equations (GEE) (Hovlid et al., 2017). One of the reasons for why we instead chose ME was that we did not use multiple imputation in study 2. ME, which fits the model using maximum likelihood (ML), is less affected by missing data than GEE (Vittinghoff et al., 2012). Another reason for choosing ME over GEE was that we had difficulties with specifying the correlation structure when we tried to fit the model using GEE, the robust correlation structure yielding large standard errors for one of the fitted models.

As a template for fitting the models, we used a standard model-based approach for estimating the outcome Y for an individual *l* from cluster *i* at time *j* that is commonly used in stepped wedge analyses (Hemming, Taljaard, & Forbes, 2017; Hussey & Hughes, 2007):

$$Y_{ijl} = \beta_0 + \beta_j + \theta X_{ij} + u_i + e_{ijl}$$

where β_0 is the intercept term, β_j is a fixed time effect, θX_{ij} is the treatment effect, u_i is a random effect for cluster, and e_{ijl} is an individual random effect. For the care process outcomes and for 30-day mortality, the model was fitted with a logit link function and effects were obtained as odd ratios. For the analysis of effects on length of stay the model was fitted with a log link and a negative binomial distribution.

In the crude analyses the model was fitted without the time effect (β_j). In the adjusted analyses the model was extended with the time effect and additional control variables (patient age, presence of organ dysfunction, and sex), following the procedure explained in appendix 3 to study 2.

4.6 Focus-group analyses

The design of study 3 was inspired by the mixed-methods approach called “nested analysis” (Lieberman, 2005). This approach combines large-N analysis (LNA) with small-N analysis (SNA), allowing for the exploration of both general associations and individual cases. Starting with a preliminary LNA, one goes on selecting a few cases

based on the outcome of the LNA for in-depth exploration. For study 3, we opted to focus on hospitals that had experienced an improvement following the inspections. Having completed the initial analyses in study 2 prior to starting the analyses for study 3, we knew the rate of improvement at each of the six hospitals included in the interview material. We used this knowledge when selecting which focus groups to include in our analysis. As we were primarily interested in how the inspections could foster improvement through an organizational or relational impact (see chapter 2), we chose to include only the interviews from the second round of interviews.

Our chosen method of data analysis was thematic analysis. In thematic analysis themes from qualitative data are identified and highlighted in order to gain understanding of the phenomena that is being researched (King & Brooks, 2018). Thematic analysis is well-suited to theoretically informed research (Braun & Clarke, 2006). This was an important reason for why we chose this approach. The theoretical framework was important to achieve interpretive power and understanding of the material.

We used Nvivo, a qualitative data analysis software, for organizing and analyzing the material. At the first day of interviews, we developed a list of initial codes to be included as nodes in Nvivo. When new interviews were completed and transcribed, I imported the transcripts and audio files from the interviews into Nvivo and performed an initial coding of the interviews to these nodes. This list of codes was revised several times and expanded with new codes, as our understanding of the data developed. Thematic analysis involves moving from the initial codes into an iterative process of reviewing and revising candidate themes (Braun & Clarke, 2013). When analyzing the data in study 3, I grouped the codes into themes and developed graphical mind maps as an aid in the analytical process. Concurrently with my analysis in study 3, a coauthor of the study independently read and analyzed the included transcripts. We then collaborated on providing a draft of the analysis along with the transcripts to the rest of the participating researchers.

5. Results

In this chapter I present the main results of the three studies, giving to each a short introduction and contextualization, and a description of its role in relation to the research project as a whole.

5.1 Study 1

5.1.1 Preliminaries and contextualization

This study was performed to document the performance metrics of the hospitals' patient care using the data gathered through the sepsis inspection. There were two main objectives for this study: (1) To assess the timeliness of diagnostic procedures for recognizing sepsis in emergency departments; and (2) to estimate associations between diagnostic procedures and time to antibiotic treatment, and to estimate associations between time to antibiotic treatment and mortality

We sought to investigate some specific themes that had not been fully addressed by previous research within this field. While it had been documented, as described in chapter 1, that screening and diagnostics were not consistently carried out in emergency departments, resulting in delayed identification of sepsis in patients (Goodwin et al., 2015), there was a need to provide more robust data documenting the extent to which diagnostics are not carried out. There was also an added value of providing such data for the Norwegian emergency care context. Additionally, there were no studies assessing the association between timeliness of diagnostic procedures and time to treatment. Hence, we wanted to assess the timeliness of care for patients with sepsis in emergency departments. Furthermore, we wanted to investigate the associations between timeliness of diagnostic procedures and treatment and between treatment and mortality.

In the context of the research project and this thesis, the study was a necessary stepping-stone, establishing a baseline from which we could evaluate the effects of the

inspections. It also provided insight into how well the chosen measures of care delivery reflected the clinical work performed within the emergency departments.

5.1.2 Results

Using data collected at the time of the inspection visit, the study included 1559 patients. When evaluating whether their care was given in line with pre-defined standards, we found that the proportion of patients who had documented triage within 15 minutes was 73%. For more than 80% of the patients, complete assessment of vital signs, blood culture, and blood samples (except for bilirubin and lactate) had been documented as complete within one hour of admission. Less than half of the patients had documentation for being examined by a physician (45%), and roughly the same percentage of patients (44%) had been adequately observed while in the emergency department according to the degree of priority assigned during triage. Treatment with antibiotics was completed within one hour for 25% and within two hours for 56% of the patients.

We also calculated the association between non-completion of diagnostic processes and time to treatment with antibiotics. We focused our analysis on four key procedures: triage, examination by physician, lactate measurement and observation regiment. For almost 50% of the sample at least two of these procedures were delayed or not completed. In an adjusted analysis where non-completion or delay of these procedures were included along with background variables (organ failure, age, comorbidity, and date of admission), we found a predicted delay of 159 minutes to first dose of antibiotics if all these four key diagnostic procedures were lacking.

The 30-day all-cause mortality was 9.9% (95% confidence interval [CI]: 8.4-11.4) for the entire study sample and 17.4% (95% CI: 14.2-20.5) for patients with documented organ failure. Analyzing the association between time to antibiotic treatment and mortality, we found a substantial variation in mortality according to time to antibiotics. Patients who started antibiotic treatment between 2 and 3 hours after admission had lower mortality than those who started antibiotics earlier or later.

5.1.3 Key findings relating to the aim of the thesis

This study documented that the hospitals had not managed to provide the recommended care for all patients with sepsis presenting to their emergency departments. Furthermore, the study documented associations between the diagnostic processes and treatment processes that the inspection assessed. These two findings are pursued further in the discussions in chapter 6 regarding the validity of the chosen indicators and the effects of the inspections.

5.2 Study 2

5.2.1 Preliminaries and contextualization

This study focused on the effects of the inspection on key care processes and patient outcomes. The two main objectives of this study were: (1) To evaluate the effects of external inspections on hospitals' clinical processes for detecting and treating sepsis, and (2) to evaluate the effects of external inspections on length of hospital stay and 30-day mortality.

At the time this study was planned, the existing research on the effects of inspections on patient care and patient outcomes had not come to any certain conclusions. According to the then most recent systematic review (Flodgren et al., 2016), there had been few rigorous quantitative studies conducted regarding effects of inspections. As explained in chapter 2, some additional studies of inspection effects have been published in the time since this study was planned. None of these studies have strengthened the case for inspections having long-lasting effects on patient care and patient outcomes. Therefore it has been central to this thesis to find out if we actually can link the inspections to quality improvements, or if inspections merely provide a form of "mindless policing" that has an adverse effect on quality improvement, as one critic has suggested (Brennan, 1998).

5.2.2 Results

Comparing patients in the pre-inspection and patients in the post-inspection group in adjusted analyses, we found statistically significant improvements for four of the process measures: being examined by a physician within the time frame set in triage (OR = 1.28, 95% confidence interval [CI]: 1.07-1.53), having complete set of vital measurements taken within one hour (OR = 1.78, 95% CI: 1.10-2.87), having lactate measured within one hour (OR = 2.75, 95% CI: 1.83-4.15), having an adequate observation regimen in the emergency department (OR = 2.20, 95% CI: 1.51-3.20), and having had antibiotics administered within one hour (OR = 2.16, 95% CI: 1.83-2.55).

In the negative binomial regression analysis where we adjusted for confounding variables but not time, the incidence rate for length of stay after the inspections was .87 (95% CI: .84-.90) times the incidence rate for length of stay before the inspections. If we obtain predictive margins using mean values for the other covariates in the model, this model predicts a drop in predicted length of stay from 7.0 (95% CI: 6.6-7.4) to 6.1 (95% CI: 5.8-6.4) days from before inspections to after inspections. Analyzing the effect of inspections on 30-day mortality, the adjusted model where we did not control for time showed that the odds for mortality in the post-inspection group was .81 (95% CI: 0.69-0.95) times the odds for mortality in the pre-inspection group. This translates to predicted probabilities of 30-day mortality of .074 (95% CI: .063-.085) in the pre-inspection patient group and .06 (95% CI: .051-.070) in the post-inspection group. In analyses where we included calendar year as a control variable, there were, however, no statistically significant effects for either of these outcomes.

5.2.3 Key findings relating to the aim of the thesis

Study 2 is in some ways the centerpiece of the present thesis. Its findings regarding the effects of the inspection on changes in care processes and patient outcomes are discussed at length in chapter 6, with regard to both the impact of the inspection and the methodological difficulties involved in measuring such effects.

5.3 Study 3

5.3.1 Preliminaries and contextualization

With study 3, we wanted to use qualitative data from the focus group interviews to dig deeper into how inspections might influence quality improvement. The objective was to study how external inspections may foster clinical improvement in hospitals.

When we started analyzing the interview data for study 3, we had already completed the analyses for study 2. We therefore knew that there were reasons to believe that the inspection had had a positive effect on the quality improvement work in the emergency departments. As described in 4.6 above, we used the quantitative data for selecting which cases should be included in study 3. Prior studies have pointed to different mechanisms involved in inspections and other regulatory external assessment schemes that could be explored further. This literature, much of which was included in a recently published narrative synthesis by Hovlid, Braut, et al. (2020), influenced our approach to the interviews.

5.3.2 Results

We identified three themes that became focal points of our analyses. First, we identified an increasing awareness about the need to improve the quality of care by providing data on clinical performance. To a degree, hospitals lacked systems and routines for monitoring and assessing completion of recommended care processes. The inspection teams reviewed and summed up data on performance in an easily digestible format. This provided a useful framing of challenges experienced in the day-to-day work in the emergency departments.

Next, we found that the inspections could help build acceptance for improvement through professional credibility and focus on clinical practice. The inspection teams included skilled experts within the field, and the hospitals felt reassured that the teams were capable of understanding both the medical issues of sepsis care and the practical

functioning of an emergency department. This practical understanding enabled the inspection teams to focus on how the care processes worked as a whole.

Finally, we found that the inspections could help fostering leadership commitment. The role of management was seen as crucial. It was perceived as very important that clinical managers showed a genuine interest in improving the processes. One of their functions was to open up channels of communication between staff and hospital management and between different groups of professionals.

5.3.3 Key findings relating to the aim of the thesis

The explorative nature of this study presented an opportunity to proffer different theoretical viewpoints and interpretations to the findings from study 2. These are brought to the fore in chapter 6. The three themes that we identified in study 3 serve as points of departure for the broader discussion of how we best can understand the role of inspection in the context of theory and previous research on this subject.

6. Discussion

The themes of this thesis have been the effects of external inspections on improvement in healthcare organizations and the mechanisms that link the two. When reviewing what our research contributes to the understanding of these themes, we should not put the sepsis inspection on a pedestal. The findings suggest that in some areas, there were no improvement effects directly attributable to the inspection. However, it was far from an abject failure: There were significant improvements in important areas following the inspection, and we identified several themes that could link the inspection to quality improvement efforts in the hospitals.

In this chapter I pull together the findings from the three studies and discuss what they can and cannot tell us about inspections and quality improvement. In the first part of the chapter, the methodological and ethical considerations, I review the strengths and weaknesses of the study design and methods; it is here that the limitations of the choices regarding data collection and analyses are laid bare. The rest of the chapter deals with answering the three main research questions that I posed in chapter 3. The discussion builds on the theoretical foundation presented in chapter 2 and the findings presented in chapter 5, but I also try to bring in other perspectives from recent research in order to provide a broader interpretation of what the studies add to the field.

6.1 Methodological and ethical conciderations

In previous chapters I have described that from a bird's eye view, the thesis can be considered a case study of statutory inspections. When discussing methodology, it is therefore natural to return to how we are to understand the sepsis inspection as a case; that is, to understand it in relation to the phenomenon I have defined it as a case of.

As noted in chapter 3, the sepsis inspection can be considered as typical of statutory inspections in Norway. While the majority of inspection reports of healthcare services are from primary care services, the County Governors perform many inspections of

specialized healthcare each year.[26] Moreover, the inspections follow the same established processes and guidelines, regardless of the theme of the inspection. One advantage of choosing a case that involves hospital emergency care, is that it is not an uncommon field in international research into external assessment of health care organizations. It has been a recurring theme in accreditation research (Hinchcliff et al., 2012), and, though the number of accreditation studies dwarfs the number of studies of statutory inspections, hospital emergency care has also been a theme some previous inspection studies (Allen et al., 2019; Castro, 2018; Smithson et al., 2018).

Not only is the inspection clearly within the bounds of the phenomenon we are investigating, i.e. it is a "typical" case of an inspection; it is also a case that is well suited to the purposes of the thesis. Our goal has been to understand how the use of measurements and the inspection teams' follow-up of the inspected hospitals influenced the quality improvement work. The inspection was designed with an eye to include a comprehensive set of clinical data in the assessment process, and to provide the hospitals with actionable evidence for improvement. This makes the sepsis inspection an information rich case for our research aims.

Having hospital emergency care as the area for research, rather than for instance family medicine or mental health services, has made it easier to include a quantitative strand in the study. Compared with an area such as family medicine, where relational work between physician and patient is central, where a multitude of non-medical issues can enter the picture along with the specific health problem, and where the physician-patient relation can go on for decades without any discernible "clinical pathway", sepsis care in emergency departments is a fairly straightforward process: It follows established guidelines, aspects of the patients' health and care processes are recorded in a way that is relatively easy to measure quantitatively, and length of stay is fairly

[26] For the years 2015 – 2019, NBHS has published 943 reports from inspections of health care services (Norwegian Board of Health Supervision, 2020b). Around one in four of these reports (246) are from specialized healthcare, and around half of the reports from specialized healthcare have been performed as a part of national inspection campaigns like the sepsis inspection.

short. This makes it possible to collect data sets of comparable information across several different organizations.

The stepped wedge design was chosen as it is a robust design to assess effects, as explained in chapter 4. This design has not been used to study the effects of statutory inspections in healthcare, and, as far as we know, only two previous studies have used this approach for assessing the effects of other types of external assessment in health care.[27] The stepped wedge design is sometimes portrayed as the next-best option after randomized controlled studies (Hovlid et al., 2017). In my opinion, this might do the stepped wedge design a disservice. We should think critically about experimental evidence. They give great control over what the intervention looks like, but at the same time experimental settings present more artificial scenarios, and it can be difficult to understand how those scenarios map onto real life. It is precisely here that the stepped wedge excels: Compared to randomized controlled studies, it is far less invasive on how the interventions work in practice. Yet it still retains much explanatory power by allowing for systematic comparisons of performance before and after the intervention while introducing an element of randomization into the mix.

As I noted in section 2.3.1, interventions that work in one context might not work in another. This project has aimed to measure and understand the effect of something that is context-dependent (inspections) on something that is both context-dependent and difficult to measure (clinical care and patient outcomes) (McWhinney, 1989). Walshe (2007) argues that critics who discount quality improvement efforts in health care on the grounds that there is weak experimental evidence for the effect of such interventions, are guilty of fundamentally misunderstanding the research methods within this domain. There is often great heterogeneity in the contexts, content, and implementations of such interventions. Says Walshe, “they can only be properly evaluated if their interconnected context, content, application and outcomes are understood” (p. 58).

[27] Stepped wedge design is used in two Danish studies of accreditation (Bogh, Falstie-Jensen, Hollnagel, Holst, Braithwaite, & Johnsen, 2017; Bogh, Falstie-Jensen, Hollnagel, Holst, Braithwaite, Raben, et al., 2017).

The issues I have raised here have to do with the internal and external validity of the studies. If we are to assess the findings of research into a complex subject, such as the sepsis inspection, it is not possible to rely only on statistical testing and generalization. What we can do, additionally, is to use theory, abstraction, and conceptual modeling as ways to extract out the significant components of the intervention and extend our inferences beyond the single case (Pawson, 2013). It is through interpreting and advancing the theoretical concepts that the study was based on that we reach new ways of understanding the phenomenon (Yin, 2018). This holds true for the sepsis inspection. The theoretical frameworks we have applied when interpreting the associations that were reported in study 2 have been important for our understanding of inspections as a phenomenon.

The mixed methods approach strengthens our interpretations. In chapter 4, I argued that the research design is made more credible due to its potential for triangulation on methods (Patton, 1999, 2015). The thesis comprises an observational quantitative study (study 1), a quasi-experimental study (study 2), and a qualitative focus-group study (study 3).[28] This provided opportunities for assessing the consistency of findings, thus contributing to the validation of the analyses (Patton, 1999). The strengths of the qualitative approach offset the weaknesses of the quantitative approach, and vice versa, and additionally triangulation can also enhance understanding and improve the usefulness of those findings (Creswell & Plano Clark, 2011). Embedding the qualitative strand into the design was especially important for strengthening validity, as it signaled a switch from an implicit successionist theory of causation to an investigation of substantial relations, which situated practices within contexts (Sayer, 2000). We cannot be content with counting the effects of inspections; we need to understand the effects of inspection, and to do so, we need insight into how the participants and practitioners understand the inspections.

[28] The third approach can in fact be considered a case study whose main approach is a holistic qualitative approach but additionally utilizes quantitative data analysis and document analysis (for the sample strategy). It could thus be considered a mixed-methods case study within a mixed-methods case study.

We did not exploit the full potential of analyzing and evaluating the sepsis inspection as a holistic case study with a mixed methods approach within a single study. None of the three published studies relied on an explicit mixed methods analysis and integration. Such an approach would involve merging or connecting the two strands in a systematic joint analysis (Creswell & Plano Clark, 2011). It is, however, important to note that we gathered and analyzed the qualitative and quantitative data simultaneously. Thus, we retained important advantages from the mixed methods approach, such as improved understanding of context and increased credibility through triangulation.

An obvious methodological challenge involved in mounting a project like the present comes from the extensive scope of data collection and analyses. Stepped wedge design can by their very nature become expensive and long-lasting (Kotz, Spigt, Arts, Crutzen, & Viechtbauer, 2012; Pawson, 2013). The inclusion of a qualitative strand of inquiry, in addition to the stepped wedge study, increased the workload and cost of the project, and required additional skills in conducting and analyzing interviews. In the present project we were able to utilize the data gathered by the inspection teams, and the requirements of the stepped wedge study did not lead to any protraction of the time it took to complete the inspection. Furthermore, the project was greatly helped by the inclusion of co-researchers with expertise on areas as diverse as regulation, sepsis diagnosis and care, and statistics.

Having discussed the overall research design, the next sections address some specific issues regarding the validity of observational data, confounding, interviewing, and research ethics.

6.1.1 The limits of observational patient data

As with any research relying on observational data, the quantitative analyses in the present studies could potentially be biased due to confounding. This threat to internal validity was especially precarious when it came to our estimates of the associations between time to antibiotic treatment and 30-day mortality. It is difficult to say for

certain what to expect of this association. A large observational study found a 4% increase in odds of in-hospital death for patients with sepsis per hour after arrival in the emergency departments (Seymour et al., 2017).[29] Hence, a natural assumption in our study would be that the predicted mortality was lowest for those who received antibiotics early, and that at one point, it would gradually increase with increasing delay to antibiotic treatment. Our results however indicated that there was significant confounding of unobserved heterogeneity in our study sample. The association was predicted as a parabolic, U-shaped curve, where mortality was predicted to be highest for those who received antibiotics earliest (see Fig. 2 in study 1). A plausible explanation for this association is that the patients who were worst off got earlier treatment than the rest of the patients, and furthermore that the variables we included as controls for such heterogeneity (age, organ dysfunction, and comorbidity) were not able to fully control for this effect. Hence, our results are of limited value in contributing to the debate of the exact pivotal timepoint for when antibiotic treatment should be started. As there is no reason to believe that antibiotic treatment was postponed for the most ailing patients, the upward trend in the right-hand side of this curve nonetheless indicates a tendency for an increasing mortality when antibiotic treatment is delayed up to four hours or more.

The caveats regarding observational data also apply to study 2. Chance variability in patient selection has probably had less impact, as the sample sizes were larger in study 2 than in study 1. However, there might be systematic differences between the patient groups before and after inspection. We know for instance that the average age of patients in the post-inspection was higher and that a larger proportion had organ failure, as compared to the pre-inspection group. While these differences are controlled for, we

[29] There is a scientific debate raging about how urgently antibiotic treatment should be given to patients with sepsis (more on this debate in section 6.2.1). Where exactly the pivotal point is for when mortality increases significantly was an open question. Prior to the study by Seymour et al. (2017), a review of eleven observational studies on the effect of time to antibiotics on mortality did not find a statistically significant increase in patient mortality for those receiving antibiotic treatment more than three hours after triage in the emergency department (Sterling, Miller, Pryor, Puskarich, & Jones, 2015). Howbeit, the previous studies were also observational studies; Therefore, they may be afflicted by the same types of biases and threats to internal validity as our study 1 was.

cannot rule out the possibility of unobserved differences between the before and after groups.

One potential source of bias that can pose problems in observational studies is that an increasing focus on the medical condition can lead to more mild cases being discovered. This might show itself as an apparent rise in prevalence and incidence, and at the same time outcomes appear to improve even if the underlying efficacy of the treatment does not (Fisher & Welch, 1999). Such effects are controlled for, as all patients were included using the same clinical eligibility criteria and the initial inclusion to be considered for eligibility was very wide. In addition to codes for sepsis, the list used to identify patients included several codes pertaining to infectious diseases.

6.1.2 Trials and tribulations with the stepped wedge

The central statistical design issue in study 2 was how to avoid the *post hoc ergo propter hoc* fallacy, i.e. concluding that changes occurring after the inspections by necessity must be caused by the inspections. There are four main ways in which we have dealt with this issue: (1) including potential confounders related to patient characteristics in the regression models, (2) assessing the effects of a patient safety program that ran concurrently with the sepsis inspection, (3) use of the stepped wedge design, and (4) including time as a covariate in the regression models. Potential confounders related to patient characteristics are discussed above. The patient safety program, which was a voluntary sepsis management improvement, was found to have a negligible impact (as explained in study 2). We will now look at the two remaining issues in turn.

The stepped wedge design employed in this project defies traditional taxonomies. In contrast to most stepped wedge studies, there are two levels of clusters. The sequencing of interventions occurred at the regional level, as each regional team was randomly paired with one of six two-months steps. On the other hand, the patient samples were drawn for each individual hospital, and the inspections were performed individually for each hospital. Hence, there were either 19 or 73 uptake times, depending on whether

one counts the steps at the regional level or at the hospital level. That is: a common baseline uptake, a series of uptakes directly prior to the intervention in each hospital, and two later uptakes per hospital (see Figure 2 under heading 4.3.2 above). Though the inspections were rolled out in evenly spaced two-month time periods, as is normal for stepped wedge designs (Girling & Hemming, 2016), the uptake periods were unevenly spaced because there were fewer patients with sepsis in the smaller hospitals. The design also included unusually long before- and after-rollout periods. In standard stepped wedge trials, the time periods before and after rollout are of the same length as the periods between rollouts (Thompson, Fielding, Hargreaves, & Copas, 2017). In the sepsis inspection, the time periods between rollouts were approximately two months, while the time period before rollout was almost one year and the one after rollout was more than one year. This affected the statistical power of the study, which depends on the number of crossover points (Baio et al., 2015). While a statistically optimized complete stepped wedge has no observations outside the rollout period (Thompson et al., 2017), our study had long uptake periods before and after the rollout where no crossovers happened.

I would argue that the discrepancy between our design and a "standard approach" to stepped wedge designs is almost a moot point. Even if the study design strays from standard stepped wedge patterns and the design offers less than optimal statistical power, the analyses are still valid. The estimated models are not dependent on any specific data collection design. The long before- and after-rollout periods have, however, compounded the challenges related to the fourth issue, controlling for calendar time.

Most of the patient records included were collected from a time period where either all records were from hospitals in the pre-inspection phase or all records were in the post-inspection phase. As illustrated in Figure 2 (section 4.3.2), this was the case for all records in the p0 samples of all regions, the p1 samples of regions South and West, the p2 samples of regions Central and East, and the p3 samples of all regions. Consequently, as described in study 2, the overlap period where clusters of hospitals switched from control to intervention included only 1433 of 7407 patients in the study. This

implies a high correlation between the dichotomous inspection variable and the admission date (point-biserial correlation = 0.87) and a high, though not alarmingly so, variance inflation factor in the various regression analyses, indicating a degree of multicollinearity in the models.[30] The last record in the pre-inspection data was from February 2017, while the last record in the post-inspection data was from July 2018. In the time period between these two records, there were almost 3000 records that all, naturally, were from post-inspection hospitals. The absence of any records from non-inspected hospitals over this time period translates into a problem of over-determination. What this meant for our interpretation, was that we had limited evidence to suggest how time influenced the care processes or patient outcomes, and that we could not accurately separate the effects of time from the effects of the inspection.

6.1.3 Trustworthy interpretations

Choosing interviews as one of the sources of data means choosing to create a situation in which we, in the words of Silverman (2007), manufacture data for use in our research. I would hasten to say that no data come to us as raw slices of fact. It is, however, important to acknowledge that interview research calls for a great deal of reflexivity on the part of the researcher. We need to be cognizant of how questions are devised, answered, and followed up. First, by preparing a set of questions the interviewer has already created an interpretative frame which will influence the interview. Secondly, the interviewee's memory and biases can play crucial roles in shaping the responses given. Thirdly, the way the interviewer consciously or unconsciously (verbally or otherwise) responds to the interviewees will further steer what information ends up in the final transcript. However, even a steadfast critic such as Silverman, who laments the "almost Pavlovian tendency to identify research design with interviews" (p. 42), concedes that some research topics require information that is not readily

[30] One textbook suggests that a variance inflation factor of more than 10 indicates that multicollinearity may be a problem (Lewis-Beck & Lewis-Beck, 2016). In the adjusted analyses where we included admission year, the variance inflation factors were slightly below or above 5.

available through observation or other venues of research, thus justifying the use of interviews.

The fact that the interviews yielded interesting data does not, however, absolve me from assessing the trustworthiness of the qualitative material and analyses. Self-disclosure and reflexivity are two central elements of such an assessment (Wertz et al., 2011). Before and after each interview, I took down fieldnotes. Reviewing these notes, I found some aspects of the interview process that figured prominently in the notes and that can help shed light on the interview data. First, my background inevitably influenced the interviews. I came to the fieldwork lacking medical training and experience from statutory inspections. From prior work, I had extensive experience from interviewing healthcare personnel, and from conducting audits and evaluations at hospitals and other healthcare institutions. I therefore had some knowledge about how hospitals operate and what the roles of different professional groups are, but I had never been in the shoes of a nurse or physician.

According to Morse (1994), lay persons (a group in which I most certainly belong) should use semi structured interview guides and be given close supervision. Both of these requirements were met in this project: We utilized a semi structured interview guide, and for the first couple of interviews, my role was first and foremost to take a seat at the table and "have big ears". Occasionally I would drop in a question or two, but I started leading interviews only when I got more familiar with the salient themes and jargon of work in emergency departments.

As I got more interviews under my belt and my understanding of both sepsis care and inspections evolved, I could participate more actively in the interviews. What I came to experience after a while, was that the way I introduced myself to a certain extent framed the following discussion. If I introduced myself as a PhD student with a background in social science, the focus group often tended to broach the details of care for sepsis patients more hesitantly than if I merely introduced myself as a PhD student. To elicit richer discussions, I therefore chose not to bring up my own background if I was not asked directly.

Second, I noticed a difference in the way the interviewees received and interpreted the questions depending on whether it was a focus group from the hospitals or one of the inspection teams. The focus group is an arena for extemporaneous debate and sharing of thoughts. Though we interviewers tried to steer the conversation towards specific questions, the focus of the focus groups would be directed towards common points of reference. In the case of the interviews with the inspection teams, the conversations naturally centered around our main topic of interest, the inspections. The focus groups with clinicians and managers would not go down that train of thought so easily. Instead, these groups tended to linger on the themes most immediate to their own day-to-day work: for the clinicians, patient care; for the managers, administration and leadership.

Even though the focus group interview is an interview, not a discussion (Patton, 2015), I had to avoid being too obtrusive. I was often eager to have the interviewees address the inspections explicitly, but I found that there was more to gain by listening patiently then by cutting off their discussion. The information from the clinicians and managers was very valuable for shedding light on the context and mechanics of the inspections. Their information would provide just as many interesting perspectives on how the inspections functioned, even though the clinicians and managers to a lesser extent than the inspection teams provided overt assessments of the inspection.

A third aspect that influenced the interviews and the later analyses was theory. Theory guided the data collection from the start. The interview guides were developed in a collaboration between us interviewers, partly using experience from previous work in healthcare institutions and inspection teams, and partly using concepts derived from theory related to health governance and quality improvement. I do not believe it is possible to enter interviews or fieldwork as a *tabula rasa* (Yin, 2018). The theoretical notions that informed the study in the first place were bound to influence the interviews, and I believe that consciously including these perspectives in the interviews was beneficial. As noted above, however, the hospital interviewees had perhaps not spent a considerable time thinking about inspections. Sometimes, I therefore had to rephrase the questions and provide examples of what types of experiences, activities, or effects

that we were interested in. By giving examples in such a way, there is always a risk of suggestive questioning that Silverman (2007) cautions against.

Theory entered the stage even more prominently in the analysis phase. Even though we were mindful of letting the data speak for itself, many of our analytical concepts came from theory and were subsequently analyzed within a theoretical framework. The resulting findings in study 3 did to a certain extent confirm existing theoretical notions. Could this be a warning sign indicating that we had allowed the theory to guide both interviews and analyses to the extent that other interesting avenues of inquiry were blocked off? A potential weakness of the design of study 3 was that we only sampled "positive cases", i.e. all the three hospitals had experienced substantial improvement following the inspection. Including hospitals where no improvement had taken place might perhaps have lent more rigor to our analysis, providing us with the opportunity to test the mechanisms against a "critical" case (Yin, 1999). It was, however, not possible to find such a critical case among the hospitals from which we had interview data. None stood out as cases of "inspection failure", even if the two hospitals we left out showed less clear-cut improvement than the four we included in the study.

Thus, we chose to sample a collection of clearly positive cases instead, and to go about the analysis as carefully and methodically as possible. Analyzing interviews is a daunting task, and one that is immensely personal for the researcher. What this comes down to in the end, is in my opinion mainly reflexivity and honesty. I have to think hard about how our preconceptions have influenced the interviews, and I have to perform the analysis as honestly as possible and be transparent about the whole process. Including the coauthors in the analyses introduced a second kind of triangulation, "analyst triangulation" (in addition to the triangulation of methods), to the research process. This can be a helpful way of reducing interpretive bias and increase the credibility of qualitative studies (Patton, 1999).

6.1.4 Ethical considerations

Assessing the ethics of research involving human subjects, means answering several different questions regarding how the research was planned, conducted, and reported. We used information from two groups of humans: patients and interviewees.

Whereas all interviewees gave their informed consent in writing, we did not obtain patient consent. A waiver of informed consent from patients was granted in the approval from the ethics committee (2015/2195/REK nord). Informed consent is not an indispensable demand for the research to be ethical in cases where it is not possible to obtain informed consent, as was the case in the sepsis inspection. Furthermore, other important areas need to be included in the ethical assessment, such as the scientific soundness of the research, the overall benefit to patients, and the respect for potential and enrolled subjects (Emanuel, Wendler, & Grady, 2000). The ethics committee found that the study had ensured the welfare and integrity of the patients.

The intervention was directed at the hospitals, not individual patients. On the other hand, the intervention could impact patients at the hospitals indirectly, through its potential impact on care processes. As such, the choice of which hospitals to inspect would bestow potential benefits (or losses) to future patients of the hospital, even if patients were not targeted individually in this research project.

The research project was designed in such a way that the potential risk to patients was minimal. A degree of collective equipoise among experts, i.e. the lack of a preference for one treatment over another, is a legitimate requirement when conducting a medical trial (Lilford & Jackson, 1995). By allowing for the intervention to be administered to all clusters, a stepped wedge trial avoids ethical dilemmas in cases where there is uncertainty about equipoise (Brown & Lilford, 2006).

On the other hand, an argument against stepped wedge trials is that it would be unethical to continue with the implementation of the intervention in all clusters if it has been proven ineffective (Kotz et al., 2012). This does not apply to our study. The effectiveness of the inspections could only be measured when nearly all hospitals had

been inspected. And the inspection would be carried out for all hospitals regardless of the findings from the research project.

It is also important to bear in mind that the research project did not influence the decision to carry out the inspections. The choice of which healthcare organization to inspect will always be a potential ethical dilemma for the inspection authorities. This was also the case with the sepsis inspection. However, the area of sepsis care in hospital emergency departments was arrived upon prior to the decision of starting a research project into its effects. Furthermore, the overall design of enrolling four hospitals from each of the six different regions would probably not have been any different if the inspections were not the subject of a research project. Like in other inspections, it was the inspection teams from the County Governors who prioritized which hospitals to include within each region. What the research project did influence, was the overall timeframe of the inspection period and the sequence of regions within this timeframe. As the timeframe was comparable to that of ordinary inspections and the sequence was determined by randomization, the influence of the research project on the delivery of the intervention did not pose any ethically objectionable burdens on the patients.

6.2 Substantive research contributions

In order to appraise the contributions of the present research to the existing knowledge within the field, I will turn back to the three research questions posed in chapter 3:

1. Are the care processes evaluated in the inspection relevant for assessing the quality of care in the emergency departments, and what can the care processes tell us about the performance of the emergency departments on a systemic level?
2. How have the inspection affected the performance indicators for diagnosis and treatment of sepsis, and for patient outcomes?
3. What is the role of inspections in bringing about improvement efforts in the emergency departments?

In our effort to answer these questions, we can gain an understanding of the role of inspections in health care that goes beyond that of this specific case.

6.2.1 The relevance of the inspection

The first research question directs our attention to the performance indicators assessed in the inspection. For the inspection to succeed, these indicators must be valid as measures of quality of care in an emergency department setting; they must reflect a set of care processes that, if completed in a timely manner, represent good quality of care for patients with sepsis. Selecting the proper indicators, criterions, or standards is a core activity of any regulatory work (Scott, 2010). As van Dijk et al. (2020, p. 2) point out, a major challenge for healthcare regulators is "to identify those standards that have the most impact upon the quality of care and to identify the information that is most relevant for monitoring quality".

The chosen indicators in the sepsis inspection have face validity as measures of good care, as they have been compiled from authoritative sources such as academic journals, expert advises, and clinical guidelines (Dellinger et al., 2013; National Institute for Health and Care Excellence, 2016). Furthermore, the findings in study 1 show that the different care processes are internally linked: Delays in initial diagnostic processes lead to delay in time to treatment. This suggests that the chosen indicators indeed not only cover important aspects of quality of care for patients with sepsis, but that they do so in a comprehensive way and taken together can shed light on the systemic quality of care.

Study 1 confirmed what has been documented in previous research regarding variability in emergency department settings of initial diagnostic processes and time to treatment.[31] Thus, we have an answer to the second part of research question 1: The degree of non-completion and delays in the different care processes indicate that the performance of the emergency departments did not meet the desired standards. The findings reflect how complexities in the handling of patients have insidious consequences: For some patients, delays may occur both when waiting for triage, when waiting for the physician's examination, and when waiting for blood works. When

[31] Variability and delay in diagnostic processes and treatment in the emergency department setting had been documented by among others Goodwin et al. (2015), Houston, Sanchez, Fischer, Volz, and Wolfe (2015), Liu et al. (2017), Morr et al. (2017), Stoneking et al. (2015), and The Royal College of Emergency Medicine (2017).

piled on top of each other, the delays in the completion of diagnostic procedures result in delays in treatment that can be harmful for the patients.

It should be noted that the somewhat counterintuitive initial drop in predictive mortality as a function of time to treatment with antibiotics could be interpreted as an indication that the inspection's focus on reducing time to antibiotics was note entirely well-founded. This relationship between time to treatment and antibiotics can probably, as discussed in 6.1.1, in large part be explained by unobserved differences in severity of illness depending on time to treatment. It nonetheless shows that the set of processes by no means gives a picture-perfect image of the quality of care for patients. Focusing only on the proportion of patients who have had antibiotics within one hour obscures the fact that for some patients, some extra time waiting for antibiotics may not be all that detrimental to their health.

On the other hand, it is important to bear in mind that other studies have shown that there is a correlation between time to antibiotics and mortality where expedited administration of antibiotics on average leads to improved outcomes (Liu et al., 2017; Seymour et al., 2017). Delays seldom contribute to improved patient outcomes. This also holds true for the general patient population presenting to the emergency departments. Delays can increase the risk of death or (re)admission to hospital (Guttmann, Schull, Vermeulen, & Stukel, 2011). Reducing unnecessary delays in the initial diagnostic processes can therefore benefit all emergency department patients.

There is an ongoing debate in the field of sepsis research regarding how quick is quick enough when it comes to treatment of patients with sepsis. The tensions came to a head when the Surviving Sepsis Campaign released an update of the practical application of the general guidelines, combining treatment bundles that previously had been recommended as completed within 3 or 6 hours into a new "1-hour bundle" (Levy, Evans, & Rhodes, 2018). Critics railed against the Surviving Sepsis Campaign for failing to take into account evidence questioning the efficacy of these treatment protocols,[32] warning

[32] Though this subject is seemingly narrow in scope, the discussion has been ongoing for many years and fired up researchers of many stripes (and resulted in a vast amount of research papers and opinions). It is not possible

that the guidelines "would cause hasty management decisions, inappropriate fluid administration, and indiscriminate use of broad-spectrum antibiotics" (Marik, Farkas, Spiegel, & Weingart, 2019, p. 14).

Scholarly discussions like these can also become points of contention in the dialogue between clinicians and external assessors if the regulatory authorities have employed standards that are disputed in professional and academic circles. This was born out, albeit to a limited extent, in one of the focus group interviews, where a physician questioned the choice of performance indicators used in the inspection. The criticism was attenuated, however, by the fact that NBHS had deemphasized some contested care process indicators, such as oxygen therapy (see Table 3 in section 4.4).

The discussion regarding the appropriate timing of the different sepsis care processes exemplifies the uncertainty inherent in all forms of specialized knowledge, and the challenges this uncertainty poses for the external assessment. What is considered gold standard medical care is subject to change: What might have been considered to be relatively straightforward recommendations for proper care might turn out to be highly contested issues among experts, and what was thought to be good practice in the planning stage of an inspection could potentially end up being an Achilles heel once the project is set in motion. One could ask why setting quantitative standards in a field marked by contested clinical best practice did not cause more resentment among the clinicians. I will try to address this question later in this chapter, after we have reviewed the effects of the inspection on care processes and patient outcomes.

---

to provide a thorough examination of the arguments here. Therefore, the following must suffice as a background: An important point of contention in the debate stems from a 2001 study (Rivers et al.) that recommended "early goal-directed treatment" relying on fluid resuscitation, vasopressors, and vasodilators, guided by monitoring of central venous pressure and central venous oxygen saturation. These recommendations were included in the first Surviving Sepsis Campaign guidelines (Dellinger et al., 2004), and were introduced as a part of the sepsis bundle, i.e. a selected set of interventions or processes of care for patients with sepsis (Levy et al., 2004). Though some researchers have found evidence suggesting that bundle care may increase the chances of survival (Barochia et al., 2010; Levy et al., 2015), others have failed to find such associations and sown doubts about whether bundle care represents an improvement over regular care for patients with sepsis (The ProCESS Investigators, 2014). Additionally, when it comes to early therapy with broad-spectrum antibiotics, the guidelines have been criticized for being based on too-weak evidence (Patel & Bergl, 2019). When the 2016 update of the Surviving Sepsis Campaign guidelines were published (Rhodes et al., 2017), the Infectious Diseases Society of America did not support the new guidelines. They pointed to the risk of guideline mandated early antibiotics leading to inappropriate antibiotics use as one of the reasons for why (IDSA Sepsis Task Force, 2017).

6.2.2 The effects of the inspection

The second research question is how the inspection have affected the performance indicators for diagnosis and treatment of sepsis, and for patient outcomes. For the sake of clarity, we can break down this research question into two sub-questions: (1) Can any changes in performance indicators for sepsis care be attributed to the inspection? (2) Do the changes in care processes result in improved patient outcomes?

Regarding the first question, study 2 found a significant improvement in 5 of 11 care processes when controlling for potential confounders. It should be noted that the effect estimates of the inspection were in a positive direction for all but one of these processes. Adding to this, all care processes showed an improvement in relative terms from before to after the inspection, and in the unadjusted analyses 9 of 11 processes were found to be significantly associated with the inspection. Hence, it seems safe to presume that the inspection had an effect in improving the care processes. This finding is contrary to that of previous research using experiments or other statistically robust approaches, which in general has found little or no evidence of improvement attributable to inspection.[33]

Regarding the second question, study 2 did not find any significant improvement in mortality or length of stay when controlling for other variables, including time. As discussed in section 6.1.2, controlling for time is a tricky issue in our statistical models and one that can potentially skew our estimates of inspection effects. If we however set aside these questions of possible biases for now, the lack of significant effects on patient outcomes is interesting.

We can hypothesize a couple of different mechanisms related to how improvement of care processes might affect the two outcome measures. While an improvement of care should lead to lower mortality, it might have mixed effects on length of stay. Shorter length of stay is what we would expect for patients who otherwise would have stayed

[33] Such studies include randomized controlled experiments (Salmon, Heavens, Lombard, & Tavrow, 2003; Wesselink et al., 2014; Wesselink et al., 2015) and time-series design (Castro-Avila et al., 2019; Castro, 2018). See section 2.3.2 for an overview of the research on inspection effects.

for longer until adequately healed and discharged. Patients whose lives are saved by the improved care processes, on the other hand, might have longer length of stay than they would have had had they died while in hospital. For a condition with high in-hospital mortality such as sepsis these effects could be of a consequential magnitude. Looking at the averages for different subgroups in our material, we find that for the 980 patients who had both a Charlson comorbidity index of above 3 and organ dysfunction (including both pre and post inspection groups), the average length of stay was 8.5 days. For those within this group who had not died within 30 days, the average length was 8.9 days, while for those who had died within 30 days, the average length was 6.9 days. For these high-risk patients, at least, it does not seem unrealistic to expect that a reduction in in-house mortality could offset a reduction in average length of stay brought on by improved care processes.

As for the effects of the inspection on mortality, we could have expected that improved care processes led to lower mortality. This was suggested by a study by Seymour et al. (2017), mentioned in 6.2.1, of the effects of a 3-hour sepsis bundle that included obtaining blood culture before antibiotics, measurement of serum lactate, and administration of antibiotics. They found that time to completion of the bundle and time to antibiotics were associated with increased risk-adjusted in-hospital mortality. Since we found that these processes were among those that improved following the inspection, there were reasons to believe that this would affect the mortality in our study as well. However, the lack of significant effects of the inspection is not altogether surprising, as there are studies suggesting that the association between improved sepsis management and mortality is not straightforward. Investigating the effects of a protocol of sepsis screening and patient management in the emergency department, Gatewood, Wemple, Greco, Kritek, and Durvasula (2015) found expedited care delivery following the intervention but not reduced mortality.

It is instructive to again consider the findings reported in study 1 related to the association between time to antibiotics and mortality. The mortality was highest for those who received treatment within one hour, and it rose again for those who received antibiotics later than three hours from admission. As pointed out in 6.2.1, the

importance of immediate broad-spectrum antibiotic treatment for patients with sepsis is contested (Patel & Bergl, 2019). It is possible that for a good portion of the patients in our sample, starting antibiotic treatment as early as within one hour after admission would not have impacted their chance of survival significantly. Timing of antibiotics is found to be especially crucial for patients with septic shock, i.e. the patients who are worst off (Kumar et al., 2006; Puskarich et al., 2011). The lack of significant effects of the inspection on mortality may in part be explained by the fact that most of the patients in our sample did not belong to this high-risk group.

Furthermore, the effect might get distorted because of the complex chain of events leading from the inspection to the patient outcomes. We can anticipate that an initial effect on care processes will be distorted along the way, due to other factors influencing the mortality rate. Compare this, metaphorically, to listening to a recording. No matter how pure the tone produced by the musicians in the recording studio, the sound is going to be affected through the signal processing chain, going through microphones, meters of cable, sound processing, compression, and in the end squeezed through the speakers in your living room. The inspection process is a much messier affair when it comes to getting the signal through without gaining to much noise on the way.

We should remember that in studying the sepsis inspection we have used indicators specifically related to patients with sepsis. Castro (2018) argues that if we are to compare the effect of inspections, we should focus on broader, hospital-level quality indicators, such as in-hospital mortality, rather than disease specific indicators. When it came to care processes, we used the same indicators as was used by the inspection, namely a wide set of indicators related to diagnosing and treating patients with sepsis. To be sure, these variables also included aspects of patient care that are important for other groups of patients in the emergency department, but the indicators were tailored to the standards for good care for patients with sepsis. In addition, we also included the patient outcome variables, but we restricted the measurements of outcomes to patients with sepsis. Thus, it was less likely that the effect estimates could be influenced by other events (i.e. “distorted”), as the effect of an inspection into one specific care process could have been drowned out if we had studied the effects on a hospital-level.

On the other hand, it is important to note the difference in level of analysis if we are to compare our results to those of studies using hospital-level measurements.

6.2.3 How the inspections worked

The third research question was: What is the role of inspections in bringing about improvement efforts in the emergency departments? I must admit that I will arrive at an answer only partially and conditionally.

The use of indicators

In study 3, which was aimed at understanding how inspections could promote improvement efforts, we emphasized the importance of indicators. Through the medium of clinically relevant indicators, the hospitals were made aware of deficiencies in their current practices and then set about trying to improve these practices.

The very existence of performance indicators can be a form of "action at a distance" (Rose & Miller, 1992, p. 187). By introducing mechanisms of recording and scoring the performance of hospitals, the government can instigate changes without having to resort to blunt enforcement. The hospitals themselves engage with their own performance metrics and come up with plans to improve. The potential power inherent in the selection of performance metrics is not to be trifled with. Inspections may very well be a forceful instrument of change but is it the desired change? In order to better understand the potential benefits of good indicators, we need to consider the potential drawbacks of not-so-good indicators. I want to address three closely related pitfalls of using performance metrics for regulatory purposes: gaming, validity, and prioritization.

Firstly, one could have a situation of gaming or "managing to the audit". As described in chapter 2, gaming is the process whereby organizations manipulate or strategically adapt their work in a way so that performance targets are reached, though, in reality, there is no overall improvement of the quality of services. This concern has been much debated over the last decades in the study of healthcare inspections (Bevan & Hood, 2006; Toffolutti et al., 2017).

The improvement of performance indicators from before announcement to after announcement can be a telltale sign of gaming. One would expect gaming to occur between announcement and inspection, as the hospitals are judged on the basis of the performance at the time of inspection. Of course, improvement of performance metrics after announcement can also be a sign of genuine improvement, and thus of legitimate *anticipatory impact* (Smithson et al., 2018). However, if the hospitals were able to game the performance indicators, it would most probably happen before the inspections. Analyzing the announcement effects was not within the scope of this study. We can, however, see the change in performance from pre-inspection to post-inspection for the hospitals that belonged to the last two steps of the stepped wedge. These hospitals received the announcement that they had been selected for inspection after the inspection reports had been completed and published for the inspections in the first step. These later hospitals then had between two and a half and five months to prepare for the inspection. If gaming was the sole reason for improvement, we should have seen little or no improvement from pre-inspection to post-inspection. This was not the case. These hospitals also had significant improvements in key care processes. Examples of this include two of the four hospitals that were included as cases of successful inspection processes in study 3.

These results indicate that Goodhart's law is not ironclad. The problem of gaming does not stem from performance measures per se but from the performance measurement systems that, in the words of Bevan (2010), punish "knights" and reward "knaves". The sepsis inspection included a set of indicators that could not be manipulated easily. As argued in studies 1 and 3, the care processes were closely related to patient care, and they were congruously linked in such a way that a concerted effort to improve all processes would be commensurate with a systemic quality improvement of sepsis care. Save from outright fabricating the data or purging electronic health records, the hospitals had to *de facto* improve their care processes in order to improve the indicators. We could obviously argue that we do not know whether the changes were instigated to advance the patients' welfare or to be seen as successful in the eyes of the powers that be. It is futile to try to empirically distinguish one from the other, as people tend to hide

any disreputable motivations. Also, ethically, at least from a consequentialist standpoint, one could argue that any action that produces a desired result is fine, whatever motivations lay behind the improvement efforts.

The problem of gaming is closely related to the second theme: whether the performance measures are, in fact, valid measurements of quality of care. The organizations might achieve changes that the inspectors (and perhaps the hospitals) perceive to be good results but that in fact are not. Or, the chosen indicators may be known to have poor validity but still be the best alternative out there. Thus, operationalizations of quality of care can in reality measure something completely different and clinically irrelevant – what Power (1994, p. 28) describes as "structures of auditability embodying performance measures which increasingly do not correspond to the first order reality of practitioners' work".

In the sepsis inspection, this kind of risk could potentially arise if a singular focus on a select few care processes led to other important processes being overlooked. This, however, does not seem to have been born out in the sepsis inspection, possibly due to the comprehensive set of indicators that were selected for review. Study 3 showed that the hospitals welcomed the way the inspection focused on the patient-care process as a whole. As mentioned above, the performance measures in the sepsis inspection were closely tied to the practical realities of work in emergency departments, and in the interviews the measures were lauded as being related to issues critical to patient care. External specialists were involved at an early stage in order to secure the most representative set of indicators possible. In study 3 we argue that this effort resulted in improved legitimacy for the inspection. This backs the argument referred to in chapter 2, about the importance of focusing on issues and indicators that are relatable for the healthcare personnel (Campbell et al., 2010; Hovlid, Braut, et al., 2020; Åsprang et al., 2015).

Thirdly, the choice of indicators can change the prioritization from one group or one service to another. Prioritization is a legitimate aspect of inspection: The areas under

inspection are given special attention. However, prioritization can also happen inadvertently. A study by Helgesen and Hanssen (2014) of healthcare governance in Norwegian municipalities gives an example of this. They found that the focus of County Governors in their inspections were metrics for institutional care, because these metrics were more easily applicable as audit criteria, even if the government prioritized other forms of care such as home care services. Moreover, the focus of one set of indicators over another could lead to a situation where care is in fact improved for the target group of patients but where some other groups are worse off than what they would have been without these new rules. Hood (2011) describes how "protocolization" can lead to iatrogenic risk:[34] ostensibly aiming at reducing risks, stricter protocolization leads to higher risks for some groups. This would be the case if, for instance, the changes in care processes benefited patients with sepsis through more timely care but this prioritization made other patients of the emergency department fall between the cracks.

As was the case for the validity of the performance indicators, the choice of a comprehensive set of performance indicators seemed to mitigate the risk of prioritization at the expense of other groups. The sepsis inspection included several indicators related to early diagnosis, such as time to triage, time to examination by physician, and measurements of vital signs. A more expedited completion of these processes would not only benefit sepsis patients but all patients presenting to the emergency department.

Two more points are to be added to the discussion of indicators. First, indicators need to be sensitive to improvement (Hovlid, Braut, et al., 2020). In her study of the effects of the Care Quality Commission's inspections in England, Castro (2018) argues that improvements will be slower and more subtle if the standard of services already are adequate. Our findings from the sepsis inspection support her reasoning: As explained in studies 1 and 3, the emergency departments clearly had a room for improvement.

[34] Here, "iatrogenic" refers to regulatory iatrogenesis, analogous to medical iatrogenesis (care induced injury). Coincidentally, one of the earliest investigations of medical iatrogenesis was in fact on (puerperal) sepsis (Wiener, 1998).

For almost half of the patients before inspection, two or more key diagnostic procedures were not completed in a timely manner, and about the same proportion of patients received antibiotics within two hours.

Second, as alluded to in chapter 2, inspections can be construed as "inscriptions". Power (1997) noted that in the case of audits, the efficiency and effectiveness of organizations in some cases could be seen as 'constructed around' the audit, rather than verified by the audit. When the organization lacks clear performance measures, the audit defines and operationalizes such measures. In a similar vein, de Kam's (2020) notion of "regulatory objects" offers a perspicacious take on regulation and inspections. A regulatory object "defines a particular quality issue as the (legitimate) object of regulatory scrutiny" (p. 14). Expanding on the idea of quality inscriptions and the constitutive effects of performance indicators (Dahler-Larsen, 2014, 2019), de Kam argues that regulatory objects, along with the instruments that render the quality of these objects documentable, partake in constituting the notion of quality.

This is not to say that performance measurement systems are omnipotent. Within the regulatory realm, constitutive effects of performance measurements are mediated by other sources of information (Kok, Leistikow, & Bal, 2019). Furthermore what happens "on the floor" can be decoupled from the goals and performance measurement systems (de Bree & Stoopendaal, 2018; Meyer & Rowan, 1977). What the notion of regulatory objects can provide, is a corrective: It encourages us to be open to the idea that in some instances the conceptions of the quality we want the inspections to improve are influenced by the inspections themselves.

Organizational learning and context

I have now described a list of potential pitfalls and argued that the sepsis inspection managed to address these risks, at least partially, by way of choosing a comprehensive set of indicators. But this only provides a limited glimpse into the workings of the inspections. We also need a deeper understanding of the *organizational impact* (Smithson et al., 2018); how the meeting between hospital and inspection authorities

influences the hospitals' actions, especially those pertaining to quality improvement efforts.

What we have come to learn from the sepsis inspection supports Walshe's argument about inspection authorities having to rely on quality improvement mechanisms within the hospital (Walshe, 2003). Hence, inspection effects are inextricably linked to the organizations and their potential for change and learning. This means that all effects are contextual, to the extent that the potential for change and learning are part of the organization's culture, i.e. its previous shared learning (Schein & Schein, 2016). Therefore, the mechanisms will most likely operate differently from organization to organization.

One of the ways in which organizations differ in their response to inspections, is in who drives the improvement efforts. Whether the inspection is to function via anticipatory effect or through more direct routes, its success is dependent on people within the organization who are willing to either put quality themes on the agenda themselves or who take up the mantel of the improvement work once the inspection authorities have forced the issue.

In study 3 we found that managers as well as healthcare professionals can be empowered by inspections. This echoes an upcoming study by Weenink, Wallenburg, Leistikow, and Bal (2020), who found that inspection frameworks are being used for quality improvement in different ways: by managers to check if quality meets the basic standards, or by the professionals to stimulate discussion and learning. One should avoid viewing this as a purely dichotomous relationship, however. There are good reasons to believe that improvement is best served by a culture of mutual learning and cooperation between clinicians and managers. A recent study into how healthcare organizations prepared in advance of upcoming NBHS inspections found that managers who involved clinicians in the assessment of care delivery tended to initiate improvement efforts to a greater extent than managers who did not involve clinicians (Hovlid, Teig, et al., 2020). Another recent study that is interesting in this respect, is a mixed-methods inquiry into how sentinel event reports, mandated and assessed by the

Dutch healthcare inspectorate, affected learning in hospitals (de Kam et al., 2020). To meet the increasing need for producing reports following incidents, hospitals would set up dedicated investigation teams.[35] The researchers found that the reports failed to adequately connect the learning process of the investigation team to that of the healthcare professionals.

Acceptance of the legitimacy of the inspection and a culture that can support improvement efforts are not the only prerequisites for quality improvement, however. The inspected organizations must also have adequate resources to initiate improvement efforts (Castro, 2018).

Our studies have not focused on the economic aspect of the inspection. Therefore, it is not possible to say anything definite about the use and availability of resources, other than that hospitals for the most part seems to have had necessary resources for following up the inspections and initiating improvement efforts. However, in the phase following the inspections we found an example of disagreement between one of the hospitals and the inspection team regarding how to address the deficiencies. To correct the nonconformities listed in the inspection report, the hospital had to adjust the work schedules so that resident physicians were available to the emergency department at all hours. According to the managers, this adjustment increased the toll on resident physicians. They said it could lead residents to resign, which was perceived as a serious risk to maintaining full capacity at the hospital, as it already was difficult to recruit enough physicians into these positions. The managers agreed that, ideally, resident physicians should be available at all times, but implementing these changes quickly left the hospital in a precarious position. One of the managers said: "this [change] is the right thing to do but you can't do it over night". This is an important reminder that there is a cost dimension to inspections, and that workload and opportunity costs can

[35] The dedicated investigation teams, it should be noted, lend support to Downs' "corollary to the Law of Counter Control […] that any increase in the number of persons monitoring a given bureau will normally evoke an even larger increase in the number of bureau members assigned to deal with the monitors" Downs (1967, p. 152).

represent a negative impact of inspections (Walshe, Wallace, Freeman, Latham, & Spurgeon, 2001).

Next, we turn to the inspectors and the *relational impact* (Smithson et al., 2018) of the inspections. What do the inspection teams bring to the table other than providing "hard" numbers? It seems from study 3 that they put a human face on those numbers. The legitimacy of the inspection was not all down to the acceptance of the indicators as relevant to the daily care processes. Another important factor was that the managers and clinicians acknowledged the expertise of the inspection team, and especially the specialists' expertise. As noted in chapter 2, the importance of the quality of the inspection team has been underscored in previous research (Hilarion et al., 2009; Walshe & Boyd, 2007). We found this to hold true for the sepsis inspection. It was the combination of hard numbers with a clinical understanding and real-life experience of the inspection teams that brought the clinicians onboard. We also found that the inspection teams' interpretations of the data were important in helping the hospital identifying potential areas for improvement. With their outside-in perspective, they can aid in contextualizing the external requirements. This role of the inspection authorities was also noted in a recent case study of the legal framework for internal control and quality improvement in Norwegian healthcare (Øyri et al., 2020), which found that the healthcare organizations wanted guidance from the inspectors regarding how to apply the internal control requirements to their own context.

Moreover, and perhaps more hidden from view, is the importance of the on-site visit for the inspection teams' assessment of the quality of care at the different emergency departments. This includes all the information gathered through documentation, interviews, and observations which are not included in the "hard" indicators (Walshe, 2011). Kok, Wallenburg, Leistikow, and Bal (2020) explain how information that is not easily classified within the existing data management system – what they refer to as "soft signals" – is essential in that it provides the necessary context to understand performance metrics and assess risks. The softness of this information is related to how it is used and made sense of (Weick, 1993), not necessarily to any inherent characteristics of the types of data or seriousness of the information.

By giving the clinicians room for reflecting on the metrics, the inspectors facilitate the learning processes within the hospitals. Preparing for and following up external assessment requires communication across organizational and professional boundaries, which can be a catalyst for learning (Baskind, Kordowicz, & Chaplin, 2010; Greenfield, Pawsey, & Braithwaite, 2011). As described in study 3, the inspections created opportunities for groups of clinicians and managers to come together to discuss patient care. These meetings could act as what Nonaka (1994) terms "communities of interaction", essential to the processing and development of knowledge within the organization. Furthermore, the inspectors are themselves learning and revising their approach to the present and upcoming inspections; they are engaging in triple-loop learning (Healy & Braithwaite, 2006).

We thus find support for the notion of facilitating change by providing arenas for discussion and reflection (Hovlid, Braut, et al., 2020). Even if inspections formally address the hierarchical chain of management, this does not mean that the change exclusively works through a top-down implementation. Networks of clinicians and managers are also crucial, and these professional networks have further ties to the wider professional community and to the inspection authorities. To borrow a metaphor from Niall Ferguson (2017), the drive for change from the “public square” of such networks, can be just as strong, if not stronger, than that from the “tower” of management.

Accountability

Discussing how the inspections worked, I have touched on themes such as choosing clinically relevant indicators, bringing clinicians onboard, and providing arenas for reflection. We also need, however, to address the role of the inspector. The importance of having a perspective from “the outside” was emphasized by one of the managers quoted in study 3. But it is not merely “anyone” who comes along with interesting performance data and opens discussions about clinical practices. It is the inspector; and that counts for something. Understanding a regulatory scheme like NBHS’s inspections requires grasping the underlying accountability structures. Accountability is an

important reason for why we have inspections in the first place. Demands for greater accountability has been a main force behind the introduction of external reviews in healthcare systems (Shaw, 2000). An ever more complex society necessitates trust between strangers, and the inspection authorities are "the new guardians of trust" (Power, 2007).

This role as a "guardian of trust" speaks to the fact that the performative qualities of the inspections are important. The inspection authorities signal to the healthcare organizations what is expected of them, and at the same time, they promote trust in the healthcare system as a whole. Making inspection reports available on the Internet further secures a form of public horizonal accountability for the organization (Bovens, 2009). Describing the role of the Care Quality Commission in England, Castro (2018, p. 250) argues that the commission mainly exerts its effect "through its existence by creating a constant pressure to maintain standards to avoid regulatory actions and to perform well during an inspection without any additional effort being required" and that it "may be viewed as acting as an invisible (but ever present) reminder to meet standards of practice and strive for improvement." We can call this the "speed camera justification" for healthcare inspections, which basically resonates with non-responsive regulatory strategies. Whether the inspection authorities identify subpar care or not, and whether they actually provide valuable information to the healthcare organization or not, is less important than the fact that people know they can be busted if their performance drops below the approved levels.

So how does this conception of accountability compare with the approach of the sepsis inspection? It seems clear that the role of the inspection teams appears not to be based on a strategy of deterrence, where the inspection authorities, armed with a discombobulating array of forms, protocols, and guidelines, clamp down on the hospital. Neither is it exclusively focused on compliance. As Leistikow (2018) reminds us, promoting quality improvement requires different regulatory strategies from those used for promoting compliance. However, even if NBHS has improvement as a goal, it would be stretching it too far to call this regulatory approach a "hybrid", i.e. providing direct action improvement support contingent on the organization's performance

(Furnival, Boaden, & Walshe, 2018b). But the approach of the sepsis inspection can broadly be termed reflexive or responsive.[36]

This brings us back to the inherent dangers of enforcing rigid regulatory standards. Sepsis care is an especially interesting case when it comes to this discussion. One of the reasons why the Infectious Diseases Society of America chose not to support the 2016 Surviving Sepsis Campaign guidelines was that they deemed the guidelines' one hour time limit on administration of antibiotics to potentially cause uninfected patients to receive antibiotics "in the rush to meet the fixed frame stipulated for infected patients" (IDSA Sepsis Task Force, 2017, p. 1632). Along the same lines but in a more dramatized fashion, lamenting the effect of what they judge to be non-contextualized or fictitious claims about sepsis, Singer et al. (2019, p. 1513) write:

> Patients and families fear the so-called hidden killer and their confidence in health-care providers is undermined. Hospitals are criticised, penalised, and litigated against for failing to give patients antibiotics within 1 h of presumptive diagnosis. Doctors are reported for not giving antibiotics to patients they deem noninfected.

What these authors are describing are the processes that Smithson et al. (2018) refer to as *informational impact* and *stakeholder impact*. Publicizing information influences the perception of the hospitals in the eyes of the public and stakeholders. But as Singer et al. point out, there can be a downside to this. It can lead providers to avoid providing what they believe to be the best care, because they fear retaliation. The argument echoes a central tenet of prospect theory, namely that potential losses are weighted more heavily than equivalent gains (Kahneman & Tversky, 1979). Such negativity bias can be institutionalized into control mechanisms. Failure gets all attention, which results in risk aversion on the part of the practitioners (Hood, 2011).

[36] Classifying the style of regulation in this way must be approached with some caution. We cannot use one specific inspection as evidence for characterizing the entirety of the regulatory regime. The inspection is just one tool from a larger box or regulatory tools, and to properly describe the regulatory governance, we need to understand how the different tools work together as a whole.

Accountability in health care cannot, however, be equated with this "blame game". There are clear limits to the power of hierarchical chains of accountability in health care and other public services dominated by professionals (Lipsky, 2010). Healthcare personnel and healthcare organizations have multiple accountabilities (Hupe & Hill, 2007). In Norway, for example, hospitals are not only subject to the kind of administrative accountability performed by the inspection authorities. They are also politically accountable as subordinates of popularly elected bodies, they are socially accountable to users and user groups, and they are professionally accountable to their peers (Byrkjeflot, Christensen, & Lægreid, 2014). The resulting network of accountabilities can be described as a "combination, mingling, and competition between accountability mechanisms" (van Erp, Wallenburg, & Bal, 2020, p. 48).

It has been argued that such patchworked accountability structures for healthcare professionals have evolved from an accountability structure primarily based on collegiality and trust (Duckett, 1983; Ewert, 2020). This shift, it is said, entails a renegotiation of the public mission entrusted to the professionals (Grimen, 2006). From the healthcare personnel's perspective, the perceived shift can of course be experienced as an erosion of trust in their profession. As Doc Deeneka laments in Joseph Heller's *Catch 22*: "It's a terrible thing when even the word of a licensed physician is suspected by the country he loves" (Heller, 2011 [1961], p. 46).

Ewert (2020, pp. 320-321) argues that "measuring healthcare performance through checklists and protocols fails to comply with physicians' cross-cutting identity facet as a professional", and that for physicians, "rendering account to somewhat anonymous regulatory bodies […] hardly promotes healthcare quality." I do not support this claim unreservedly. While I agree that giving a human face to the inspection probably is important, the measurements and protocols are part and parcel of what it is to be a physician in a specialized healthcare setting. What is important for the legitimacy of the inspection process, and hence the accountability of the healthcare personnel and healthcare organizations, is that the measurements are seen as relevant to the patient care, which they were in the case of the sepsis inspection, as explained in study 3. Thus, accountability salience, i.e. the perceived importance of the outcomes that are at stake

(Hall et al 2017), can be an important determinant on the felt accountability of the healthcare professionals and healthcare organizations.

At the heart of accountability lies the felt obligation of the accountor (the one rendering account) to inform the accountee (the one taking account) about performance, procedures, or outcome. The accountee must then able to question the accountor about information and conduct, and also to pass judgement about this conduct (Bovens, 2009). In the sepsis inspection, we saw that hospitals gave account both in the form of providing documentation and responses on request and on their own volition, as exemplified in study 3 by how staff and managers discussed challenges in one of the closing meetings. In such cases, the inspector can act out the role of a consultant (Kagan & Scholz, 1984) or a teacher (Kok et al., 2019).

There is a tension, though, between active improvement support and independent assessment that can lead to confusion regarding regulatory roles and diminish accountability (Furnival, 2017). This raises an interesting question about accountability in cases where the inspectors refrain from making judgements or recommendations altogether. Leistikow (2018) discusses the regulatory approach of "feedback without recommendations". The Dutch healthcare inspectorate had carried out a pilot project where they used a new tool for monitoring client experiences. They provided the results to the care institutions but did not issue any opinion. The institutions had then used the results to improve the quality of care, resulting in the hypothesis that feedback without recommendations might be an effective intervention. Leistikow explains the simple logic behind why this approach could work: "people are more enthusiastic about working on improvements they themselves have devised than on externally imposed measures" (p. 22).

It is not necessarily so that the healthcare organizations do not feel accountable to the inspection authorities that adopt a learning-based approach or refrain from making judgements. Accountability can function according to a "logic of proximity", where the closeness between the accountor and the accountee influences the degree of felt accountability (Schillemans et al., 2020). In this regard, it is interesting to note that

Weenink et al. (2020), who investigated the impact of publishing inspection frameworks, found that such frameworks are used more actively in the improvement efforts if the ties between the inspectorate and the sector are close. In part, this mechanism seems to be tied to the "grip" the inspection authorities have on the sector. If it is likely that providers will be inspected, they will use the frameworks more actively, according to Weenink et al. I believe, however, that this logic of proximity is also related to the relational impact of the inspections. As reported in study 3, the clinicians and managers had expectations of inspection team's clinical insight. Accountability "presupposes agreement […] about what constitutes an acceptable performance" (Day & Klein, 1987, p. 5). When the healthcare organizations are cognizant of the inspection authorities' role and see them as credible and authoritative when it comes to assessing quality of care, the healthcare organizations will be more likely to take action on the basis of the results provided by the inspection authorities.

Even if the inspection authorities do not resort to enforcement, they have provided the healthcare organizations with a "template" for assessing the quality of care within a given area. To better understand this, we can turn to Goffman's (1986) concept of "keying", which refers to a systematic transformation of one activity into another. Through the process of keying, the meaning of the activity for the participants is changed but the activity is patterned on the original one. Quality improvement following an inspection, as well as improvement in advance of an inspection (anticipatory impact) or inspired by inspections in other organizations (lateral impact), can be understood as keying. The improvement can take the form of what Goffman calls *technical redoing*, i.e. "practicing" the inspection, or of *regrounding*, where the healthcare organization uses the same indicators and standards as the inspection authorities but they perform the assessment not with the motivation of complying with requirements but to improve quality. This, then, can give us an idea of how the interpretation schemes of the inspection authorities can be adopted, internalized, and used by the healthcare organizations, and of how the administrative and professional accountability can align.

7. Concluding remarks

Inspection is a governance tool with a practical end goal, and this thesis is meant to provide insights that can be of practical value. This chapter therefore starts out with considering the possible contributions to inspection practice that our research can offer. As much as I would like to provide failsafe guidance to inspection authorities, this thesis is unfortunately not a panacea for succeeding with inspections. I nonetheless point to some areas that could be worthy of the attention of those involved in governing and developing healthcare inspections. I then go on to discuss some themes and approaches that I believe could be worth pursuing in future research efforts in this field.

7.1 Implications for the work with inspections

Even though regulators perhaps spend more time concerned with setting minimum standards than driving improvement (Walshe, 2011), there has been an international trend where states have taken stronger regulatory roles in promoting quality improvement in healthcare organizations (Healy, 2011). So too has the Norwegian state. If the present thesis can provide any valuable input for the governance and regulation of health care, two questions seem especially pertinent: Is the use of inspections a viable solution for improving quality of care, and, if so, what can be done to maximize the effects of inspections in this regard?

As described in the previous chapters, research has – with some exceptions (including findings from the present studies) – painted a somewhat bleak picture of the inspections' contribution to quality improvement. This has led some researchers to doubt the efficiency of inspections as a tool for improvement (Castro, 2018). Defenders of inspections have a last resort in what I in chapter 6 called "speed camera justification" of regulation. The mere knowledge that the inspection authorities exist and might come knocking one day increases the accountability of the health organizations and, hence,

the quality of care they provide. Therefore, it does not matter whether they actually contribute to improvement for the organizations that are being inspected.

In my opinion, we should not conclude on such a Panglossian note. The inspection authorities should strive to make an impact. And they do strive. Facing a continuous changing of circumstances, supervisory bodies are required to constantly adapt to new challenges and opportunities (van Oijen, Wallenburg, Bal, & Grit, 2020). As I hope this thesis has shown, inspections can in fact contribute to improving the quality of care for a group of patients suffering a life-threatening acute illness. This improvement process has not only taken place in a select few hospitals. It has happened across hospitals, resulting in an average improvement of important care processes for the whole patient group. Thus, the answer to the question of whether inspections are viable solution to quality improvement is in my opinion an emphatic *yes*.

The next question is then how the inspections should be designed and delivered to make sure that they foster quality improvement in the healthcare organizations. Here I want to highlight three important themes.

The first theme is related to the use of performance indicators. This was a main focus of study 3. Our findings suggest that it was important to identify indicators that could capture the care processes in a comprehensive way. The key to achieve this in the sepsis inspection was to include external clinical specialists at an early stage. The specialists had both research-based knowledge of sepsis and practical experience from treating patients with sepsis. As argued in study 3 and described in the previous chapter, the inclusion of such clinical specialists on the inspection teams further bolstered the legitimacy of the inspections. Including top notch clinical knowledge on the inspection team will also help the inspection with a common problem in healthcare regulation, namely integrating and weighing “hard” and “soft” data on performance (Kok et al., 2020; Walshe, 2011). The specialists could use their real-world experience and practical knowledge in helping identify areas that needed improving.

The second theme is the importance of addressing and engaging the healthcare personnel during inspections. The role of clinicians in bringing about positive quality

improvement following the inspections has been underlined in this research project. Any improvement of care must, by necessity, involve the personnel. Regulators should therefore make sure that the inspections address the employees' contribution to quality improvement directly during inspections. This echoes core concepts in Safety-II regarding the role of humans in bolstering safety and resilience in organizations, and it requires a regulatory focus on *inter alia* interdisciplinary teamwork and enabling professionals to voice quality concerns within their organizations (Leistikow & Bal, 2020).

Though it has not been a focus of this study, the importance of user participation, a companion theme to the importance of engaging healthcare personnel, is worth mentioning in this respect. Leistikow (2018, p. 8) reminds us that the "added value of governmental regulation of healthcare quality should always be assessed from the perspective of those who receive care." Yet, a study of regulatory practices in the UK found that among eight different dimensions of improvement capability, service-user focus, i.e. "the identification and meeting of current and emergent needs and expectations for service users", was the one that received the least attention in policy documents and interviews (Furnival, Boaden, & Walshe, 2018a). User participation in inspections has received more focus in recent scholarly work, though. This research has advocated the further development of new types of user involvement (Rutz, Bovenkamp, Buitendijk, Robben, & Bont, 2018; Wiig, Haraldseid-Driftland, Tvete Zachrisen, Hannisdal, & Schibevaag, 2019; Wiig et al., 2020). A channel for user participation that has become more relevant over the last years, is social media. There is a growing number of consumer-led online sources related to health, including patient rating sites targeted at sharing users' experiences with healthcare providers and healthcare personnel (Laukka, Rantakokko, & Suhonen, 2019). Inspection authorities need to decide whether to make use of data from such sites, like the Dutch healthcare inspectorate has done (Kool, Kleefstra, Borghans, Atsma, & van de Belt, 2016), and, should they decide to use them, devise ways to harness the data for regulatory purposes.

The third theme has to do with the overall inspection approach. There seems to be an agreement that in healthcare, the inspection authorities should choose some sort of responsive approach (Furnival, 2017; Leistikow, 2018; OECD, 2014; Schaefer & Wiig,

2017; Smithson et al., 2018). No one wants the inspection authorities to be enforcers of byzantine rules, and if the inspections are to foster improvement and organizational change, their approach needs to be sensitive to the challenges met by the inspected organization. This recommendation, however, will be found wanting for lack of actionable information. What should the overall priority be, between setting minimum standards and promoting quality in the sector as a whole (Leistikow, 2018)? Should one choose light controls, which can be ineffective and easy to game, or heavier controls that will capture more but are more costly to operate (Bouckaert & Pollitt, 2011)? Should the regulator strive to take the hybrid role of both doing inspections and providing improvement support, even if this entails a risk of undermining the inspection authorities' role as independent assessor (Furnival, 2017; Furnival et al., 2018b)? Or, to use an example from the sepsis inspection regarding the recommended time to antibiotics, should inspection authorities opt to be more dynamic and responsive to criticism towards criteria and indicators, or would this "setting their sails to every wind" approach lead to hopeless relativism where all inspection findings can and will be contested?

Regrettably, I cannot tell the inspection authorities how they should square the circle of these questions. What I can say, is that the proposal of Baldwin and Black (2008) about developing "really responsive" regulation makes intuitive sense to me. For regulation to be really responsive, they argue, regulators need to be responsive towards more than just how compliant the organization is. They need to respond to their attitudinal settings, i.e. the cognitive framework of the regulated organization. They also need to be responsive to the regulatory regime they themselves are a part of: They need to understand the institutional environment and the logic of the tools and strategies they employ, and they need to be responsive to their own performance and sensitive to how the challenges of regulation shift over time. I think one of the keys to success for the sepsis inspection was that they strived to fully understand the existing care processes of the emergency departments and to frame the inspection in a way that was meaningful to the inspected organizations. This is, I would argue, exactly what being responsive to the cognitive framework of the regulated organization is all about. To reach the next

step of "responsiveness", being responsive to the regulatory regime and their own performance, the regulator needs to learn from previous experience (Black & Baldwin, 2010). To achieve this, I think it is necessary to prioritize both a self-evaluative culture within the inspection authorities, and also to promote research of many stripes into regulatory regimes and the effects of inspections.

There is an important drawback to adjusting and improving the approach to inspection in the ways described above, namely cost. Including experts during the planning stages, engaging personnel and users, and providing a reflexive approach to inspection requires time and resources. If the inspection authorities are to operate within existing budgetary constraints, a change in inspection approach could mean having to scale back on the number of inspections carried out. In other words, doing fewer inspections in order to do better. This solution is feasible only if the inspection authorities are willing to and allowed to cut back on other activities. The inherent danger in such retrenchment is that the "speed camera" loses credibility followed by a weakened potential for instilling accountability in the healthcare organizations.

7.2 Future research

Some stones are left unturned. There is a need to more thoroughly investigate the mechanisms that link inspections and quality improvement. A natural progression of the work in the present thesis is to re-analyze the focus group interviews and expand this analysis to include all six hospitals and both pre and post inspection interviews. Gathering additional data from documents and including these together with the data from the electronic health records and interviews, would provide a rich material for a theory-focused, context-sensitive case study in the form of a realist evaluation (Pawson, 2013).

Another option is to broaden the scope. In chapter 4, I described how we utilized a "nested analysis" (Lieberman, 2005), where the large-N analysis (LNA) was followed by a small-N analysis (SNA). Studies 1 and 2 provided the initial LNAs, and we then

went on to explore the inspections in-depth using SNA in study 3. The next logical step in this approach is to go back to the LNA, using the insights we have gained through the three studies.

One theme flagged in our research that could be a candidate for a new LNA is anticipatory effects. To date there are no studies from the Norwegian context quantitatively measuring the effects of announcing an inspection. Another potential avenue of research is the effect of inspections in services outside the hospital-setting. In primary care there are fewer evidence-based standards and data on care processes are not always as readily available as they are from hospitals. Developing an LNA for inspection effects in primary care would perhaps require generating a set of performance indicators that can be collected independently from the inspection teams' data collection, akin to the method devised by Allen et al. (2019) and Allen et al. (2020), and use these indicators to assess inspection effects.

This touches on a point that is of more general relevance to the current state of inspection research: the importance of allowing for a multitude of methodological approaches to the study of inspections. There are many exciting methods that are being tested out, from approaches as disparate as discrete choice experiments (van Dijk et al., 2020) and ethnography (de Bree & Stoopendaal, 2018). I want to call special attention to the merits of comparative designs. There is a story, integral to each and one inspection, of how the inspection succeeded or failed in its ambition to promote quality improvement. If we refrain from paying attention to these different stories of improvement, and merely extrapolate our knowledge of one specific inspection (whether that knowledge is from an LNA or a SNA), we are left with very little understanding of how inspections work.

In trying to understand how inspections work, I think we also should focus intently on theory. Involved in the study of regulation in healthcare is a ragtag collection of disciplinary fields. We find influences from medicine, law, and a host of social science disciplines such as economics, political science, sociology, and psychology. To my mind, this is a strength that we should nourish. We should provide clear, theoretical

descriptions of how we think inspections work. This can provide us with what Pawson (2013) calls reusable conceptual platforms. Some frameworks and conceptual platforms have been provided, such as those of Smithson et al. (2018) and Hovlid, Braut, et al. (2020). Building further investigations off platforms such as these can bring research on healthcare regulation and inspections further.

An important reason for why I stress the need for comparative research, is that I think it is important to resist tendencies towards parochial outlooks on inspections. It is easy to turn a blind eye to the lessons from other countries and other systems, reasoning that our way is so particular that those lessons do not apply to us. This is what makes research efforts like those of Wiig et al. (2020) so important. Touting the blessings of having an international outlook may seem to contradict my emphasis of the importance of context in explaining and understanding inspections. After all, I have been constantly arguing that understanding the context and system within which inspections take place is paramount. My argument, however, is that our objective should not be wholesale import of quick and ready inspection recipes from one country to another. Rather, we should carefully analyze the inspections that take place in different systems, aiming to leverage the slight varieties in inspection schemes, contexts, historical trajectories, and outcomes to gain a deeper understanding of how inspections can be best adapted to different circumstances and aims.

References

Allen, T., Walshe, K., Proudlove, N., & Sutton, M. (2019). Measurement and improvement of emergency department performance through inspection and rating: an observational study of emergency departments in acute hospitals in England. *Emergency Medicine Journal, 36*(6), 326. doi:10.1136/emermed-2018-207941

Allen, T., Walshe, K., Proudlove, N., & Sutton, M. (2020). The measurement and improvement of maternity service performance through inspection and rating: an observational study of maternity services in acute hospitals in England. *Health Policy*. doi:10.1016/j.healthpol.2020.08.007

Allison, P. D. (2001). *Missing Data*. Thousand Oaks, CA: Sage.

Allison, P. D. (2009). Missing Data. In R. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE Handbook of Quantitative Methods in Psychology*. doi:10.4135/9780857020994

Arianson, H., Elvbakken, K., & Malterud, K. (2008). Hvordan opplevde helsepersonell tilsynet med fødeinstitusjoner? [How did the healthcare personnel experience the inspection of maternity institutions?]. *Tidsskrift for Den norske legeforening, 128*(10), 1179-1181.

Aristotle. (1910). *On the Generation of Animals*. Retrieved from https://web.archive.org/web/20040628121253/http://etext.library.adelaide.edu.au/a/a8ga/

Ayres, I., & Braithwaite, J. (1992). *Responsive regulation: Transcending the deregulation debate*. Oxford: Oxford University Press.

Baio, G., Copas, A., Ambler, G., Hargreaves, J., Beard, E., & Omar, R. Z. (2015). Sample size calculation for a stepped wedge trial. *Trials, 16*, 354-354. doi:10.1186/s13063-015-0840-9

Baldwin, R., & Black, J. (2008). Really Responsive Regulation. *The Modern Law Review, 71*(1), 59-94. doi:10.1111/j.1468-2230.2008.00681.x

Baldwin, R., Cave, M., & Lodge, M. (2011). *Understanding regulation: theory, strategy, and practice* (2nd ed.). Oxford: Oxford University Press.

Bardsley, M. (2017). Learning how to make routinely available data useful in guiding regulatory oversight of hospital care. *BMJ Quality & Safety, 26*(2), 90-92. doi:10.1136/bmjqs-2016-005311

Barker, D., McElduff, P., D'Este, C., & Campbell, M. J. (2016). Stepped wedge cluster randomised trials: a review of the statistical methodology used and available. *BMC Medical Research Methodology, 16*, 69-69. doi:10.1186/s12874-016-0176-5

Barochia, A. V., Cui, X., Vitberg, D., Suffredini, A. F., O'Grady, N. P., Banks, S. M., . . . Eichacker, P. Q. (2010). Bundled care for septic shock: an analysis of clinical trials. *Critical Care Medicine, 38*(2), 668-678. doi:10.1097/CCM.0b013e3181cb0ddf

Baskind, R., Kordowicz, M., & Chaplin, R. (2010). How does an accreditation programme drive improvement on acute inpatient mental health wards? An exploration of members' views. *J Ment Health, 19*(5), 405-411. doi:10.3109/09638230903531118

Batalden, P. B., & Davidoff, F. (2007). What is "quality improvement" and how can it transform healthcare? *Quality and Safety in Health Care, 16*(1), 2-3. doi:10.1136/qshc.2006.022046

Benson, L. A., Boyd, A., & Walshe, K. (2006). Learning from regulatory interventions in healthcare; The Commission for Health Improvement and its clinical governance review process. *Clinical Governance: An International Journal, 11*(3), 213-224. doi:10.1108/14777270610683146

Berwick, D. (1998). Crossing the boundary: changing mental models in the service of improvement. *International Journal for Quality in Health Care, 10*(5), 435-441. doi:10.1093/intqhc/10.5.435

Bevan, G. (2010). Performance Measurement of "Knights" and "Knaves": Differences in Approaches and Impacts in British Countries after Devolution. *Journal of Comparative Policy Analysis: Research and Practice, 12*(1-2), 33-56. doi:10.1080/13876980903076187

Bevan, G., & Hood, C. (2006). What's Measured Is What Matters: Targets and Gaming in the English Public Health Care System. *Public Administration, 84*(3), 517-538. doi:10.1111/j.1467-9299.2006.00600.x

Black, J., & Baldwin, R. (2010). Really responsive risk-based regulation. *Law Policy, 32*. doi:10.1111/j.1467-9930.2010.00318.x

Bogh, S. B., Falstie-Jensen, A. M., Hollnagel, E., Holst, R., Braithwaite, J., & Johnsen, S. P. (2017). Improvement in quality of hospital care during accreditation: A nationwide stepped-wedge study. *International Journal for Quality in Health Care, 28*(6), 715-720. doi:10.1093/intqhc/mzw099

Bogh, S. B., Falstie-Jensen, A. M., Hollnagel, E., Holst, R., Braithwaite, J., Raben, D. C., & Johnsen, S. P. (2017). Predictors of the effectiveness of accreditation on hospital performance: A nationwide stepped-wedge study. *International Journal for Quality in Health Care, 29*(4), 477-483. doi:10.1093/intqhc/mzx052

Bone, R. C., Balk, R. A., Cerra, F. B., Dellinger, R. P., Fein, A. M., Knaus, W. A., . . . Sibbald, W. J. (1992). Definitions for Sepsis and Organ Failure and Guidelines for the Use of Innovative Therapies in Sepsis. *Chest, 101*(6), 1644-1655. doi:10.1378/chest.101.6.1644

Bouckaert, G., & Pollitt, C. (2011). *Public Management Reform. A Comparative Analysis–New Public Management, Governance and the Neo-Weberian State*: Oxford: Oxford University Press.

Bourdieu, P. (1977). *Outline of a theory of practice* (R. Nice, Trans.). Cambridge: Cambridge University Press.

Bovens, M. (2009). Public Accountability. In E. Ferlie, L. E. Lynn, & C. Pollitt (Eds.), *The Oxford Handbook of Public Management*. doi:10.1093/oxfordhb/9780199226443.003.0009

Boyd, A., Ross, S., Robertson, R., Walshe, K., & Smithson, R. (2018). How hospital survey teams function: An analysis of Care Quality Commission inspections of acute hospitals in England. *Journal of Health Organization and Management, 32*(2), 206-223. doi:10.1108/JHOM-07-2017-0175

Braithwaite, J., Greenfield, D., Westbrook, J., Pawsey, M., Westbrook, M., Gibberd, R., . . . Lancaster, J. (2010). Health service accreditation as a predictor of clinical and organisational performance: a blinded, random, stratified study. *Qual Saf Health Care, 19*(1), 14-21. doi:10.1136/qshc.2009.033928

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology, 3*(2), 77-101. doi:10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. Los Angeles, Calif: Sage.

Braut, G. S. (2019). Helsetilsynet i fylket [NBHS in the county]. *Store medisinske leksikon.* Retrieved from https://sml.snl.no/Helsetilsynet_i_fylket

Brennan, T. A. (1998). The role of regulation in quality improvement. *Milbank Quarterly, 76*(4), 709-731. doi:10.1111/1468-0009.00111

Brown, C. A., & Lilford, R. J. (2006). The stepped wedge trial design: A systematic review. *BMC Medical Research Methodology, 6*(1), 54. doi:10.1186/1471-2288-6-54

Brubakk, K., Vist, G. E., Bukholm, G., Barach, P., & Tjomsland, O. (2015). A systematic review of hospital accreditation: the challenges of measuring complex intervention effects. *BMC Health Services Research, 15*(1), 280. doi:10.1186/s12913-015-0933-x

Byrkjeflot, H., Christensen, T., & Lægreid, P. (2014). The Many Faces of Accountability: Comparing Reforms in Welfare, Hospitals and Migration. *Scandinavian Political Studies, 37*(2), 171-195. doi:10.1111/1467-9477.12019

Cameron, A. C., & Miller, D. L. (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources, 50*(2), 317-372. doi:10.3368/jhr.50.2.317

Campbell, S. M., Chauhan, U., & Lester, H. (2010). Primary Medical Care Provider Accreditation (PMCPA): pilot evaluation. *British Journal of General Practice, 60*(576), 295-304. doi:10.3399/bjgp10X514800

Castro-Avila, A., Bloor, K., & Thompson, C. (2019). The effect of external inspections on safety in acute hospitals in the National Health Service in England: A controlled interrupted time-series analysis. *Journal of Health Services Research and Policy, 24*(3), 182-190. doi:10.1177/1355819619837288

Castro, A. C. (2018). *Evaluating the effect of external inspections of the Care Quality Commission of acute NHS hospitals in England.* (PhD thesis). University of York, York.

Charlson, M. E., Pompei, P., Ales, K. L., & MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Diseases, 40*(5), 373-383. doi:10.1016/0021-9681(87)90171-8

Collopy, B. T. (2000). Clinical indicators in accreditation: an effective stimulus to improve patient care. *International Journal for Quality in Health Care, 12*(3), 211-216. doi:10.1093/intqhc/12.3.211

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Los Angeles: Sage.

Czura, C. J. (2011). "Merinoff symposium 2010: Sepsis" - Speaking with One Voice. *Molecular Medicine, 17*(1-2), 2-3. doi:10.2119/molmed.2010.00001.commentary

Dahler-Larsen, P. (2014). Constitutive Effects of Performance Indicators: Getting beyond unintended consequences. *Public Management Review, 16*(7), 969-986. doi:10.1080/14719037.2013.770058

Dahler-Larsen, P. (2019). *Quality: From Plato to Performance*. doi:10.1007/978-3-030-10392-7

Day, P., & Klein, R. (1987). *Accountabilities: Five public services*. London: Tavistock.

de Bree, M., & Stoopendaal, A. (2018). De- and Recoupling and Public Regulation. *Organization Studies, 41*(5), 599-620. doi:10.1177/0170840618800115

de Kam, D. (2020). *Through the Regulator's Eyes: On the effects of making quality and safety of care inspectable.* (PhD thesis). Erasmus University, Rotterdam.

de Kam, D., Kok, J., Grit, K., Leistikow, I., Vlemminx, M., & Bal, R. (2020). How incident reporting systems can stimulate social and participative learning: A mixed-methods study. *Health Policy, 124*(8), 834-841. doi:10.1016/j.healthpol.2020.05.018

Dellinger, R. P., Carlet, J. M., Masur, H., Gerlach, H., Calandra, T., Cohen, J., . . . Levy, M. M. (2004). Surviving Sepsis Campaign guidelines for management of severe sepsis and septic shock. *Intensive Care Medicine, 30*(4), 536-555. doi:10.1007/s00134-004-2210-z

Dellinger, R. P., Levy, M. M., Carlet, J. M., Bion, J., Parker, M. M., Jaeschke, R., . . . Beale, R. (2008). Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2008. *Intensive Care Medicine, 34*(1), 17-60. doi:10.1007/s00134-007-0934-2

Dellinger, R. P., Levy, M. M., Rhodes, A., Annane, D., Gerlach, H., Opal, S. M., . . . Moreno, R. (2013). Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock, 2012. *Intensive Care Medicine, 39*(2), 165-228. doi:10.1007/s00134-012-2769-8

Donabedian, A. (1988). The quality of care. How can it be assessed? *JAMA, 260*(12), 1743-1748. doi:10.1001/jama.1988.03410120089033

Downs, A. (1967). *Inside bureaucracy*. Boston: Little, Brown and Company.

Duckett, S. (1983). Changing hospitals: the role of hospital accreditation. *Social Science and Medicine, 17*(20), 1573-1579.

Emanuel, E. J., Wendler, D., & Grady, C. (2000). What Makes Clinical Research Ethical? *JAMA, 283*(20), 2701-2711. doi:10.1001/jama.283.20.2701

Ewert, B. (2020). Focusing on quality care rather than 'checking boxes': How to exit the labyrinth of multiple accountabilities in hybrid healthcare arrangements. *Public Administration, 98*(2), 308-324. doi:10.1111/padm.12556

Ferguson, N. (2017). *The square and the tower: Networks, hierarchies and the struggle for global power*: Penguin.

Fisher, E. S., & Welch, H. G. (1999). Avoiding the unintended consequences of growth in medical care: how might more be worse? *JAMA, 281*(5), 446-453. doi:10.1001/jama.281.5.446

Flodgren, G., Goncalves-Bradley, D. C., & Pomey, M. P. (2016). External inspection of compliance with standards for improved healthcare outcomes. *Cochrane Database of Systematic Reviews, 12*, Cd008992. doi:10.1002/14651858.CD008992.pub3

Flyvbjerg, B. (2001). *Making Social Science Matter: Why social inquiry fails and how it can succeed again*. Cambridge: Cambridge University Press.

Furnival, J. (2017). *Regulation for Improvement? A Study of How Improvement Capability is Conceptualised by Healthcare Regulatory Agencies in the United Kingdom.* (PhD thesis). The University of Manchester, Manchester.

Furnival, J., Boaden, R., & Walshe, K. (2018a). Assessing improvement capability in healthcare organisations: a qualitative study of healthcare regulatory agencies in the UK. *International Journal for Quality in Health Care*, 1-9. doi:10.1093/intqhc/mzy085

Furnival, J., Boaden, R., & Walshe, K. (2018b). Emerging Hybridity: A Comparative Analysis of Regulatory Arrangements in the Four Countries of the UK. In A. M. McDermott, M. Kitchener, & M. Exworthy (Eds.), *Managing Improvement in Healthcare: Attaining, Sustaining and Spreading Quality* (pp. 59-75). Cham: Springer International Publishing.

Gatewood, M. O. K., Wemple, M., Greco, S., Kritek, P. A., & Durvasula, R. (2015). A quality improvement project to improve early sepsis care in the emergency department. *BMJ Qual Saf, 24*(12), 787-795. doi:10.1136/bmjqs-2014-003552

George, A. L., & Bennett, A. (2005). *Case Studies and Theory Development in the Social Sciences*. Cambridge, Mass: MIT Press.

Giddens, A. (1995). *Politics, sociology and social theory: Encounters with classical and contemporary social thought*. Cambridge: Polity Press.

Gilad, S. (2010). It runs in the family: Meta-regulation and its siblings. *Regulation & Governance, 4*(4), 485-506. doi:10.1111/j.1748-5991.2010.01090.x

Girling, A. J., & Hemming, K. (2016). Statistical efficiency and optimal design for stepped cluster studies under linear mixed effects models. *Statistics in Medicine, 35*(13), 2149-2166. doi:10.1002/sim.6850

Goffman, E. (1986). *Frame analysis: An essay on the organization of experience*. Boston: Northeastern University Press.

Goodhart, C. (1981). Problems of Monetary Management: The U.K. Experience. In A. S. Courakis (Ed.), *Inflation, Depression, and Economic Policy in the West*. New Jersey: Barnes & Noble Books.

Goodwin, A., Srivastava, V., Shotton, H., Protopapa, K., Butt, A., & Mason, M. (2015). *Just say sepsis: A review of the process of care received by patients with sepsis*. Retrieved from https://www.ncepod.org.uk/2015report2/downloads/JustSaySepsis_FullReport.pdf

Greenfield, D., Pawsey, M., & Braithwaite, J. (2011). What motivates professionals to engage in the accreditation of healthcare organizations? *International Journal for Quality in Health Care, 23*(1), 8-14. doi:10.1093/intqhc/mzq069

Grimen, H. (2006). *Profesjon og profesjonsmoral [Profession and professional ethics]* (Arbeidsnotat nr. 04/2006). Retrieved from https://fagarkivet.oslomet.no/jspui/bitstream/20.500.12199/3081/4/4-2006%2C%20Grimen%2C%20Profesjon%20og%20moral.pdf

Gunningham, N. (2010). Enforcement and Compliance Strategies. In R. C. Baldwin, M. Cave, & M. Lodge (Eds.), *The Oxford Handbook of Regulation*. London: Oxford University Press.

Guo, G., & Zhao, H. (2000). Multilevel Modeling for Binary Data. *Annual Review of Sociology, 26*(1), 441-462. doi:10.1146/annurev.soc.26.1.441

Guttmann, A., Schull, M. J., Vermeulen, M. J., & Stukel, T. A. (2011). Association between waiting times and short term mortality and hospital admission after departure from emergency department: population based cohort study from Ontario, Canada. *BMJ, 342*, d2983. doi:10.1136/bmj.d2983

Haines, T. P., & Hemming, K. (2018). Stepped-wedge cluster-randomised trials: level of evidence, feasibility and reporting. *Journal of Physiotherapy, 64*(1), 63-66. doi:10.1016/j.jphys.2017.11.008

Healy, J. (2011). *Improving Health Care Safety and Quality: Reluctant Regulators*. Farnham Surrey: Ashgate.

Healy, J., & Braithwaite, J. (2006). Designing safer health care through responsive regulation. *Medical Journal of Australia, 184*(10), S56-S59. doi:10.5694/j.1326-5377.2006.tb00364.x

Helgesen, M. K., & Hanssen, G. S. (2014). Kommunalt handlingsrom etter Samhandlingsreformen [The municipality as negotiator in the Norwegian Coordination Reform]. *Tidsskrift for velferdsforskning, 17*(3), 41-53.

Heller, J. (2011). *Catch 22*. London: Vintage Books.

Hemming, K., & Girling, A. (2014). A menu-driven facility for power and detectable-difference calculations in stepped-wedge cluster-randomized trials. *Stata Journal, 14*(2), 363-380. doi:10.1177/1536867x1401400208

Hemming, K., Haines, T. P., Chilton, P. J., Girling, A. J., & Lilford, R. J. (2015). The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ, 350*. doi:10.1136/bmj.h391

Hemming, K., Lilford, R., & Girling, A. J. (2015). Stepped-wedge cluster randomised controlled trials: a generic framework including parallel and multiple-level designs. *Statistics in Medicine, 34*(2), 181-196. doi:10.1002/sim.6325

Hemming, K., Taljaard, M., & Forbes, A. (2017). Analysis of cluster randomised stepped wedge trials with repeated cross-sectional samples. *Trials, 18*(1). doi:10.1186/s13063-017-1833-7

Hilarion, P., Suñol, R., Groene, O., Vallejo, P., Herrera, E., & Saura, R. M. (2009). Making performance indicators work: the experience of using consensus indicators for external assessment of health and social services at regional level in Spain. *Health Policy, 90*(1), 94-103. doi:10.1016/j.healthpol.2008.08.002

Hinchcliff, R., Greenfield, D., Moldovan, M., Westbrook, J. I., Pawsey, M., & Mumford, V. (2012). Narrative synthesis of health service accreditation literature. *BMJ Qual Saf, 21*. doi:10.1136/bmjqs-2012-000852

Hood, C. (2011). *The Blame Game: Spin, Bureaucracy, and Self-Preservation in Government*. Princeton: Princeton University Press.

Hood, C., Rothstein, H., & Baldwin, R. (2001). *The government of risk: Understanding risk regulation regimes*. Oxford: Oxford University Press.

Hood, C., Scott, C., James, O., Jones, G., & Travers, T. (1999). *Regulation inside government: Waste watchers, quality police, and sleaze-busters*. Oxford: Oxford University Press.

Houston, C., Sanchez, L. D., Fischer, C., Volz, K., & Wolfe, R. (2015). Waiting for Triage: Unmeasured Time in Patient Flow. *Western Journal of Emergency Medicine, 16*(1), 39-42. doi:10.5811/westjem.2014.11.22824

Hovlid, E., Braut, G. S., Hannisdal, E., Walshe, K., Bukve, O., Flottorp, S., . . . Frich, J. C. (2020). Mediators of change in healthcare organisations subject to external assessment: a systematic review with narrative synthesis. *BMJ Open*. doi:10.1136/bmjopen-2020-038850

Hovlid, E., Frich, J. C., Walshe, K., Nilsen, R. M., Flaatten, H. K., Braut, G. S., . . . Harthug, S. (2017). Effects of external inspection on sepsis detection and treatment: a study protocol for a quasiexperimental study with a stepped-wedge design. *BMJ Open, 7*(9), e016213. doi:10.1136/bmjopen-2017-016213

Hovlid, E., Høifødt, H., Smedbråten, B., & Braut, G. S. (2015). A retrospective review of how nonconformities are expressed and finalized in external inspections of health-care facilities. *BMC Health Services Research, 15*(1), 405. doi:10.1186/s12913-015-1068-9

Hovlid, E., Teig, I. L., Halvorsen, K., & Frich, J. C. (2020). Inspecting teams' and organisations' expectations regarding external inspections in health care: a qualitative study. *BMC Health Services Research, 20*(1), 627. doi:10.1186/s12913-020-05475-0

Hupe, P., & Hill, M. (2007). Street-level bureaucracy and public accountability. *Public Administration, 85*(2), 279-299. doi:10.1111/j.1467-9299.2007.00650.x

Hussey, M. A., & Hughes, J. P. (2007). Design and analysis of stepped wedge cluster randomized trials. *Contemporary Clinical Trials, 28*(2), 182-191. doi:10.1016/j.cct.2006.05.007

IDSA Sepsis Task Force. (2017). Infectious Diseases Society of America (IDSA) POSITION STATEMENT: Why IDSA Did Not Endorse the Surviving Sepsis Campaign Guidelines. *Clinical Infectious Diseases, 66*(10), 1631-1635. doi:10.1093/cid/cix997

Inspection. (2020). *Online Etymology Dictionary.* Retrieved from https://www.etymonline.com/word/inspection

International Organization for Standardization. (2012). Conformity assessment -- Requirements and recommendations for content of a third-party audit report on management systems (ISO/IEC TS 17022:2012). In.

Johannesen, D. T. S., & Wiig, S. (2017). Why adopt ISO 9001 certification in hospitals? A case study of external triggers and sensemaking in an emergency department in Norway. *Safety in Health, 3*(1), 7. doi:10.1186/s40886-017-0058-5

Kagan, R. A., & Scholz, J. T. (1984). "The Criminology of the Corporation" and Regulatory Enforcement Strategies. In K. Hawkins & J. M. Thomas (Eds.), *Enforcing Regulation* (pp. 67-95). Boston, M.A.: Kluwer-Nijhoff.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica, 47*(2), 263-291. doi:10.2307/1914185

Keegan, J. (1976). *The face of battle*. London: Jonathan Cape.

King, N., & Brooks, J. (2018). Thematic Analysis in Organisational Research. In C. Cassell, A. L. Cunliffe, & G. Grandy (Eds.), *The SAGE Handbook of Qualitative Business and Management Research Methods: Methods and Challenges*. doi:10.4135/9781526430236.n14

Knoop, S. T., Skrede, S., Langeland, N., & Flaatten, H. K. (2017). Epidemiology and impact on all-cause mortality of sepsis in Norwegian hospitals: A national retrospective study. *PloS One, 12*(11), e0187990. doi:10.1371/journal.pone.0187990

Kok, J., Leistikow, I., & Bal, R. (2019). Pedagogy of regulation: Strategies and instruments to supervise learning from adverse events. *Regulation & Governance, 13*(4), 470-487. doi:10.1111/rego.12242

Kok, J., Wallenburg, I., Leistikow, I., & Bal, R. (2020). "The doctor was rude, the toilets are dirty." Utilizing 'soft signals' in the regulation of patient safety. *Safety Science, 131*, 104914. doi:10.1016/j.ssci.2020.104914

Konisky, D. M., & Teodoro, M. P. (2016). When Governments Regulate Governments. *American journal of political science, 60*(3), 559-574. doi:10.1111/ajps.12221

Kool, R. B., Kleefstra, S. M., Borghans, I., Atsma, F., & van de Belt, T. H. (2016). Influence of Intensified Supervision by Health Care Inspectorates on Online Patient Ratings of Hospitals: A Multilevel Study of More Than 43,000 Online Ratings. *Journal of Medical Internet Research, 18*(7), e198. doi:10.2196/jmir.5884

Koop, C., & Lodge, M. (2017). What is regulation? An interdisciplinary concept analysis. *Regulation & Governance, 11*(1), 95-108. doi:10.1111/rego.12094

Kotz, D., Spigt, M., Arts, I. C. W., Crutzen, R., & Viechtbauer, W. (2012). Use of the stepped wedge design cannot be recommended: A critical appraisal and comparison with the classic cluster randomized controlled trial design. *Journal of Clinical Epidemiology, 65*(12), 1249-1252. doi:10.1016/j.jclinepi.2012.06.004

Kringen, J., Lindøe, P. H., & Braut, G. S. (2015). Innledning [Introduction]. In P. H. Lindøe, J. Kringen, & G. S. Braut (Eds.), *Risiko og tilsyn: Risikostyring og rettslig regulering* (2 ed.). Oslo: Universitetsforlaget.

Kumar, E. A., Roberts, E. D., Wood, E. K., Light, E. B., Parrillo, E. J., Sharma, E. S., . . . Cheang, E. M. (2006). Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Critical Care Medicine, 34*(6), 1589-1596. doi:10.1097/01.CCM.0000217961.75225.E9

Latour, B., & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts* (2nd ed.). Princeton: Princeton University Press.

Laukka, E., Rantakokko, P., & Suhonen, M. (2019). Consumer-led health-related online sources and their impact on consumers: An integrative review of the literature. *Health Informatics Journal, 25*(2), 247-266. doi:10.1177/1460458217704254

Lawrie, J., Carlin, J. B., & Forbes, A. B. (2015). Optimal stepped wedge designs. *Statistics & Probability Letters, 99*, 210-214. doi:10.1016/j.spl.2015.01.024

Leistikow, I. (2018). *The proof of the pudding: The value of governmental regulation of healthcare quality and safety*. Rotterdam: Erasmus School of Health Policy & Management.

Leistikow, I., & Bal, R. A. (2020). Resilience and regulation, an odd couple? Consequences of Safety-II on governmental regulation of healthcare quality. *BMJ Quality & Safety*, bmjqs-2019-010610. doi:10.1136/bmjqs-2019-010610

Levy, M. M., Evans, L. E., & Rhodes, A. (2018). The Surviving Sepsis Campaign Bundle: 2018 update. *Intensive Care Medicine, 44*(6), 925-928. doi:10.1007/s00134-018-5085-0

Levy, M. M., Fink, M. P., Marshall, J. C., Abraham, E., Angus, D., Cook, D., . . . Ramsay, G. (2003). 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Intensive Care Medicine, 29*(4), 530-538. doi:10.1007/s00134-003-1662-x

Levy, M. M., Pronovost, P. J., Dellinger, R. P., Townsend, S., Resar, R. K., Clemmer, T. P., & Ramsay, G. (2004). Sepsis change bundles: converting guidelines into meaningful change in behavior and clinical outcome. *Critical Care Medicine, 32*(11 Suppl), S595-597. doi:10.1097/01.CCM.0000147016.53607.C4

Levy, M. M., Rhodes, A., Phillips, G. S., Townsend, S. R., Schorr, C. A., Beale, R., . . . Dellinger, R. P. (2015). Surviving Sepsis Campaign: association between performance metrics and outcomes in a 7.5-year study. *Critical Care Medicine, 43*(1), 3-12. doi:10.1097/ccm.0000000000000723

Lewis-Beck, C., & Lewis-Beck, M. (2016). *Applied Regression: An Introduction* (2nd ed.). doi:10.4135/9781483396774

Lieberman, E. S. (2005). Nested Analysis as a Mixed-Method Strategy for Comparative Research. *American Political Science Review, 99*(3), 435-452. doi:10.1017/S0003055405051762

Lilford, R. J., & Jackson, J. (1995). Equipoise and the ethics of randomization. *Journal of the Royal Society of Medicine, 88*(10), 552-559. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1295354/

Lipsky, M. (2010). *Street-level bureaucracy: Dilemmas of the individual in public services* (30th Anniversary ed.). New York: Russell Sage Foundation.

Liu, V. X., Fielding-Singh, V., Greene, J. D., Baker, J. M., Iwashyna, T. J., Bhattacharya, J., & Escobar, G. J. (2017). The Timing of Early Antibiotics and Hospital Mortality in Sepsis. *American Journal of Respiratory and Critical Care Medicine, 196*(7), 856-863. doi:10.1164/rccm.201609-1848OC

Lægreid, P., Roness, P. G., & Rubecksen, K. (2008). Controlling Regulatory Agencies. *Scandinavian Political Studies, 31*(1), 1-26. doi:10.1111/j.1467-9477.2008.00193.x

Majno, G. (1991). The ancient riddle of sigma eta psi iota sigma (sepsis). *Journal of Infectious Diseases, 163*(5), 937-945. doi:10.1093/infdis/163.5.937

Majone, G. (1994). The rise of the regulatory state in Europe. *West European Politics, 17*(3), 77-101. doi:10.1080/01402389408425031

Makkai, T., & Braithwaite, J. (1992). In and Out of the Revolving Door: Making Sense of Regulatory Capture. *J. Pub. Pol, 12*(1), 61-78. doi:10.1017/S0143814X00005006

Marik, P. E., Farkas, J. D., Spiegel, R., & Weingart, S. (2019). POINT: Should the Surviving Sepsis Campaign Guidelines Be Retired? Yes. *Chest, 155*(1), 12-14. doi:10.1016/j.chest.2018.10.008

Mayntz, R. (2004). Mechanisms in the Analysis of Social Macro-Phenomena. *Philosophy of the Social Sciences, 34*(2), 237-259. doi:10.1177/0048393103262552

McWhinney, I. R. (1989). 'An acquaintance with particulars...'. *Family Medicine, 21*(4), 296.

Meyer, J. W., & Rowan, B. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology, 83*(2), 340-363. doi:10.1086/226550

Ministry of Government Administration and Labour. (2003). *Om statlige tilsyn [On government supervision]*. (St.meld. nr. 17). Retrieved from https://www.regjeringen.no/no/dokumenter/stmeld-nr-17-2002-2003-/id134860/?ch=1

Ministry of Health. (2002). *Om lov om endringer i lov 30. mars 1984 nr. 15 om statlig tilsyn med helsetjenesten og i enkelte andre lover [On (..) changes in Act on government supervision of healthcare services and some other acts]*. (Ot.prp. nr. 105. (2001-2002)). Retrieved from https://www.regjeringen.no/contentassets/259095b00a4b4da786a56a55201174f5/no/pdfa/otp200120020105000dddpdfa.pdf

Ministry of Health and Care Services. (2019). *Prop. 1 S*. (2019 – 2020). Retrieved from https://www.regjeringen.no/no/dokumenter/stmeld-nr-17-2002-2003-/id134860/?ch=1

Morr, M., Lukasz, A., Rübig, E., Pavenstädt, H., & Kümpers, P. (2017). Sepsis recognition in the emergency department – impact on quality of care and outcome? *BMC Emergency Medicine, 17*(1), 11. doi:10.1186/s12873-017-0122-9

Morse, J. M. (1994). Going in "Blind". *Qualitative Health Research, 4*(1), 3-5. doi:10.1177/104973239400400101

Moses, J., & Knutsen, T. (2012). *Ways of knowing: Competing methodologies in social and political research* (2nd ed.). New York: Palgrave Macmillan.

National Institute for Health and Care Excellence. (2016). *Sepsis: recognition, diagnosis and early management (NICE guideline 51)*. Retrieved from https://www.nice.org.uk/guidance/ng51

Nguyen, H. B., Rivers, E. P., Abrahamian, F. M., Moran, G. J., Abraham, E., Trzeciak, S., . . . Talan, D. A. (2006). Severe Sepsis and Septic Shock: Review of the Literature and Emergency Department Management Guidelines. *Annals of Emergency Medicine, 48*(1), 54.e51. doi:10.1016/j.annemergmed.2006.02.015

Noah, Z. L. (2014). The History of Sepsis Management Over the Last 30 Years. *Clinical Pediatric Emergency Medicine, 15*(2), 116-119. doi:10.1016/j.cpem.2014.05.007

Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. *Organization Science, 5*(1), 14-37. doi:10.1287/orsc.5.1.14

Norwegian Board of Health Supervision. (2018). *Sepsis – ingen tid å miste [Sepsis – no time to lose]*. (Rapport fra Helsetilsynet 1/2018). Retrieved from https://www.helsetilsynet.no/globalassets/opplastinger/publikasjoner/rapporter2018/helsetilsynetrapport1_2018.pdf

Norwegian Board of Health Supervision. (2019). Introduction to the supervisory authorities and the supervision of child welfare services, social services and health and care services in Norway. Retrieved from https://www.helsetilsynet.no/en/introduction-to-the-supervisory-authorities-and-the-supervision-of-child-welfare-services-social-services-and-health-and-care-services-in-norway

Norwegian Board of Health Supervision. (2020a). Organisasjonskart [Organizational chart]. Retrieved from https://www.helsetilsynet.no/om-oss/organisasjonskart/

Norwegian Board of Health Supervision. (2020b, 2020/12/02/). Tilsynsrapporter – søkeside [Inspection reports - search page]. Retrieved from https://www.helsetilsynet.no/tilsyn/tilsynsrapporter/

Norwegian Directorate of Health. (2018). 2. Sepsis. In *Antibiotika i sykehus*. Retrieved from https://www.helsedirektoratet.no/retningslinjer/antibiotika-i-sykehus/sepsis

OECD. (2014). *Regulatory enforcement and inspections, OECD Best Practice Principles for Regulatory Policy*. doi:10.1787/9789264208117-en

Patel, J. J., & Bergl, P. A. (2019). COUNTERPOINT: Should Broad-Spectrum Antibiotics Be Routinely Administered to All Patients With Sepsis as Soon as Possible? No. *Chest, 156*(4), 647-649. doi:10.1016/j.chest.2019.05.031

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health Services Research, 34*(5 Pt 2), 1189. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1089059/pdf/hsresearch00022-0112.pdf

Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). Thousand Oaks, California: SAGE Publications, Inc.

Pawson, R. (2013). *The Science of Evaluation: A Realist Manifesto*. London: SAGE Publications.

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). Realist review - a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research and Policy, 10*(Suppl 1), 21-34. doi:10.1258/1355819054308530

Plsek, P. E., & Greenhalgh, T. (2001). Complexity science - The challenge of complexity in health care. *British Medical Journal, 323*, 625-628. doi:10.1136/bmj.323.7313.625

Pollitt, C. (2006). Discipline and punish – or trust? Contrasting bases for performance management in executive agencies. In T. Christensen & P. Lægreid (Eds.), *Autonomy and regulation: Coping with agencies in the modern state* (pp. 301-327). Cheltenham: Edward Elgar.

Power, M. (1994). *The audit explosion*: Demos.

Power, M. (1997). *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.

Power, M. (2000). The audit society—Second thoughts. *International Journal of Auditing, 4*(1), 111-119. doi:10.1111/1099-1123.00306

Power, M. (2007). *Organized Uncertainty: Designing a World of Risk Management*. Oxford: Oxford University Press.

Puskarich, M. A., Trzeciak, S., Shapiro, N. I., Arnold, R. C., Horton, J. M., Studnek, J. R., . . . Emergency Medicine Shock Research Network (EMSHOCKNET). (2011). Association between timing of antibiotic administration and mortality from septic shock in patients treated with a quantitative resuscitation protocol. *Critical Care Medicine, 39*(9), 2066-2071. doi:10.1097/CCM.0b013e31821e87ab

Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and longitudinal modeling using Stata* (2nd ed.). College Station: Stata Press.

Regulation on management and quality improvement. (2016). *Forskrift om ledelse og kvalitetsforbedring i helse- og omsorgstjenesten [Regulation on management and quality improvement in the healthcare services]*. (FOR-2016-10-28-1250). Retrieved from https://lovdata.no/dokument/SF/forskrift/2016-10-28-1250

Reiss, A., Jr. (1984). Selecting Strategies of Social Control Over Organizational Life. In K. Hawkins & J. M. Thomas (Eds.), *Enforcing Regulation* (pp. 23-35). Boston, M.A.: Kluwer-Nijhoff.

Rhodes, A., Evans, L. E., Alhazzani, W., Levy, M. M., Antonelli, M., Ferrer, R., . . . Dellinger, R. P. (2017). Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. *Intensive Care Medicine, 43*(3), 304-377. doi:10.1007/s00134-017-4683-6

Rivers, E., Nguyen , B., Havstad , S., Ressler , J., Muzzin , A., Knoblich , B., . . . Tomlanovich , M. (2001). Early Goal-Directed Therapy in the Treatment of Severe Sepsis and Septic Shock. *New England Journal of Medicine, 345*(19), 1368-1377. doi:10.1056/NEJMoa010307

Rose, N., & Miller, P. (1992). Political power beyond the state: Problematics of government. *British Journal of Sociology*, 173-205.

Rudd, K. E., Johnson, S. C., Agesa, K. M., Shackelford, K. A., Tsoi, D., Kievlan, D. R., . . . Naghavi, M. (2020). Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study. *Lancet, 395*(10219), 200-211. doi:10.1016/S0140-6736(19)32989-7

Rutz, S., Bovenkamp, H., Buitendijk, S., Robben, P., & Bont, A. (2018). Inspectors' responses to adolescents' assessment of quality of care: A case study on involving adolescents in inspections. *BMC Health Services Research, 18*, 226. doi:10.1186/s12913-018-2998-9

Salmon, J., Heavens, J., Lombard, C., & Tavrow, P. (2003). *The impact of accreditation on the quality of hospital care: KwaZulu-Natal Province, Republic of South Africa*. Bethesda, MD: University Research Co.

Sayer, A. (2000). *Realism and Social Science*. Thousand Oaks: Sage.

Schaefer, C., & Wiig, S. (2017). Strategy and practise of external inspection in healthcare services – a Norwegian comparative case study. *Safety in Health, 3*(1), 1-9. doi:10.1186/s40886-017-0054-9

Schein, E. H., & Schein, P. A. (2016). *Organizational Culture and Leadership*. New York: John Wiley & Sons, Incorporated.

Schillemans, T., Overman, S., Fawcett, P., Flinders, M., Fredriksson, M., Laegreid, P., . . . Wood, M. (2020). Understanding Felt Accountability The institutional antecedents of the felt accountability of agency-CEO's to central government. *Governance*. doi:10.1111/gove.12547

Scholz, J. T. (1991). Cooperative Regulatory Enforcement and the Politics of Administrative Effectiveness. *The American Political Science Review, 85*(1), 115-136. doi:10.2307/1962881

Scott, C. (2010). Standard-Setting in Regulatory Regimes. In R. Baldwin, M. Cave, & M. Lodge (Eds.), *The Oxford Handbook of Regulation*. London: Oxford University Press.

Selznick, P. (1985). Focusing Organizational Research on Regulation. In R. G. Noll (Ed.), *Regulatory Policy and the Social Sciences* (pp. 363 - 367). Berkely: University of California Press.

Seymour, C. W., Gesten, F., Prescott, H. C., Friedrich, M. E., Iwashyna, T. J., Phillips, G. S., . . . Levy, M. M. (2017). Time to Treatment and Mortality during Mandated Emergency Care for Sepsis. *New England Journal of Medicine, 376*(23), 2235-2244. doi:10.1056/NEJMoa1703058

Shaw, C. D. (2000). External quality mechanisms for health care: summary of the ExPeRT project on visitatie, accreditation, EFQM and ISO assessment in European Union countries. *International Journal for Quality in Health Care, 12*. doi:10.1093/intqhc/12.3.169

Shaw, C. D., Groene, O., & Berger, E. (2019). External institutional strategies: accreditation, certification, supervision. In R. Busse, N. Klazinga, D. Panteli, & W. Quentin (Eds.), *Improving healthcare quality in Europe* (pp. 203-231): World Health Organization.

Silverman, D. (2007). *A very short, fairly interesting and reasonably cheap book about qualitative research*. Los Angeles: Sage.

Singer, M., Deutschman, C. S., Seymour, C. W., & et al. (2016). The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA, 315*(8), 801-810. doi:10.1001/jama.2016.0287

Singer, M., Inada-Kim, M., & Shankar-Hari, M. (2019). Sepsis hysteria: excess hype and unrealistic expectations. *The Lancet, 394*(10208), 1513-1514. doi:10.1016/S0140-6736(19)32483-3

Slade, E., Tamber, P. S., & Vincent, J.-L. (2003). The Surviving Sepsis Campaign: raising awareness to reduce mortality. *Critical Care, 7*(1), 1. doi:10.1186/cc1876

Smithson, R., Richardson, E., Roberts, J., Walshe, K., Wenzel, L., Robertson, R., . . . Proudlove, N. (2018). *Impact of the Care Quality Commission on provider performance: Room for improvement?* Retrieved from https://www.kingsfund.org.uk/publications/impact-cqc-provider-performance

Sollecito, W. A., & Johnson, J. K. (2019). The Global Evolution of Continuous Quality Improvement: From Janpanese Manufacturing to Global Health Services. In W. A. Sollecito & J. K. Johnson (Eds.), *Mclaughlin and Kaluzny's continuous quality improvement in health care* (5th ed., pp. 1-31). Burlington, MA: Jones & Bartlett Learning.

Sommerfeldt, P. A., Rognsvåg, S., & Anthun, M. (2020, 2020/12/04/). Tolv tidlegare overlegar går ut mot helseleiinga. Fylkeslegen ventar at Bergen ryddar opp [Twelve former chief physicians take the health management to task. The County Medical Officer expects that Bergen turns things around]. *NRK*. Retrieved from https://www.nrk.no/vestland/tolv-tidlegare-overlegar-i-bergen-gar-ut-mot-helseetaten-i-felles-uttale-1.15273747

Sterling, S. A., Miller, W. R., Pryor, J., Puskarich, M. A., & Jones, A. E. (2015). The Impact of Timing of Antibiotics on Outcomes in Severe Sepsis and Septic Shock: A Systematic Review and Meta-Analysis. *Critical Care Medicine, 43*(9), 1907-1915. doi:10.1097/ccm.0000000000001142

Stoneking, L. R., Winkler, J. P., DeLuca, L. A., Stolz, U., Stutz, A., Luman, J. C., . . . Denninghoff, K. R. (2015). Physician documentation of sepsis syndrome is associated with more aggressive treatment. *Western Journal of Emergency Medicine, 16*(3), 401-407. doi:10.5811/westjem.2015.3.25529

Strathern, M. (1997). 'Improving ratings': audit in the British University system. *European Review, 5*(3), 305-321. doi:10.1017/s1062798700002660

Sutherland, K., & Leatherman, S. (2006). *Regulation and quality improvement: A review of the evidence*. London: The Health Foundation.

The Gambia Hepatitis Study Group. (1987). The Gambia Hepatitis Intervention Study. *Cancer Research, 47*(21), 5782-5787.

The Health Personnel Act. (1999). *Lov om helsepersonell m.v. [Act relating to Health Personnel etc.]*. (LOV-1999-07-02-64). Retrieved from https://lovdata.no/dokument/NL/lov/1999-07-02-64

The ProCESS Investigators. (2014). A Randomized Trial of Protocol-Based Care for Early Septic Shock. *New England Journal of Medicine, 370*(18), 1683-1693. doi:10.1056/NEJMoa1401602

The Royal College of Emergency Medicine. (2017). *Severe Sepsis and Septic Shock* (Clinical Audit 2016/17). Retrieved from http://allcatsrgrey.org.uk/wp/wpfb-file/rcem-severe-sepsis-and-septic-shock-national-audit-report-pdf/

Thompson, J. A., Fielding, K., Hargreaves, J., & Copas, A. (2017). The optimal design of stepped wedge trials with equal allocation to sequences and a comparison to other trial designs. *Clinical Trials, 14*(6), 639-647. doi:10.1177/1740774517723921

Toffolutti, V., McKee, M., & Stuckler, D. (2017). Evidence points to 'gaming' at hospitals subject to national health service cleanliness inspections. *Health Affairs, 36*(2), 355-361. doi:10.1377/hlthaff.2016.1217

van Dijk, P., Schellings, R., Essers, B. A. B., Kessels, A. G., Leistikow, I., & Zeegers, M. P. (2020). A discrete choice experiment to identify the most efficient quality indicators for the supervision of psychiatric hospitals. *BMC Health Services Research, 20*(1), 192. doi:10.1186/s12913-020-4993-1

van Erp, J., Wallenburg, I., & Bal, R. (2020). Performance regulation in a networked healthcare system: From cosmetic to institutionalized compliance. *Public Administration, 98*(1), 46-61. doi:10.1111/padm.12518

van Oijen, J. C. F., Wallenburg, I., Bal, R., & Grit, K. J. (2020). Institutional work to maintain, repair, and improve the regulatory regime: How actors respond to external challenges in the public supervision of ongoing clinical trials in the Netherlands. *PloS One, 15*(7), e0236545. doi:10.1371/journal.pone.0236545

Vedung, E. (1997). *Public Policy and Program Evaluation*. New Brunswick, New Jersey: Transaction Publishers.

Vittinghoff, E., Glidden, D. V., Shiboski, S. C., & McCulloch, C. E. (2012). *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models* (2nd ed.). New York: Springer.

Walshe, K. (2003). *Regulating Healthcare: A Prescription for Improvement?* Berkshire: McGrawHill Education.

Walshe, K. (2007). Understanding what works—and why—in quality improvement: the need for theory-driven evaluation. *International Journal for Quality in Health Care, 19*(2), 57-59. doi:10.1093/intqhc/mzm004

Walshe, K. (2011). Managing performance. In K. Walshe & J. Smith (Eds.), *Healthcare Management* (2nd ed., pp. 515-535). Maidenhead: McGraw-Hill Education.

Walshe, K., & Boyd, A. (2007). *Designing regulation: A review*. Retrieved from https://www.escholar.manchester.ac.uk/uk-ac-man-scw:118421

Walshe, K., & Phipps, D. (2013). Developing a strategic framework to guide the Care Quality Commission's programme of evaluation. *London: CQC*.

Walshe, K., Wallace, L., Freeman, T., Latham, L., & Spurgeon, P. (2001). The external review of quality improvement in health care organizations: a qualitative study. *International Journal for Quality in Health Care, 13*(5), 367-374. doi:10.1093/intqhc/13.5.367

Weenink, J.-W., Wallenburg, I., Leistikow, I., & Bal, R. (2020). *Publication of inspection frameworks: a qualitative study exploring the impact on quality improvement and regulation in three healthcare settings. Forthcoming.*

Weick, K. E. (1993). The collapse of sensemaking in organizations: The Mann Gulch disaster. *Administrative Science Quarterly*, 628-652. doi:10.2307/2393339

Wertz, F. J., Charmaz, K., McMullen, L. M., Josselson, R., Anderson, R., & Emalinda, M. (2011). *Five Ways of Doing Qualitative Analysis* New York: Guilford Press.

Wesselink, S. F. O., Lingsma, H. F., Erasmus, V., Wentzel, H. R., Reulings, P. G. J., Robben, P. B. M., & Mackenbach, J. P. (2014). Does Government Supervision Improve Stop-Smoking Counseling in Midwifery Practices? *Nicotine & Tobacco Research, 17*(5), 572-579. doi:10.1093/ntr/ntu190

Wesselink, S. F. O., Lingsma, H. F., Ketelaars, C. A., Mackenbach, J. P., & Robben, P. B. (2015). Effects of Government Supervision on Quality of Integrated Diabetes Care: A Cluster Randomized Controlled Trial. *Medical Care, 53*(9), 784-791. doi:10.1097/Mlr.0000000000000399

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine, 30*(4), 377-399. doi:10.1002/sim.4067

Wiener, J. B. (1998). Managing the iatrogenic risks of risk management. *Risk Health, Safety & Environment, 9*(39), 39-82. Retrieved from https://scholarship.law.duke.edu/faculty_scholarship/135/

Wiig, S., Haraldseid-Driftland, C., Tvete Zachrisen, R., Hannisdal, E., & Schibevaag, L. (2019). Next of Kin Involvement in Regulatory Investigations of Adverse Events That Caused Patient Death: A Process Evaluation (Part I - The Next of Kin's Perspective). *J Patient Saf*. doi:10.1097/pts.0000000000000630

Wiig, S., Rutz, S., Boyd, A., Churruca, K., Kleefstra, S., Haraldseid-Driftland, C., . . . van de Bovenkamp, H. (2020). What methods are used to promote patient and family involvement in healthcare regulation? A multiple case study across four countries. *BMC Health Services Research, 20*(1), 616. doi:10.1186/s12913-020-05471-4

Yin, R. K. (1999). Enhancing the quality of case studies in health services research. *Health Services Research, 34*(5 Pt 2), 1209-1224. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1089060/pdf/hsresearch00022-0132.pdf

Yin, R. K. (2018). *Case study research and applications: Design and methods* (Sixth ed.). Thousand Oaks: Sage publications.

Øvretveit, J. (2009). Public Service Quality Improvement. In E. Ferlie, L. E. Lynn, & C. Pollitt (Eds.), *The Oxford Handbook of Public Management*. doi:10.1093/oxfordhb/9780199226443.003.0024

Øyri, S. F., Braut, G. S., Macrae, C., & Wiig, S. (2020). Exploring links between resilience and the macro-level development of healthcare regulation - a Norwegian case study. *BMC Health Services Research, 20*(1), 762. doi:10.1186/s12913-020-05513-x

Åsprang, A. F., Frich, J. C., & Braut, G. S. (2015). Organizational impact of governmental audit of blood transfusion services in Norway: A qualitative study. *Transfusion and Apheresis Science, 53*(2), 228-232. doi:10.1016/j.transci.2015.04.015

I

PLOS ONE

RESEARCH ARTICLE

Early diagnosis of sepsis in emergency departments, time to treatment, and association with mortality: An observational study

Gunnar Husabø[1,2]*, Roy M. Nilsen[3], Hans Flaatten[4], Erik Solligård[5,6], Jan C. Frich[7], Gunnar T. Bondevik[2,8], Geir S. Braut[9,10], Kieran Walshe[11], Stig Harthug[12,13], Einar Hovlid[1,2,10]

1 Department of Social Science, Western Norway University of Applied Sciences, Sogndal, Norway, **2** Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway, **3** Department of Health and Functioning, Western Norway University of Applied Sciences, Bergen, Norway, **4** Department of Clinical Medicine, University of Bergen, Bergen, Norway, **5** Clinic of Anaesthesia and Intensive Care, St. Olavs Hospital, Trondheim, Norway, **6** Department of Circulation and Medical Imaging and Mid-Norway Sepsis Research Group, Norwegian University of Science and Technology, Trondheim, Norway, **7** Institute of Health and Society, University of Oslo, Oslo, Norway, **8** National Centre for Emergency Primary Health Care, NORCE Norwegian Research Centre, Bergen, Norway, **9** Stavanger University Hospital, Stavanger, Norway, **10** Norwegian Board of Health Supervision, Oslo, Norway, **11** Alliance Manchester Business School, University of Manchester, Manchester, England, United Kingdom, **12** Department of Research and Development, Haukeland University Hospital, Bergen, Norway, **13** Department of Clinical Science, University of Bergen, Bergen, Norway

* gunnar.husabo@hvl.no

OPEN ACCESS

Citation: Husabø G, Nilsen RM, Flaatten H, Solligård E, Frich JC, Bondevik GT, et al. (2020) Early diagnosis of sepsis in emergency departments, time to treatment, and association with mortality: An observational study. PLoS ONE 15(1): e0227652. https://doi.org/10.1371/journal.pone.0227652

Editor: Martina Crivellari, Vita Salute University of Milan, ITALY

Received: August 27, 2019

Accepted: December 22, 2019

Published: January 22, 2020

Data Availability Statement: As this study used data from the Norwegian Patient Registry, there are legal restrictions on sharing the data set. The registry is regulated according to the Act relating to Personal Health Data Registries (2014) and specific regulation for the registry under the provision of this act. These laws do not permit us to share these data for secondary use. Third parties can apply to obtain data from the Norwegian Directorate of Health, which is the body granting access to the registry. The Directorate responds to inquiries

Abstract

Background

Early recognition of sepsis is critical for timely initiation of treatment. The first objective of this study was to assess the timeliness of diagnostic procedures for recognizing sepsis in emergency departments. We define diagnostic procedures as tests used to help diagnose the condition of patients. The second objective was to estimate associations between diagnostic procedures and time to antibiotic treatment, and to estimate associations between time to antibiotic treatment and mortality.

Methods

This observational study from 24 emergency departments in Norway included 1559 patients with infection and at least two systemic inflammatory response syndrome criteria. We estimated associations using linear and logistic regression analyses.

Results

Of the study patients, 72.9% (CI 70.7–75.1) had documented triage within 15 minutes of presentation to the emergency departments, 44.9% (42.4–47.4) were examined by a physician in accordance with the triage priority, 44.4% (41.4–46.9) were adequately observed through continual monitoring of signs while in the emergency department, and 25.4% (23.2–27.7)

regarding the Norwegian Patient Registry via email helseregistre@helsedir.no. Data obtained from the Norwegian Patient Registry can be used to identify electronic health records at the hospitals, pending approval from ethics committee and hospitals.

Funding: The Norwegian Board of Health Supervision funded the expenses incurred by the clinical experts on the inspection teams and part of the costs related to data collection. Erik Solligård was one of the clinical experts. Members of the inspection teams are employed by the County Governors in Norway. Western Norway University of Applied Sciences funds the PhD scholarship for Gunnar Husabø. Einar Hovlid, Gunnar Husabø and Geir Sverre Braut are part-time employees at the Norwegian Board of Health Supervision. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

received antibiotics within 1 hour. Delay or non-completion of these key diagnostic procedures predicted a delay of more than 2.5 hours to antibiotic treatment. Patients who received antibiotics within 1 hour had an observed 30-day all-cause mortality of 13.6% (10.1–17.1), in the timespan 2 to 3 hours after admission 5.9% (2.8–9.1), and 4 hours or later after admission 10.5% (5.7–15.3).

Conclusions

Key procedures for recognizing sepsis were delayed or not completed in a substantial proportion of patients admitted to the emergency department with sepsis. Delay or non-completion of key diagnostic procedures was associated with prolonged time to treatment with antibiotics. This suggests a need for systematic improvement in the initial management of patients admitted to emergency departments with sepsis.

Introduction

Sepsis is a major challenge, being present in a large proportion of hospitalizations that culminate in death [1–3]. Most sepsis cases seem to arise outside hospital settings [4], and these patients present to emergency departments with heterogeneous signs and symptoms, making detection and diagnosis challenging [5]. New sepsis criteria and early antibiotic treatment has been a major focus of research and debate over the last years [6] but factors associated with delayed treatment in the emergency departments have received less attention.

Previous research, mostly based on single case studies and smaller patient cohorts, suggests that systematic screening and diagnostic procedures for recognizing sepsis are not consistently carried out according to current guidelines [5, 7] and that sepsis is not recognized early enough [7]. Early recognition of sepsis is of critical importance for timely treatment [8–10], and compliance with sepsis guidelines is associated with improved outcomes [11–13]. However, no studies have assessed the association between timeliness of diagnostic procedures and time to treatment [11]. More knowledge about such associations can prove useful in improving initial care of the many patients admitted to emergency departments with sepsis. Moreover, there is a need for robust data documenting the extent to which diagnostic procedures are delayed or not carried out for patients with sepsis presenting to the emergency room.

The objectives of the study were to assess the timeliness of diagnostic procedures for recognizing sepsis in emergency departments and to evaluate associations between timeliness of procedures and time to initial administration of antibiotics and association between time to antibiotic administration and 30-day all-cause mortality.

Methods

Setting and participants

We conducted a multicenter, observational study based on data in electronic health records of 24 Norwegian hospitals.

The Norwegian health care system is publicly funded, and it scores relatively high on the Organisation for Economic Co-operation and Development's quality indicators [14]. In Norway, primary care physicians decide whether to refer patients with suspected sepsis to an emergency department for further assessment and treatment.

This study is a part of a 4-year longitudinal research project to assess the effects and outcomes of inspections on early detection of sepsis and time to treatment in emergency departments. The project was initiated in 2015 by the Norwegian Board of Health Supervision, which is the body delegated with the overall responsibility for external inspections of health care in Norway. The protocol for this project has been published previously [15]. In the present article, we report the results of the first part of the study, establishing baseline levels of compliance with sepsis guidelines, and assessing the associations between delayed diagnostic procedures and time to treatment and between time to treatment and mortality.

Measures of care delivery and outcome

The Norwegian Board of Health Supervision identified key clinical practices involved in recognition of sepsis by examining international guidelines [16, 17] and receiving advice from experts on sepsis. The choice of which care processes to include was based on key elements of the guideline: screening for sepsis and diagnosing sepsis, source control, and treatment. Operationalizing these elements into process measures of care delivery was a pragmatic decision based on what data that could be expected to be available in the electronic health records. The data were collected by inspection teams, who evaluated the practices at each hospital. We operationalized these practices into process measures of care delivery, which we used as the study variables (see Box 1). We defined mortality as all-cause mortality within 30 days from hospital admission.

Study cohort

We sampled data from electronic health records of patients admitted to 24 Norwegian hospitals from May 2015 through February 2017. Hospital size and geographic location were the main inclusion criteria. The 24 hospitals were representative of Norwegian hospitals with emergency departments and included all university and regional hospitals in Norway and a geographically based selection of local hospitals, together serving 75% of the total Norwegian population of 5 million. The hospitals ranged in size from 58 to 1640 beds and had emergency departments that served their local or regional communities.

We defined sepsis as *the suspicion of infection together with two systemic inflammatory response syndrome signs*, in accordance with internationally established and widely adopted definitions of sepsis at the time the protocol was developed [16]. The inclusion criteria were clinically suspected infection on presentation to an emergency department and at least two systemic inflammatory response syndrome signs, not including high leukocyte counts. We excluded high leukocyte counts as a criterion because the result of the blood sample in many cases would not be available for the clinicians when they do their initial judgment of severity of the patient's condition.

Organ failure was defined as fulfilling one of the following criteria at arrival to the emergency department: oxygen saturation <90% or PaO2/FiO2 <40 kPa, altered mental status, urine output <0.5 mL/kg/hour or increase in serum creatinine >50 micro mol/L, international normalized ratio >1.5 or activated partial thromboplastin time > 60 seconds, platelet count < 100 or 50% reduction in previous three days, serum bilirubin >70 mmol/L, serum lactate >4 mmol/L, blood pressure <90 systolic, mean arterial pressure <60, or fall in mean arterial pressure >40 mm Hg. We did not use a Sequential Organ Failure Assessment (SOFA) score for inclusion, as this was not in use at the emergency departments, and it was not possible to collect data retrospectively in order to evaluate the patients' SOFA score.

Box 1. Clinical processes of delivery of sepsis care.

- Proportion of patients triaged within 15 minutes of arrival at an emergency department.*
- Proportion of patients assessed by a physician in accordance with the urgency specified in the initial triage.
- Proportion of patients whose vital signs were measured within 1 hour of arrival at an emergency department.
- Proportion of patients whose blood lactate was measured within 1 hour of arrival at an emergency department.
- Proportion of patients from whom blood samples† were taken within 1 hour of arrival at an emergency department.
- Proportion of patients from whom blood cultures were taken before administration of antibiotics.
- Proportion of patients with adequate supplementary investigations to detect the focus of infection.
- Proportion of patients adequately observed‡ while in an emergency department.
- Proportion of patients who had received antibiotics within 1, 2, 4, and more than 4 hours.

* Norwegian hospitals are required to establish a system for prioritizing patients admitted to emergency departments. The scales that are in use are based on the South African Triage Scale (SATS) and the Rapid Emergency Triage and Treatment System (RETTS).

† Leukocyte count, hemoglobin, C-reactive protein, creatinine, electrolytes, platelet count, glucose, bilirubin, blood lactate

‡ 'Adequate' is defined as continual observation and measurement and documentation of vital signs at least every 15 minutes in critically ill patients with sepsis and organ failure, measurement and documentation of vital signs every 15 minutes if a physician has not examined a patient with sepsis but no documented organ failure, and every 30 minutes after first examination in such patients unless the physician decides otherwise.

Data collection

We used a two-step case ascertainment approach to identify eligible patients. First, we searched the Norwegian Patient Registry using a predefined list of the ICD-10 diagnostic codes that are most commonly used in Norway to classify sepsis and infections [18] (see S1 File). The patient registry contains diagnostic and therapeutic codes for all hospital admissions. The search produced a list of patients who had been discharged from the participating hospitals with a sepsis and/or infection code, together with an identification number that enabled access to the corresponding health records. Second, information about the patients' clinical status upon presentation to the emergency department from the individual patient records was assessed on-site at the hospitals to determine eligibility. Out of 5188 patients initially screened for eligibility, 1559 patients were included in the study (see S1 Fig). The sample size

was arrived at through power calculations to detect changes between the baseline measurements (which is the study sample used for this study) and post-inspection measurements. The power calculations are explained in detail in the study protocol [15].

Six regional inspection teams from the County Governors, who carry out inspections on behalf of the Norwegian Board of Health Supervision, were tasked with assessing eligibility and collecting data. The inspections were headed by experienced team leaders with particular training in performing similar inspections. The teams consisted of a minimum of four inspectors with medical and legal expertise, including an independent senior consultant physician in internal medicine or critical care medicine.

Each team performed four inspections within a time frame of about seven weeks in four geographically proximate hospitals. The inspections were rolled out sequentially from March 2016 to February 2017, averaging two per month. The sequence of inspections was randomized to facilitate comparison of outcomes before and after the inspections. The details of and rationale for the study design are described in the published study protocol [15].

The teams collected the data retrospectively during their inspection site visits. To allow for possible changes in clinical performance over time, they sampled data from two different time intervals for each hospital. For each time interval, we aimed to include the last 33 consecutive patients with sepsis who fulfilled the inclusion criteria on presentation to the emergency departments. The first sample included the last 33 patients admitted at each hospital before 1 October 2015, which was immediately before the Norwegian Board of Health Supervision announced the inspection campaign. The second sample included the last 33 patients before inspection of each hospital.

Data were recorded manually on case record forms, and subsequently digitized and saved to a single database containing information from all 24 participating hospitals. Upon completion of the database, we obtained information from the National Patient Registry on 30 day all-cause mortality and Charlson Comorbidity Index [19] for all patients, based on their national identity number and date of hospital admittance.

Data analyses

To assess the timeliness of diagnostic procedures, we calculated the percentages and 95% confidence intervals of patients with sepsis who had been documented as undergoing diagnostic procedures and receiving antibiotics within specified time limits (see Box 1). We did the analysis with all patients included and for the subgroup of patients with organ failure. Because several patient records lacked data on one or more measures of care delivery, we have also provided the number of records with missing documentation.

To estimate mean difference in time to first dose of antibiotics between categories of clinical procedure variables, we performed linear regression analyses using minutes to first dose of antibiotics as the outcome variable. Following previous research [16] and knowledge of the clinical care process in emergency departments, we focused on the following clinical procedures as exposure variables: triage within 15 minutes, examination by a physician in accordance with priority ascertained by triage, blood lactate measurements within one hour, and evidence of an adequate observation regimen within the emergency department. We performed univariate analyses for each procedure and then included all factors in a multivariable analysis.

Analyzing the association between time to diagnostic procedures and time to treatment, we needed to address the question of whether all kinds of delays were the results of clinical decisions to prioritize care for the patients who had the most serious clinical condition, thus making any association between delays to diagnostic procedures and delays to treatment a

spurious one. We controlled for such confounding by indication by including age, organ failure, and comorbidity as covariates in adjusted analyses. The results were then checked against a subgroup analysis of patients who had a serious medical condition at arrival identified by a red or orange triage color (see S2 File). We made additional adjustment for elapsed time since commencement of the study. Elapsed time was measured using calendar days, and was added as a cubic term.

We used a logistic regression model to estimate the 30-day all-cause mortality rate in relation to time to antibiotic administration. In this analysis, we used all-cause mortality as the binary outcome variable and time to first dose of antibiotics as a cubic exposure term, allowing for a non-linear relationship. We also made adjustment for age, year of admission (entered as categorical variables), Charlson comorbidity index, and organ failure as adjustment variables. We present the model-predicted mortality rates by time to antibiotic administration in a graphical format.

Patients who either had no antibiotic indication or had received antibiotic treatment before admittance in the emergency department were excluded from the regression analyses. So were patients for whom we lacked information on time to antibiotic treatment.

For some patients, data were missing for one or several of the variables included as covariates in the analyses. The results from blood samples are imported to the electronic health record; we therefore coded blood lactate as not taken within one hour when it was not documented in the patient record. Similarly, we coded patients who lacked documentation on adequate observation regimen within the emergency department as not having been observed adequately. We imputed data for four variables in our data set: time to antibiotics in minutes, time to examination by a physician, time to triage, and organ failure. Missing values were imputed using fully conditional specification [20]. See S2 File for more information about the treatment of missing data and the regression models.

The regression analyses were first performed including the whole study sample, and then for the sub-group of patients with organ failure (see supplemental S3 File).

We performed the statistical analyses with *Stata* versions SE 15.1 and IC 16.0 (StataCorp LP, College Station, TX, USA). For all regression models, we obtained cluster-robust standard errors of model parameters to account for intra-cluster correlations.

Results

The study included 1559 patients from 24 Norwegian hospitals from all the 19 Norwegian counties. Table 1 shows characteristics of the study cohort.

The percentages of patients who received care in line with the pre-defined standards are shown in Table 2.

Of the patients in our sample, 72.9% (95% confidence interval 70.7 to 75.1), had documented triage within 15 minutes of presentation to the emergency department, 44.9% (42.4 to

Table 1. Selected characteristics of the study cohort.

	Male	Female	All
N	800 (51.3%)	759 (48.7%)	1559
Mean (standard deviation) age	69.3 (16.5)	64.6 (20.9)	67.0 (18.9)
Median (min—max) age	72 (18–98)	69 (18–99)	71 (18–99)
Mean (standard deviation) CCI*	3.0 (2.5)	2.2 (2.2)	2.6 (2.4)
Organ failure	313 (39.5%)	244 (32.8%)	557 (36.3%)

* Charlson Comorbidity Index

https://doi.org/10.1371/journal.pone.0227652.t001

Table 2. Proportion of patients who underwent clinical procedures and received treatment in line with pre-defined standards.

		Number of patient records		Percent of patients documented receiving recommended care (95% confidence interval)	
Measure		**Records with documentataion***	**Records lacking documentation**	**All study patients†**	**Patients with organ dysfunction**
Diagnostics					
Complete assessment of vital signs within 1 hour		1360	199	83.6 (81.8 to 85.5)	81.0 (77.7 to 84.2)
Pulse rate measured within 1 hour		1496	63	93.3 (92.9 to 94.6)	94.1 (92.1 to 96.0)
Temperature measured within 1 hour		1492	67	93.1 (91.9 to 94.4)	94.3 (92.3 to 96.2)
Blood pressure measured within 1 hour		1493	66	92.7 (91.9 to 94.0)	93.9 (91.9 to 95.9)
Respiration rate measured within 1 hour		1476	83	91.5 (90.9 to 92.9)	93.4 (91.3 to 95.4)
Mental status assessed within 1 hour		1390	169	86.0 (84.8 to 87.7)	83.1 (80.0 to 86.2)
Blood culture taken prior to administration of antibiotics		1350	95	85.3 (83.8 to 87.1)	84.6 (81.6 to 87.6)
Adequate supplementary examinations to identify source of infection		1548	11	93.7 (92.9 to 94.9)	93.7 (91.7 to 95.7)
Time to triage ($\leq$ 15 min)		1375	184	72.9 (70.7 to 75.1)	77.0 (73.5 to 80.5)
Adequate observation regimen in ED		1524	35	44.4 (41.4 to 46.9)	47.4 (43.2 to 51.6)
Examination by physician in accordance with triage urgency		1105	454	44.9 (42.4 to 47.4)	47.6 (43.4 to 51.7)
Leukocytes count	Blood samples taken within 1 hour	1534	25	87.1 (85.8 to 88.8)	88.5 (85.9 to 91.2)
Hemoglobin		1533	26	87.2 (85.8 to 88.8)	88.2 (85.5 to 90.8)
C-reactive protein		1534	25	87.0 (85.8 to 88.7)	88.5 (85.9 to 91.2)
Creatinine		1525	34	86.7 (85.8 to 88.3)	88.3 (85.7 to 91.0)
Electrolytes		1524	35	86.8 (85.8 to 88.5)	88.3 (85.7 to 91.0)
Platelet count		1510	49	85.8 (84.8 to 87.5)	87.4 (84.7 to 90.2)
Glucose		1492	67	85.1 (83.8 to 86.9)	87.6 (84.9 to 90.4)
Bilirubin		963	482	62.0 (59.6 to 64.4)	66.2 (62.3 to 70.2)
Blood lactate		955	604	48.6 (46.4 to 51.1)	58.5 (54.4 to 62.6)
Treatment					
Antibiotics within 1 hour‡		1313	132	25.5 (23.2 to 27.7)	30.4 (26.4 to 34.3)
Antibiotics within 2 hours‡		1313	132	55.5 (52.9 to 58.1)	59.4 (55.2 to 63.6)
Antibiotics within 4 hours‡		1313	132	79.7 (77.6 to 81.7)	82.5 (79.2 to 85.7)

* Total number of records: 1559

† Patients with suspected infection together with two systemic inflammatory response syndrome signs

‡ n = 1438 (patients registered as needing antibiotic treatment and not having received antibiotics prior to admission)

https://doi.org/10.1371/journal.pone.0227652.t002

47.4) were examined by a physician in accordance with the priority specified during triage, and 83.6% (81.8 to 85.5) had a complete set of vital signs recorded within one hour of presentation. Blood samples were obtained within one hour from more than 80% of the patients for all specified tests except for bilirubin and lactate, 62.0% (59.6 to 64.4) and 48.6% (46.4 to 51.1), respectively; 44.4% (42.4 to 47.4) were adequately observed while in the emergency department according to the degree of priority assigned during triage; and 25.4% (23.2 to 27.7) and 55.5% (52.5 to 58.0) of the patients received antibiotics within one and two hours, respectively.

We found an association between non-completed or delayed diagnostic procedures and prolonged time to administration of antibiotics. In adjusted analyses (Model 1 in Table 3), patients who had not been triaged within 15 minutes had in average an extra delay of 54.7 minutes (95% confidence interval 33.2 to 76.2) to administration of antibiotics. We found a similar pattern of prolonged time to administration of antibiotics of cases where patients were not examined by a physician within the time limits set in triage, 61.2 minutes (40.8 to 81.6), not having blood lactate measured within one hour, 86.2 minutes (71.5 to 100.8), and not having an adequate observation regimen, 39.3 minutes (21.8 to 56.8). When we included all four procedures in one regression analysis (Model 2 in Table 3), they together predicted a delay of 159 minutes to first dose of antibiotics.

Replicating the regression analyses for the sub-group of patients with organ failure yielded similar results, with the model including all four factors also predicting an extra delay of 159 minutes for patients with organ failure (see S3 File).

Fig 1 shows the distribution of patients according to the number of the four specified procedures that were not performed within the recommended time limits.

The 30-day all-cause mortality was 9.9% (8.4 to 11.4) for the entire study sample and 17.4% (14.2 to 20.5) for patients with documented organ failure.

Fig 2 displays the observed 30-day all-cause mortality in hourly intervals for time from admission to administration of antibiotics (bars). Patients receiving antibiotics within 1 hour had an observed mortality of 13.6% (10.1 to 17.1), whereas those receiving antibiotics in the timespan 2 to 3 hours after admission had an observed mortality of 5.9% (2.8 to 9.1) and those receiving antibiotics 4 hours or later after admission had an observed mortality of 10.5% (5.7 to 15.3).

Fig 2 also shows the model-predicted 30-day all-cause mortality according to time to antibiotic treatment in minutes, adjusted for patient's age, date of admission and presence of organ failure (shown by the solid black curve).

Table 3. Linear regression for factors associated with delay in antibiotic treatment.

	Unadjusted	Model 1*	Model 2†
	b (95% CI)	*b (95% CI)*	*b (95% CI)*
Not triaged within 15 minutes	54.4 (32.9 to 75.9)	54.7 (33.2 to 76.2)	25.8 (3.8 to 47.8)
Examination by physician not in accordance with priority	60.0 (39.2 to 80.9)	61.2 (40.8 to 81.6)	38.0 (16.1 to 59.8)
Lactate not measured within 1 hour	81.6 (65.9 to 97.2)	86.2 (71.5 to 100.8)	71.4 (56.0 to 86.8)
Inadequate observation regimen	41.3 (22.3 to 60.4)	39.3 (21.8 to 56.8)	23.9 (10.5 to 37.3)

Outcome variable: Time to antibiotics measured in minutes. n = 1307

* Adjusted for organ failure, patient age, comorbidity, and time to admission

† Adjusted for the other variables in this table, and organ failure, age, comorbidity, and time to admission

https://doi.org/10.1371/journal.pone.0227652.t003

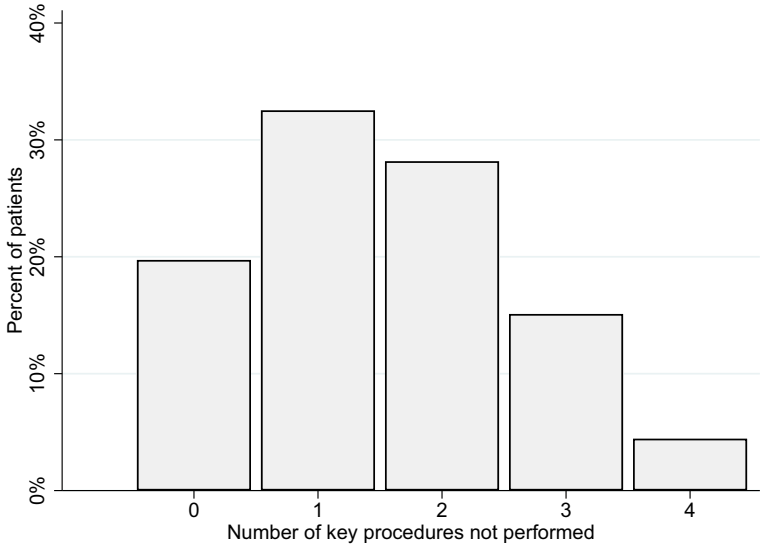


Fig 1. Distribution of patients according to number of non-completed or delayed key diagnostic procedures. Key procedures: triage within 15 minutes, examination by physician in accordance with urgency specified during triage, blood lactate measured within 1 hour, adequate observation regimen. N = 1559.

https://doi.org/10.1371/journal.pone.0227652.g001

Replicating the analysis of the association between time to antibiotic treatment and 30-day mortality for the sub-group of patients with organ failure, we found a similar curvilinear trend where predicted mortality was highest for patients who received antibiotics within one hour (see S3 File).

Discussion

Principal findings

In this study of 24 emergency departments, we found that they frequently failed to perform important diagnostic procedures in time, and that delays in or non-completion of diagnostic procedures were associated with prolonged time to administration of antibiotics. In 46% of the study patients, two or more of the following four key procedures had not been carried out in a timely manner: triage within 15 minutes, examination by physician in accordance with priority as set in triage, measuring blood lactate within one hour, and adequate observation. Non-completion or delay of these procedures together predicted a delay of 159 minutes to administration of antibiotics. We also found a substantial variation in mortality according to time to antibiotics. Patients who started antibiotic treatment between 2 and 3 hours after admission had lower mortality than those who started antibiotics earlier or later.

Strengths and limitations

The main strengths of our study are the combination of inclusion procedures and the size of the patient cohort from 24 different study sites. The hospitals that the study patients attended

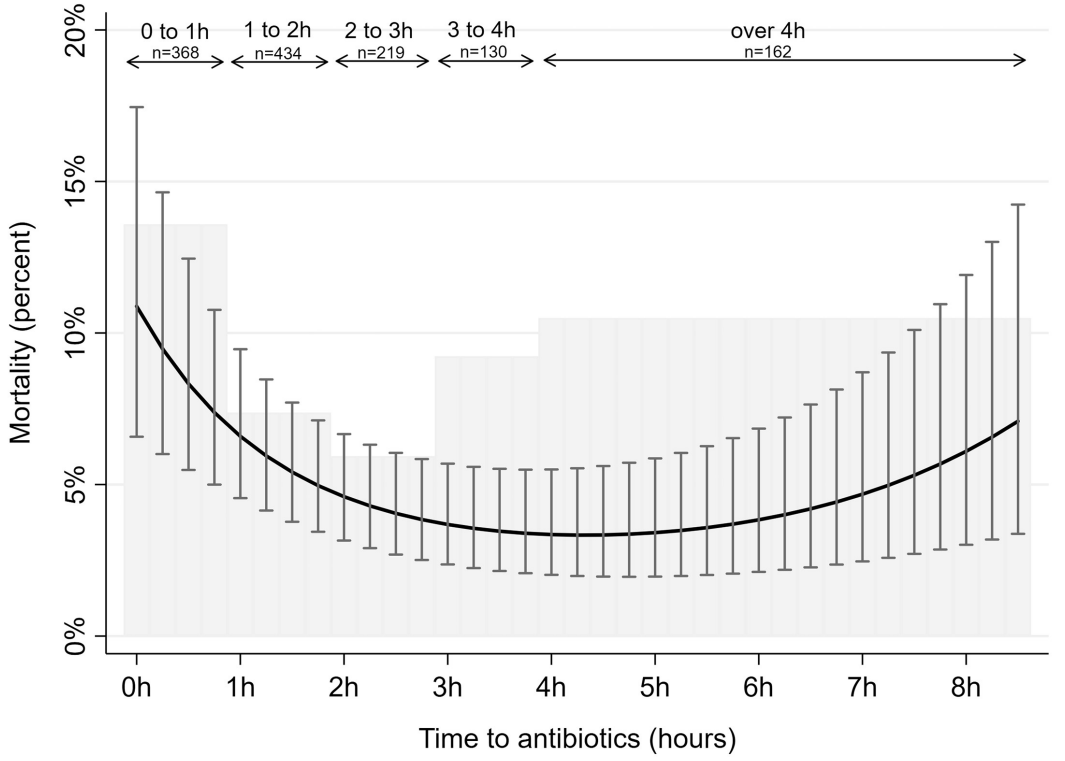


Fig 2. All-cause 30-day mortality by time to antibiotic treatment. Gray shaded histogram represents mortality rates according to time to antibiotic treatment in hours. Solid black curve with bars represents model-predicted mortality rates with 95% confidence intervals according to time to antibiotic treatment in minutes using logistic regression models, adjusted for patient's age, date of admission, comorbidity, and presence of organ failure. Date of admission was measured using calendar days since study start, entered as a polynomial function with first (b -0.011 $p<0.001$), second (b 2.5e-5 $p<0.001$) and third degree (b -1.2e-8 $p<0.01$) variables. The model prediction uses average values for adjustment values.

https://doi.org/10.1371/journal.pone.0227652.g002

can be considered representative of Norwegian hospitals because they included all university and regional hospitals and a geographically based selection of local hospitals, serving 75% of the total Norwegian population. Moreover, we included consecutive patients admitted to the emergency departments during two different time periods by using a cluster randomized sampling approach. Therefore, our study cohort is representative of patients admitted to Norwegian emergency departments with infection and meeting two or more systemic inflammatory response syndrome criteria.

Another important strength of our study is that we manually reviewed all patient records and established eligibility on the basis of recorded clinical data rather than on diagnostic codes alone. The latter approach can cause ascertainment bias and misleading inferences because coding practices for sepsis can vary over time and between hospitals [21].

A limitation of our study is the use of systematic inflammatory response syndrome (SIRS) as an inclusion criterion, in line with the *Sepsis-2* definition. Since the protocol was initially

drafted and the project started, a *Sepsis-3* definition as a *life-threatening organ dysfunction caused by a dysregulated host response to infection* was proposed [6]. As compared to Sepsis-2, Sepsis-3 represents only a minority of patients with infection [22]. Thus, our findings cannot be directly generalized to patient groups that have sepsis according to the Sepsis-3 definition. We have performed sub analyses of patients with organ failure, which is a group of patients that more closely overlaps with the Sepsis-3 definition, to make it easier to compare our findings to those of studies relying on the Sepsis-3 definition. These analyses show similar results for the sub-group of patients with organ failure as those of the whole study sample.

Another limitation of our study is that we did not have data on severity of sepsis in the form of commonly used severity scores like SAPS 2 (simplified acute physiology score) or APACHE II (acute physiology and chronic health evaluation), or a detailed organ failure assessment score like SOFA (sequential organ failure assessment). We did control for age, presence of organ failure, and comorbidity, which are three important variables associated with severity of sepsis [23]; however, even when controlling for these variables, the associations we found between time to antibiotics and mortality were probably subject to confounding by unmeasured variations in severity of illness and patient characteristics. As such, the study design does not allow for unbiased estimation of treatment effects.

Comparison with other studies

The delays we identified in diagnostic procedures are consistent with previous research findings of delay in time to triage [24], recording of vital signs [5, 7], measurement of blood lactate [25, 26], and delays in recognition of sepsis in general [7]. The processes we found to be lacking are essential for reaching an accurate diagnosis and institution of treatment in a timely manner [27].

We found that 25.5% and 55.5% of patients received antibiotics within one and two hours, respectively. This is in line with or slightly faster than the timing reported in previous studies with comparable patient cohorts in emergency department settings, which found that 28% of patients received antibiotics within 1 hour [7] and that median times to commencing antibiotics were 2.1 hours [28] and 182 minutes [29].

No previous studies have demonstrated the extent of delay or non-completion of diagnostic procedures for patients with sepsis in a large, representative cohort of patients admitted to emergency departments, or assessed the association between non-completed or delayed procedures and prolonged time to antibiotic administration.

The mortality rate in our study was in line with mortality rates reported in previous research in an emergency department setting [11]. However, we found a curvilinear association, where patients receiving early treatment and treatment later than four hours after admission had higher mortality rates than those receiving treatment between two and four hours after admission. This parabolic trend conflicts with a previous report of a linear increase in mortality with increasing time to antibiotics [13].

Interpretation of findings and implications

There is an ongoing debate concerning how timing of antibiotics for patients with sepsis should be operationalized in guidelines. The guidelines have come under criticism for not being adequately based in empirical evidence and being overly reliant on treatment protocols mandating antibiotic initiation within one hour of triage [30] and early administration of broad-spectrum antibiotics to all patients with sepsis [31]. Commenting on the sepsis guidelines, the Infectious Disease Society of America recommends administration of antibiotics as soon as possible to patients with severe infections. However, they warn that rigid guideline

recommendations with fixed time frames might increase the likelihood of broad-spectrum antibiotics will be given to uninfected patients [32]. In line with this argument, we maintain that the timing of antibiotic treatment should be an informed clinical decision rather than a consequence of unintended delays in diagnostic procedures. Our study indicates that the latter might often be the case: Delays in diagnostic procedures are common and they might lead to delayed treatment.

Emergency departments must therefore attend to optimizing diagnostic screening to improve time to treatment and overall management of sepsis. The Surviving Sepsis Campaign recommends a performance improvement program that includes screening for sepsis [33]; however, it is still necessary to define more precisely what screening measures should be implemented and how they should be monitored as part of the improvement program. We argue that our findings can inform this work.

We assert that the non-linear association we found between antibiotic treatment and mortality reflects the fact that many patients with sepsis are already critically ill when they present to an emergency department and that these patients are more easily recognized and given aggressive treatment earlier. This is an observation study, and the associations we found between time to antibiotics and mortality were probably subject to confounding by unmeasured variations in severity of illness and patient characteristics. Thus, one should not draw conclusions regarding the efficiency of antibiotic treatment at specific time intervals based on these analyses.

To our knowledge, this is the first multicenter study that assesses the association between a wide array of diagnostic procedures and antibiotic treatment. Previous research, mostly based on single case studies and smaller patient cohorts, has found delays in time to treatment comparable to those we found, suggesting that emergency departments elsewhere in Europe and the USA face challenges regarding variability of performance of initial screening procedures to detect sepsis. We therefore argue that our findings might have relevance for emergency departments outside of Norway.

Conclusions

We found that key procedures for recognizing sepsis and organ failure in the emergency department were delayed or not carried out in a substantial proportion of patients with sepsis. Delay or non-completion of key diagnostic procedures together predicted a delay of 2.5 hours to the first dose of antibiotics. Initiation of antibiotic treatment should be an informed clinical decision. Delays in antibiotic treatment could potentially have a negative effect on patient outcomes. These findings have important implications for managers and health professionals. The extent of delay and non-completion of important diagnostic procedures suggests that there is a need for systematic improvement efforts in the initial management of patients with sepsis presenting to emergency departments.

Supporting information

S1 File. ICD 10 codes. List of ICD 10 codes used to search the National Patient Register.
(PDF)

S2 File. Statistical analyses. Description of how missing data are treated and of how we fitted the regression models.
(PDF)

S3 File. Sub-analyses. Sub-analyses of association between diagnostic measures and time to treatment and between time to treatment and mortality for the sub-group of patients with

organ failure.
(PDF)

S1 Fig. Data collection process. Description Patient flow diagram showing the number of patients included and excluded.
(TIF)

Acknowledgments

We thank Paul Barach for reading a draft of the manuscript and providing valuable input. We also thank the inspection teams from the County Governors and the clinical experts for participating in data collection, and the staff at the Norwegian Board of Health Supervision who developed the sepsis inspection. An earlier draft of the manuscript was reviewed by Trish Reynolds and edited by Karyn Heavner, both from Edanz Group.

Disclaimer: Data from the Norwegian Patient Registry has been used in this publication. The interpretation and reporting of these data are the sole responsibility of the authors. No endorsement by the Norwegian Patient Registry is intended nor should be inferred.

Author Contributions

Conceptualization: Gunnar Husabø, Roy M. Nilsen, Hans Flaatten, Erik Solligård, Jan C. Frich, Gunnar T. Bondevik, Geir S. Braut, Kieran Walshe, Stig Harthug, Einar Hovlid.

Data curation: Gunnar Husabø.

Formal analysis: Gunnar Husabø, Roy M. Nilsen, Einar Hovlid.

Investigation: Einar Hovlid.

Methodology: Gunnar Husabø, Roy M. Nilsen, Hans Flaatten, Erik Solligård, Jan C. Frich, Gunnar T. Bondevik, Geir S. Braut, Kieran Walshe, Stig Harthug, Einar Hovlid.

Project administration: Gunnar Husabø, Einar Hovlid.

Resources: Einar Hovlid.

Supervision: Einar Hovlid.

Visualization: Gunnar Husabø, Einar Hovlid.

Writing – original draft: Gunnar Husabø, Einar Hovlid.

Writing – review & editing: Gunnar Husabø, Roy M. Nilsen, Hans Flaatten, Erik Solligård, Jan C. Frich, Gunnar T. Bondevik, Geir S. Braut, Kieran Walshe, Stig Harthug, Einar Hovlid.

References

1. Rhee C, Dantes R, Epstein L, Murphy DJ, Seymour CW, Iwashyna TJ, et al. Incidence and Trends of Sepsis in US Hospitals Using Clinical vs Claims Data, 2009–2014. JAMA. 2017; 318(13):1241–9. https://doi.org/10.1001/jama.2017.13836 PMID: 28903154
2. Prescott HC, Osterholzer JJ, Langa KM, Angus DC, Iwashyna TJ. Late mortality after sepsis: propensity matched cohort study. BMJ. 2016; 353:i2375. https://doi.org/10.1136/bmj.i2375 PMID: 27189000
3. Reinhart K, Daniels R, Kissoon N, Machado FR, Schachter RD, Finfer S. Recognizing Sepsis as a Global Health Priority—A WHO Resolution. N Engl J Med. 2017; 377(5):414–7. https://doi.org/10.1056/NEJMp1707170 PMID: 28658587
4. Page DB, Donnelly JP, Wang HE. Community-, Healthcare-, and Hospital-Acquired Severe Sepsis Hospitalizations in the University HealthSystem Consortium. Crit Care Med. 2015; 43(9):1945–51. https://doi.org/10.1097/CCM.0000000000001164 PMID: 26110490

5. Morr M, Lukasz A, Rübig E, Pavenstädt H, Kümpers P. Sepsis recognition in the emergency department–impact on quality of care and outcome? BMC Emerg Med. 2017; 17(1):11. https://doi.org/10.1186/s12873-017-0122-9 PMID: 28330460
6. Singer M, Deutschman CS, Seymour C, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016; 315(8):801–10. https://doi.org/10.1001/jama.2016.0287 PMID: 26903338
7. Goodwin A, Srivastava V, Shotton H, Protopapa K, Butt A, Mason M. Just say sepsis. A review of the process of care received by patients with sepsis London: National Confidential Enquiry into Patient Outcome and Death. 2015.
8. Gatewood MOK, Wemple M, Greco S, Kritek PA, Durvasula R. A quality improvement project to improve early sepsis care in the emergency department. BMJ Qual Saf. 2015; 24(12):787–95. https://doi.org/10.1136/bmjqs-2014-003552 PMID: 26251506
9. Pruinelli L, Westra BL, Yadav P, Hoff A, Steinbach M, Kumar V, et al. Delay Within the 3-Hour Surviving Sepsis Campaign Guideline on Mortality for Patients With Severe Sepsis and Septic Shock. Crit Care Med. 2018; 46(4):500–5. https://doi.org/10.1097/CCM.0000000000002949 PMID: 29298189
10. Torsvik M, Gustad LT, Mehl A, Bangstad IL, Vinje LJ, Damås JK, et al. Early identification of sepsis in hospital inpatients by ward nurses increases 30-day survival. Critical Care. 2016; 20(1):244. https://doi.org/10.1186/s13054-016-1423-1 PMID: 27492089
11. Damiani E, Donati A, Serafini G, Rinaldi L, Adrario E, Pelaia P, et al. Effect of performance improvement programs on compliance with sepsis bundles and mortality: a systematic review and meta-analysis of observational studies. PLoS One. 2015; 10(5):e0125827. https://doi.org/10.1371/journal.pone.0125827 PMID: 25946168
12. Levy MM, Rhodes A, Phillips GS, Townsend SR, Schorr CA, Beale R, et al. Surviving Sepsis Campaign: association between performance metrics and outcomes in a 7.5-year study. Crit Care Med. 2015; 43(1):3–12. https://doi.org/10.1097/CCM.0000000000000723 PMID: 25275252
13. Seymour CW, Gesten F, Prescott HC, Friedrich ME, Iwashyna TJ, Phillips GS, et al. Time to Treatment and Mortality during Mandated Emergency Care for Sepsis. N Engl J Med. 2017; 376(23):2235–44. https://doi.org/10.1056/NEJMoa1703058 PMID: 28528569
14. OECD. Health at a Glance 2017: OECD Indicators. Paris: OECD Publishing; 2017.
15. Hovlid E, Frich JC, Walshe K, Nilsen RM, Flaatten HK, Braut GS, et al. Effects of external inspection on sepsis detection and treatment: a study protocol for a quasiexperimental study with a stepped-wedge design. BMJ Open. 2017; 7(9).
16. Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM, et al. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock, 2012. Intensive Care Med. 2013; 39(2):165–228. https://doi.org/10.1007/s00134-012-2769-8 PMID: 23361625
17. National Institute for Health and Care Excellence. Sepsis: recognition, diagnosis and early management (NICE guideline 51). 2016.
18. Knoop ST, Skrede S, Langeland N, Flaatten HK. Epidemiology and impact on all-cause mortality of sepsis in Norwegian hospitals: A national retrospective study. PLoS One. 2017; 12(11):e0187990. https://doi.org/10.1371/journal.pone.0187990 PMID: 29149187
19. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. J Chronic Dis. 1987; 40(5):373–83. https://doi.org/10.1016/0021-9681(87)90171-8 PMID: 3558716
20. Van Buuren S. Flexible imputation of missing data: Chapman and Hall/CRC; 2018.
21. Gohil SK, Cao C, Phelan M, Tjoa T, Rhee C, Platt R, et al. Impact of policies on the rise in sepsis incidence, 2000–2010. Clin Infect Dis. 2016; 62(6):695–703. https://doi.org/10.1093/cid/civ1019 PMID: 26787173
22. Vincent J-L, Martin GS, Levy MM. qSOFA does not replace SIRS in the definition of sepsis. Crit Care. 2016; 20(1):210–. https://doi.org/10.1186/s13054-016-1389-z PMID: 27423462
23. Nasa P, Juneja D, Singh O. Severe sepsis and septic shock in the elderly: An overview. World J Crit Care Med. 2012; 1(1):23–30. https://doi.org/10.5492/wjccm.v1.i1.23 PMID: 24701398
24. Houston C, Sanchez LD, Fischer C, Volz K, Wolfe R. Waiting for Triage: Unmeasured Time in Patient Flow. West J Emerg Med. 2015; 16(1):39–42. https://doi.org/10.5811/westjem.2014.11.22824 PMID: 25671006
25. The Royal College of Emergency Medicine. Severe Sepsis and Septic Shock, Clinical Audit 2016/17. National Report.; 2017.
26. Hayden GE, Tuuri RE, Scott R, Losek JD, Blackshaw AM, Schoenling AJ, et al. Triage sepsis alert and sepsis protocol lower times to fluids and antibiotics in the ED. Am J Emerg Med. 2016; 34(1):1–9. https://doi.org/10.1016/j.ajem.2015.08.039 PMID: 26386734

27. National Institute for Health and Care Excellence. Sepsis. Quality standard QS1612017.

28. Liu VX, Fielding-Singh V, Greene JD, Baker JM, Iwashyna TJ, Bhattacharya J, et al. The Timing of Early Antibiotics and Hospital Mortality in Sepsis. Am J Respir Crit Care Med. 2017; 196(7):856–63. https://doi.org/10.1164/rccm.201609-1848OC PMID: 28345952

29. Stoneking LR, Winkler JP, DeLuca LA, Stolz U, Stutz A, Luman JC, et al. Physician documentation of sepsis syndrome is associated with more aggressive treatment. West J Emerg Med. 2015; 16(3):401–7. https://doi.org/10.5811/westjem.2015.3.25529 PMID: 25987914

30. Marik PE, Farkas JD, Spiegel R, Weingart S. POINT: Should the Surviving Sepsis Campaign Guidelines Be Retired? Yes. Chest. 2019; 155(1):12–4. https://doi.org/10.1016/j.chest.2018.10.008 PMID: 30616719

31. Patel JJ, Bergl PA. COUNTERPOINT: Should Broad-Spectrum Antibiotics Be Routinely Administered to All Patients With Sepsis as Soon as Possible? No. Chest. 2019; 156(4):647–9. https://doi.org/10.1016/j.chest.2019.05.031 PMID: 31590707

32. IDSA Sepsis Task Force. Infectious Diseases Society of America (IDSA) POSITION STATEMENT: Why IDSA Did Not Endorse the Surviving Sepsis Campaign Guidelines. Clin Infect Dis. 2017; 66 (10):1631–5.

33. Rhodes A, Evans LE, Alhazzani W, Levy MM, Antonelli M, Ferrer R, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. Intensive Care Med. 2017; 43(3):304–77. https://doi.org/10.1007/s00134-017-4683-6 PMID: 28101605

II

Open access Original research

BMJ Open

Effects of external inspections on sepsis detection and treatment: a stepped-wedge study with cluster-level randomisation

Gunnar Husabø ,[1,2] Roy Miodini Nilsen,[3] Erik Solligård,[4,5] Hans Kristian Flaatten,[6] Kieran Walshe,[7] Jan C Frich,[8] Gunnar Tschudi Bondevik,[2,9] Geir Sverre Braut,[10,11] Jon Helgeland ,[12] Stig Harthug,[13,14] Einar Hovlid [1,2,11]

To cite: Husabø G, Nilsen RM, Solligård E, *et al*. Effects of external inspections on sepsis detection and treatment: a stepped-wedge study with cluster-level randomisation. *BMJ Open* 2020;**10**:e037715. doi:10.1136/bmjopen-2020-037715

► Prepublication history and additional material for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2020-037715).

Received 14 February 2020
Revised 17 August 2020
Accepted 10 September 2020

For numbered affiliations see end of article.

Correspondence to
Dr Einar Hovlid;
einar.hovlid@hvl.no

ABSTRACT

Objective To evaluate the effects of external inspections on (1) hospital emergency departments' clinical processes for detecting and treating sepsis and (2) length of hospital stay and 30-day mortality.

Design Incomplete cluster-randomised stepped-wedge design using data from patient records and patient registries. We compared care processes and patient outcomes before and after the intervention using regression analysis.

Setting Nationwide inspections of sepsis care in emergency departments in Norwegian hospitals.

Participants 7407 patients presenting to hospital emergency departments with sepsis.

Intervention External inspections of sepsis detection and treatment led by a public supervisory institution.

Main outcome measures Process measures for sepsis diagnostics and treatment, length of hospital stay and 30-day all-cause mortality.

Results After the inspections, there were significant improvements in the proportions of patients examined by a physician within the time frame set in triage (OR 1.28, 95% CI 1.07 to 1.53), undergoing a complete set of vital measurements within 1 hour (OR 1.78, 95% CI 1.10 to 2.87), having lactate measured within 1 hour (OR 2.75, 95% CI 1.83 to 4.15), having an adequate observation regimen (OR 2.20, 95% CI 1.51 to 3.20) and receiving antibiotics within 1 hour (OR 2.16, 95% CI 1.83 to 2.55). There was also significant reduction in mortality and length of stay, but these findings were no longer significant when controlling for time.

Conclusions External inspections were associated with improvement of sepsis detection and treatment. These findings suggest that policy-makers and regulatory agencies should prioritise assessing the effects of their inspections and pay attention to the mechanisms by which the inspections might contribute to improve care for patients.

Trial registration NCT02747121.

Strengths and limitations of this study

- This is the first large-scale study using a robust design to evaluate the effects of external inspections on clinical care.
- As it was not possible to design a randomised controlled study, we used a stepped-wedge design, allowing the inspections to proceed as usual while we assessed effects based on data collected by the inspectors.
- Even though we adjusted for a range of known confounders, there is a risk that unknown external factors not included in the analyses introduced bias to the effect estimates.

INTRODUCTION

External assessment of healthcare providers is in widespread use as a policy strategy to foster improvement in the quality of care.[1] WHO defines assessment as an external institutional strategy and divides it into three subcategories: accreditation, certification and supervision.[2] According to WHO, *accreditation* generally refers to external assessment of an organisation by an accreditation body, *certification* is usually used to describe external assessment of compliance with standards published by the International Organisation for Standardisation (ISO) and *supervision* refers to an authoritative monitoring of healthcare providers' compliance with minimum standards often set by legislation.[2]

These assessment schemes represent heterogeneous, complex processes that consist of a set of activities that are introduced into varying organisational and regulatory contexts, and their origin and objectives can differ.[3] They share an important defining element in that: "some dimensions or characteristics of a health care provider organisation and its activities are assessed or analysed against a framework of ideas, knowledge, or measures derived or developed outside that organisation".[4] The phrase "external" also implies that the assessment is initiated and

conducted by an organisation external to the one being assessed.

External assessments can serve different purposes. They can represent a control strategy, emphasising whether providers meet certain standards, thereby promoting accountability and transparency in a regulated society.[5] However, they can also represent an improvement strategy, based on the assumption that externally promoted adherence to evidence-based standards contributes to higher quality of healthcare.[4] This assumption, however, seems to lack a clear scientific foundation. Although research suggests that external assessment can have a positive impact on an organisation level, for example, on improved leadership, quality systems and professional development,[6–9] less is known about the impacts of external assessment on the quality of care. According to a Cochrane review of the literature, there is a paucity of high-quality controlled evaluations on this topic.[10]

External assessments are contemporary, real-world events that involve autonomous actors, including healthcare providers, inspecting organs and policy-makers. The complexity of the settings in which external assessments take place may explain why only three randomised controlled studies have been performed to evaluate their effect on quality of care: two small-scale studies[11 12] and one study[13] whose methods have been criticised for leading to unreliable conclusions.[10] None of these studies found improvements in patient care resulting from external assessments. Three other studies that used time-series design and a before-and-after design also did not find improvement in performance indicators of care delivery that could be attributed to external assessments.[14–16] Considering the widespread and growing use of external assessments and the resources spent on conducting and participating in them, there is a need for high-quality studies evaluating their effectiveness on quality of care.[10]

The overall aim of this study is to evaluate the effect of statutory inspections at the patient level by assessing detection and treatment of sepsis in emergency departments in Norwegian hospitals. According to WHO's classification, this represents an example of assessment in the supervision category. We use the term "external inspection", in line with the Cochrane review.[10]

Sepsis is a major public health challenge and a leading cause of death.[17] The inspections in the present study assessed adherence to standards for sepsis care that have been shown to be associated with improved patient outcomes.[18 19] Specifically, the aim of our study was to evaluate the effects of the inspections on hospital emergency departments' clinical processes for detecting and treating sepsis and on length of hospital stay and 30-day mortality.

METHODS

To study the effects of the inspections, we used a pragmatic cross-sectional incomplete stepped-wedge design with cluster-level randomisation at the regional level. The study included data from all hospitals subject to the sepsis inspections.

Setting

The Norwegian healthcare system is publicly funded, and it scores high on Organization for Economic Co-operation and Development (OECD) quality indicators.[20] All provision of health services in Norway is regulated by legislation. Healthcare services should be safe, effective and provided in accordance with sound professional practice, and all organisations that provide healthcare services must have a quality management system to ensure that healthcare services are provided in accordance with the legal requirements.[21] The organisation of specialised emergency care is based on the principle of equal access to services. The government designates a specific geographical area to each hospital, within which the hospital is responsible for providing emergency care to the whole population. Patients within the designated area in need of hospital emergency services are admitted to the hospital after referral or prior contact with general practitioners or other medical professionals.[22]

The Norwegian Board of Health Supervision and its regional-level subordinate, the County Governors, are mandated by law to ensure that healthcare is provided in accordance with the legal requirements. An important way of fulfilling this mandate is conducting thematic, nationwide inspections. The themes of these inspections are decided on the basis of information about risk and vulnerability. Norway does not have a mandated system of hospital accreditation, and there are no other regulatory agencies or government bodies supervising the provision of health services.

Intervention

As an intervention, we studied inspections in 24 hospitals in Norway. The inspections addressed early detection and treatment of patients with sepsis admitted to the hospitals' emergency departments. The inspection campaign lasted from April 2016 to March 2018.

The sepsis inspections were planned and directed by the Norwegian Board of Health Supervision, and they were carried out by six inspection teams from the County Governors. Each team performed four inspections within a time frame of about 8 weeks in four geographically proximate hospitals. The inspections were headed by experienced team leaders who were trained in performing inspections. The teams consisted of a minimum of four inspectors with medical and legal expertise, including an independent senior consultant physician in internal medicine or critical care medicine.

The inspections were based on the ISO's procedures for system audits[23] and encompassed three main phases: the announcement of the inspection and collection of relevant data, the site visit, and reporting and follow-up. The information and data reviewed during the inspections comprised administrative documentation, interviews

with management staff and personnel with responsibilities related to care for patients with sepsis, and patient records. During the follow-up period, the inspection teams conducted verification of patient records at 8 and 14 months after the initial inspection.

On the basis of all the information and gathered data, the inspection team assessed whether the emergency department's clinical processes for sepsis detection and treatment were in line with the regulatory standard. A key part of the inspections was identifying and pointing out underlying reasons for substandard performance of the clinical system delivering care to patients admitted with sepsis. The inspection team also assessed to what extent the hospital management had fulfilled their legal obligation to implement a functional management system that monitors, and when necessary improves, the quality of sepsis detection and treatment. In this way, the inspections challenged the quality of performance through addressing the managerial level's responsibility for ensuring good practice through providing an expedient organisational framework for delivering sound professional practice. The inspection teams' findings and conclusions were presented in reports that were made publicly available on the Internet (one of the reports is provided here as an example—see online supplemental material 1). The reports focused on identified non-conformities in the quality of care for patients with sepsis. All inspections found instances of substandard performance, the most common being delay in antibiotic treatment. After each site visit, there was a follow-up phase. During this phase, the hospital management were held responsible for developing and implementing necessary measures to improve substandard performance.

Table 1 provides an overview of the key elements of the intervention.

Study design, participants and data collection

The inspection campaign was mandated to include all 18 counties in Norway, and each County Governor decided which hospitals to inspect in their region. The main inclusion criterion was hospital size because substandard care in larger hospitals would potentially affect more patients. The hospitals selected for inspection comprised all university and regional hospitals and a geographically based selection of local hospitals. They included 24 out of 50 hospitals in Norway with emergency services, and served 75% of the total population.

This study was developed in conjunction with the planning of the inspections. It was not feasible to establish an unexposed control group, and, for practical reasons, the intervention could not be delivered simultaneously to the entire study population. We therefore used a pragmatic cross-sectional incomplete stepped-wedge design with cluster-level randomisation where all inspected hospitals were included in the study.[24 25] The inspections were carried out sequentially in the 24 hospitals over 12 months from April 2016 to March 2017. The County Governors notified the hospitals of the inspections 2 to 6 months in advance of the site visits. The order of the inspections was randomised at the regional level using a computer-generated list of random numbers; each inspection team received a randomly assigned time slot of 8 weeks during which they were to conduct the inspection of four hospitals in their region. Figure 1 provides an overview of the trial profile. We have previously published the full protocol of the study, including the rationale for the use of the stepped-wedge design.[25]

We based inclusion into the study on the standard definition of sepsis in use when the study was developed.[26] Accordingly, the criteria were clinically suspected infection on presentation to the emergency department and at least two systemic inflammatory response syndrome

Table 1 Key elements of the intervention

Time in months	Activity
1	Inspection team announces inspection and requests the hospital to submit information
2	Inspection team reviews records of patients with sepsis and collect relevant data for the inspection criteria. Data are collected for two time periods, baseline (September 2015) and right before the site visit. Inspection team reviews information from hospital and prepares for the site visit
3	Two-day site visit at the hospital with interviews of key personnel At the end of the site visit, the inspection team presents the preliminary findings, and the hospital can comment on these preliminary findings
4–5	The inspection team writes a preliminary report of their findings. The hospital can comment on the report
6	The inspection team sends the final report to the hospital
Continuously	The hospital plans and implements improvement measures
11	Follow-up audit (8 months after site visit). The inspection team reviews records of patients with sepsis and collect the same kinds of data as they did prior to the site visit Report on findings from audit. Require the hospital to implement necessary changes
17	Follow-up audit (14 months after site visit). The inspection team reviews records of patients with sepsis and collect the same kinds of data as they did prior to the site visit. Report on findings from audit. Require the hospital to implement necessary changes

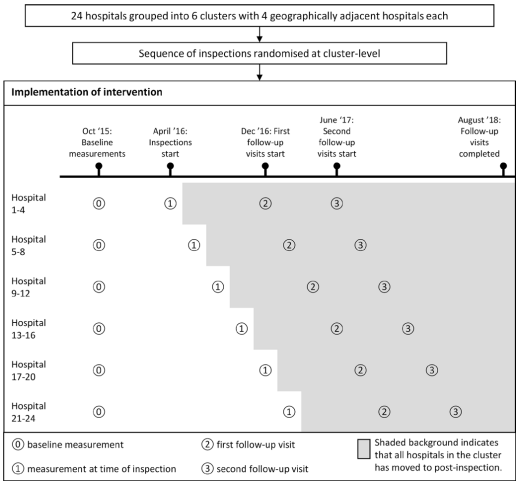


Figure 1 Trial profile.

signs, not including high leucocyte counts. The included patients were aged 18 years or older.

We used a two-step approach to identify eligible patients. First, we searched the Norwegian Patient Registry using a predefined list of the ICD-10 diagnostic codes most commonly used in Norway to classify sepsis and infections (online supplemental material 2).[27] The Patient Registry contains diagnostic and therapeutic codes for all hospital admissions. The search produced a list of patients who had been discharged from the participating hospitals with a sepsis and/or infection code, together with an identification number that enabled us to access the corresponding health records. Second, we assessed the individual patient records for eligibility by collecting information about these patients' clinical status on presentation to the emergency department.

We collected data for four time periods for each hospital: two before the inspection and two after the inspection. In the first data collection period, we included patients admitted to all hospitals before 1 October 2015. This was before the Norwegian Board of Health Supervision announced the national inspection campaign. The second collection period varied across the hospitals. The endpoint of this period was the day prior to the site visit at each hospital. The third and fourth time periods were also specific to each hospital, encompassing the 8 and 14 months after the initial site visit, respectively. For each time period, we included the last 83 consecutive patients who fulfilled the inclusion criteria on presentation to the emergency department. For all patient records, we gathered data from the electronic health records about patient age, sex, admission and discharge dates, and the presence of organ failure. Following national evidence-based guidelines,[26 28] we defined organ failure as fulfilling at least one of the following criteria at arrival to the emergency department: oxygen saturation <90% or PaO_2/FiO_2 <40 kPa, altered mental status, urine output <0.5 mL/kg/h or increase in serum creatinine >50 µmol/L, international normalised ratio >1.5 or activated partial thromboplastin time >60 s, platelet count <100 or 50% reduction in previous 3 days, serum bilirubin >70 mmol/L, serum lactate >4 mmol/L, blood pressure <90 mm Hg systolic, mean arterial pressure <60 mm Hg or fall in mean arterial pressure >40 mm Hg. For the first 33 patients, we also collected data on diagnostic measures and treatment given. The data on diagnostic measures and treatment were gathered by the inspection teams and used as audit evidence during the inspection and follow-up visits.

It was not possible to blind health personnel to the intervention, as information about the inspections was publicly known and health personnel participated in interviews and during follow-up. Nor was it possible to blind inspectors and researchers reviewing health records to the intervention or control condition, as information about time and dates was critical to the review.

We obtained data on 30-day all-cause mortality and the Charlson Comorbidity Index[29] for the included patients from the Norwegian Patient Registry by connecting data using a unique personal identifier. For patients who had multiple admissions, we used data relating to the first admission.

We performed power calculations using the *stepped-wedge* function[30] in Stata/IC, V.14.0 (StataCorp, College Station, TX, USA). For the measures of care delivery, we powered the study to detect an absolute improvement of 70% to 83% and a reduction in mortality of 5% to 11%. We assumed an intra-cluster correlation of 0.05, and type I and type II errors were assumed to be 0.05 and 0.20, respectively. See the study protocol for further details of the power calculations.[25]

Study outcomes and covariates

Previous research has emphasised the importance of early recognition of sepsis in enabling timely treatment[31 32] and demonstrated an association between compliance with evidence-based standards and improved outcomes.[18 19 33] The Norwegian Board of Health Supervision has identified key clinical processes involved in the recognition and treatment of sepsis by examining international guidelines[26 34] and soliciting advice from experts on sepsis. The indicators we used as study variables for measures of care delivery (see box 1) were operationalised from the inspection criteria used by the Board of Health Supervision. We defined mortality as all-cause mortality within 30 days of hospital admission, and we defined length of stay as the number of days from admission date to discharge date. In the study protocol, we included the percentage of patients receiving oxygen therapy as a measure of care delivery. However, information on oxygen therapy was not systematically recorded in the electronic patient records, so we had to exclude this variable from our analysis.

Box 1 Study measures

Measures of health care delivery

- Proportion of patients triaged within 15 min of arrival at the emergency department.*
- Proportion of patients assessed by a physician in accordance with the urgency specified in the initial triage.
- Proportion of patients whose vital signs were measured within 1 hour of arrival at the emergency department.
- Proportion of patients whose blood lactate was measured within 1 hour of arrival at the emergency department.
- Proportion of patients from whom blood samples† were drawn within 1 hour of arrival at the emergency department.
- Proportion of patients from whom blood cultures were taken before the administration of antibiotics.
- Proportion of patients with adequate supplementary investigations to detect the focus of infection.
- Proportion of patients who were adequately observed‡ while in the emergency department.
- Proportion of patients who were adequately discharged from the emergency department for further treatment in the hospital (written statement indicating patient status, treatment and further actions).
- Percentage of patients who received intravenous fluids within 1 hour.
- Proportion of patients who received antibiotics within 1 hour.

Outcome measures

- Length of stay in hospital.
- 30-day all-cause mortality.

*Norwegian hospitals are required to establish a system for prioritising patients admitted to emergency departments. The scales used for this are based on the South African Triage Scale and the Rapid Emergency Triage and Treatment System.
†Leucocyte count, haemoglobin, C reactive protein, creatinine, electrolytes, platelet count and glucose.
‡'Adequate' is defined as continuous observation, as well as measurement and documentation of vital signs at least every 15 min in critically ill patients with sepsis and organ failure, measurement and documentation of vital signs every 15 min if a physician has not examined a patient with sepsis but no documented organ failure, and every 30 min after first examination in such patients unless the physician decides otherwise.

Statistical analysis

We compared patient care delivery and outcomes before and after the inspection. Patient characteristics were compared using univariate analyses with Pearson's χ^2 test or the Wilcoxon rank-sum test.

Using logistic regression, we assessed the strength of the associations between having had an inspection and patients receiving adequate care, and we report the associations as ORs. To analyse changes in the patient outcome variables from before the inspection to after the inspection, we used logistic models for 30-day mortality and negative binomial models for length of stay. Because the patient data were sampled from different hospitals, we used mixed-effects models with hospital number included as a random effect to account for clustering.

In our analysis of changes in diagnosis and treatment, we report both unadjusted and adjusted models. In the adjusted models, we controlled for selections of the following variables: age, organ failure, Charlson Comorbidity Index, sex, seasonality and calendar year. We based the choice of control variables for each model on the Akaike information criteria score. Non-significant adjustment variables were kept in the models if doing so improved the overall model fit (see online supplemental material 3). Age and Charlson Comorbidity Index were entered into the model as linear terms, organ dysfunction and sex were entered as categorical variables, and seasonality and calendar year were entered as categorical variables. For patient outcomes, we report estimates from models with the adjustment variables age, organ dysfunction, Charlson Comorbidity Index and sex, with and without additional adjustment for calendar year.

Among the 3082 health records from which we had collected data on diagnostic and treatment processes, there was missing information on several of the process variables. Data on the diagnostic variables, complete vital measures, blood samples, lactate, observation regimen and supplemental diagnostic procedures were, to a great extent, automatically incorporated in the electronic patient records if they were performed. For these variables, we therefore recoded missing data as the procedure not having been performed within the given time limit. The other variables with missing information were blood culture taken prior to antibiotic treatment (5% missing), timely assessment by a physician (24% missing), time to triage (11% missing), time to fluid administration (29% missing) and time to administration of antibiotics (7% missing). For these variables, we could not assume a specific reason for missing observations. We decided not to impute missing data on these variables because running multiple imputation on dependent variables is not recommended in the absence of an auxiliary variable that correlates strongly with the imputation variable, as was the case here.[35]

The statistical analyses were performed using Stata/IC, V.16 (StataCorp).

Patient and public involvement

Patient organisations participated in a reference advisory group for the overall research programme, which included this study. They were involved from the planning stage on, but they did not directly participate in developing the research questions or outcome measures used in this article. We used their inputs to inform the overall study design. Patient organisations strongly advocated the importance of disseminating the study findings to relevant parties. The Norwegian Board of Health Supervision has held a national, public conference for hospitals, government agencies and patient representatives where we presented preliminary study findings.

RESULTS

A total of 7407 patients with sepsis were included in the study. Figure 2 shows the flow of patients through the study.

The median age of patients in the pre-inspection and post-inspection groups were 70 and 73 years, respectively.

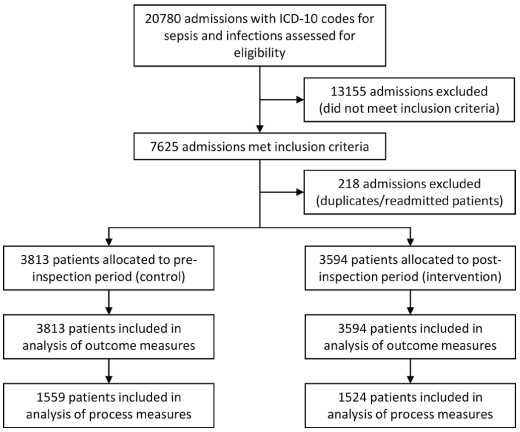


Figure 2 Patient flow.

A larger proportion of patients in the post-inspection group had organ failure (39.3%), compared with the pre-inspection group (36.6%) (see table 2).

Changes in diagnosis and treatment

Relative to the pre-inspection group, the post-inspection group had higher odds for being examined by a physician within the time frame set in triage (OR 1.28, 95% CI 1.07 to 1.53), higher odds for having complete set of vital measurements taken within 1 hour (OR 1.78, 95% CI 1.10 to 2.87), higher odds for having measured lactate within 1 hour (OR 2.75, 95% CI 1.83 to 4.15), higher odds for having adequate observation regimen (OR 2.20, 95% CI 1.51 to 3.20) and higher odds for antibiotics being administered within 1 hour (OR 2.16, 95% CI 1.83 to 2.55). Figure 3 displays changes in process measures before and after the inspections.

Changes in patient outcomes

On average, the length of stay was significantly shorter in the post-inspection group (6.2 days) than in the pre-inspection group (7.1 days). There was also a significant reduction in all-cause mortality within 30 days of admission, from 11.1 in the pre-inspection group to 10.5 in the post-inspection group. After controlling for time, neither of these changes were statistically significant (see table 3).

DISCUSSION

In the adjusted analyses, five of the measures of healthcare delivery showed significant improvement from before to after the inspection. The improvements were observed in aspects of care delivery that are of great importance to patients. We found significant improvement in time to treatment with antibiotics, which is associated with increased survival in patients with severe sepsis and septic shock.[18 19] Moreover, we also found that significant improvement for timely assessment by a physician, lactate measurement, adequate observation and taking a complete set of vital signs are all key clinical processes in the early detection of sepsis. The first three mentioned measures are significantly associated with earlier treatment with antibiotics.[36] The measures that were improved coincided with the main targets of the inspections and the key measures that the hospitals were required to improve following the inspections.

Although there were improvements also in the other measures of healthcare delivery and in the patient outcomes, 30-day mortality and length of stay, these changes were not statistically significant after adjusting for time. This implies that we cannot specifically attribute these improvements to the inspections.

Strengths and limitations

The main strength of our study is that it comprised a robust evaluation of the effects of external inspections on quality of care in real-world settings. To the best of our knowledge, the present research is the largest and most comprehensive study of inspection effects using a cluster-randomised research design.

It is, nevertheless, important to discuss whether the changes in the observed measures are attributable to causes other than the inspections. To do this, we address questions relating to coding and documentation,

Table 2 Patient characteristics

Factor		Before inspection	After inspection	P value
N		3813	3594	
Sex	Male	1939 (50.9%)	1881 (52.3%)	0.2*
	Female	1874 (49.1%)	1713 (47.7%)	
Age, years	Median (IQR)	70 (56–81)	73 (60–82)	<0.001†
	Mean (SD)	66.8 (18.8)	68.9 (18.5)	
Organ failure		1387 (36.6%)	1409 (39.3%)	0.015*
Charlson Comorbidity Index	Median (IQR)	2 (1–4)	2 (1–4)	0.49†

*Pearson's χ^2 test.
†Wilcoxon rank-sum test.

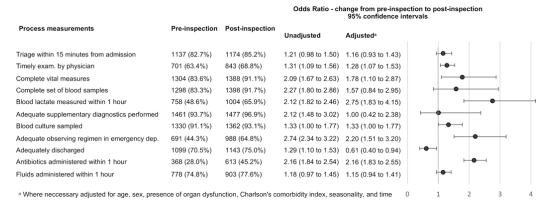


Figure 3 Process measures before and after inspection.

unknown confounders, contamination by other initiatives and time adjustments.

In 2016, the Society of Critical Care Medicine and the European Society of Intensive Care Medicine launched a new sepsis definition.[37] This initiative may have led to changes in the coding practice for sepsis. However, inclusion in the present study was based on the assessment of the clinical status of patients with infection at arrival in the emergency department. Patients at Norwegian hospitals are discharged with one main diagnosis and up to seven secondary diagnoses. If a patient has had any kind of infection during the hospital stay, this will be coded as a primary or secondary diagnosis at discharge. Therefore, changes in the coding practice for sepsis could not have biased our analyses. There is a risk that by oversight, no infection or sepsis code was assigned in a health record, resulting in that the record was not included in our screening. However, there is no reason to believe that the relative frequency of such errors would differ between the pre-inspection and post-inspection data. Therefore, such errors are not likely to have influenced our estimates of the inspection effects.

For some of the clinical processes examined in this study, the degree of documentation in the electronic patient records improved after the inspections. To assess whether the observed improvements could be caused by improvements in documentation, we analysed the association between time to treatment and missing data on diagnostic procedures. We found that having missing data on diagnostic procedures was associated with prolonged time to treatment (see online supplemental material 3). Given the association between time to diagnosis and time to treatment, we expect observations with missing data to be, on average, more delayed, compared with observations for which we have data. Thus, improvements in the process variables are, in all probability, not caused by improvements in documentation.

Due to the preconditions and constraints provided by doing the study in a real-world setting, it was not possible to conduct a randomised controlled trial.[25] As with any observational study, there was an inherent risk of confounding from unknown factors. A limitation of our study is that we did not have data on severity of sepsis in the form of commonly used severity scores like SAPS 2 (simplified acute physiology score) or APACHE II (acute physiology and chronic health evaluation), or a detailed organ failure assessment score like SOFA (sequential organ failure assessment). We did control for age, presence of organ failure and comorbidity, which are three important variables associated with severity of sepsis.[38]

Another potential source of confounding was influence on emergency department practices by other external factors. The stepped-wedge design reduced the risk of such biases. The overlap period where clusters of hospitals switched from control to intervention according to a randomised schedule encompassed 1433 patients. In this period, other factors besides the inspections that could contribute to the observed improvements would affect both the intervention group and the control group simultaneously. In addition, we also controlled for seasonality and year of admittance.

When it comes to confounding by external factors, a particular concern in our study was the possibility of other nationwide initiatives influencing sepsis management at

Table 3 Patient outcomes before and after inspection

	Before	After	Adjusted for background variables*	Adjusted for background and time†
All-cause mortality, 30 days from admission	424 (11.1%)	377 (10.5%)	0.81 (0.69 to 0.95)‡	1.25 (0.86 to 1.80)‡§
Length of stay in days, mean (SD)	7.1 (7.8)	6.2 (6.1)	0.87 (0.84 to 0.90)¶	0.94 (0.86 to 1.03)¶
n	3813	3594	7371	7371

*Adjusted for age, organ dysfunction, sex and Charlson Comorbidity Index.
†Adjusted for secular trends entered as categorical variables per year, plus the other background variables.
‡OR.
§n=7360.
¶Incidence rate ratio.

the emergency departments. There were no other regulatory initiatives that could affect these practices, as the Norwegian Board of Health Supervision and the County Governors are the only bodies assessing compliance with regulatory standards. During the period the present study was conducted, the Norwegian government initiated a voluntary sepsis-management improvement programme. Some of the included hospitals chose to participate in this programme. We specifically analysed the possible impacts of participation in this programme. Participation did not have a significant effect in terms of explaining the improvements in the process measures, and it had a negligible impact on the estimated sizes of the inspection effects. We therefore chose not to include participation in the patient safety programme as a control variable.

'Anticipatory effects'—stemming from improvement initiatives by the hospitals made in preparation of an upcoming inspection—may also represent a source of bias. Such effects can be considered as constituent parts of the total impact of the inspections.[39] Anticipatory initiatives could have influenced the emergency departments' processes in the timespan between announcement and inspection. As some of the pre-inspection data stems from this period, we could expect our estimates of the inspection effect to be attenuated by the influence of anticipatory effects on the pre-inspection data.

Finally, the time variable may introduce a bias to the regression model which results in an underestimated effect of the intervention. Due to the constraints of doing the research in a real-world setting, the majority of observations were collected from the periods before and after inspection where, respectively, none and all of the hospitals had been inspected. This results in correlation between time and intervention. There are too few observations of inspected hospitals in the early period of the study and of uninspected hospitals in the later period to consistently estimate the true effect of time, independent of the intervention effect. There were significant reductions in mortality and length of stay after the inspections when adjusting for age, organ dysfunction, sex and Charlson Comorbidity Index. When additionally adjusting for time, there were no longer any significant reductions. The lack of significant effects on patient outcome was not expected, as previous research has suggested that the improvments we observed in care delivery would lead to improved patient outcomes,[18] and challenges with modelling of secular trends is one potential explanation for this finding.

Interpretation of findings and comparison with other studies

In contrast to previous studies, which were unable to detect association between external inspections and improvement in quality of care,[11–13] we found improvements in key measures of care delivery, including time to treatment. The lack of significant associations between inspection and the outcome measures when adjusting for time might be due to the heterogeneity of the patient group included in the study. The effect of earlier treatment on reduced mortality has been documented in patients with severe sepsis and septic shock.[18] As only a proportion of patients in our study had severe sepsis and septic shock, it seems reasonable to expect more modest reduction in mortality. Furthermore, we need to be careful when reviewing the results from the models where time is included as an adjustment variable. Regardless of whether the improvements in patient outcomes in this study can be attributed to the inspections or not, we found significant improvements in key processes of care delivery including diagnostic processes, observation and time to treatment. These are key processes of care delivery that will enable medical personnel to make sound clinical decisions and initiate treatment processes that have shown to be important to patient outcomes.[18 36]

The fact that our study showed such improvements after the inspections, whereas previous studies did not, might be explained by contextual factors and how the inspections were conducted. WHO has described a generic framework for how external assessment can contribute to quality improvement: (1) development of standards addressing requirements that will lead to improvement in patient care, (2) reliable identification of performance gaps, (3) involvement of managers and professionals in developing action plans in response to the assessment, and (4) implementation of the plans in a way that lead to improvement.[2]

Several aspects of the Norwegian sepsis inspections are noteworthy in relation to this framework. Compared with the previously studied inspections,[15 16] the sepsis inspections in Norway had a narrower target with requirements that were closely related to patient care. Previous work has suggested that, to contribute to improvement, the inspection process should be translated into something meaningful and understandable for clinical practice.[14 39] The inspections in our study explicitly targeted the early detection and treatment of sepsis, which are crucial for patients and highly relevant and understandable from a clinical perspective. The methods used during the inspections provided the hospitals with reliable and valid data on how their emergency departments performed in terms of the early detection and treatment of patients with sepsis.[36] The hospitals were also provided with quantitative and qualitative data that could shed light on possible reasons for substandard performance of the clinical care delivery system.

The inspections addressed the hospital management staff's responsibility for ensuring evidence-based practice through providing an expedient organisational framework for the clinical care delivery system. During the two follow-up visits, the management was provided with feedback on the progress of the improvement efforts and held accountable for implementing the necessary changes. Previous research has indicated that holding the inspected organisation accountable for implementing changes can contribute to the creation of momentum for implementing the necessary improvement measures.[39] Together with orienting the target of the inspections

towards quality of care delivery, this emphasis on following up and holding the hospital management accountable for making changes to the clinical care delivery system may have contributed to the improvements we found.

Adequate discharge, the one process measure which showed a negative trend after the inspections, was not emphasised during the inspections because it was not directly related to early detection of the sepsis diagnosis and treatment. Neither was it a measure that the hospitals were required to report on following the inspection. The fact that this measure was not emphasised during the inspection process might in part explain the negative trend. This finding is also in line with previous research indicating that inspections can have a negative impact on measures that are not within the main purview of the inspections.[15]

Our findings also bring into view a larger concern regarding internal and external assessment schemes. There is always a risk that the use of quantitative indicators instigates managerial 'gaming', whereby hospitals put efforts into improving their scoring on specific indicators rather than on improving the system delivering care.[40] The inspections in our study used a set of performance indicators that together provided information about how the clinical system delivering care for patients with sepsis performed as a whole.[36] Improving these indicators would thus contribute to an overall systemic improvement in the quality of care for patients with sepsis. If the inspection maintains a clear focus on the overall goal of quality of care by assessing a set of performance indicators that matters for patient care, gaming behaviour by inspected hospitals can be dissuaded.

CONCLUSIONS

Comparing a range of measures before and after inspection, we found improvements in both processes and patient outcomes following the inspections. Though the improvements in patient outcomes cannot be specifically attributed to the inspections, we find substantial improvements in care delivery for patients with sepsis. Our findings indicate that inspections can be used to foster large-scale improvements in quality of care. However, it does not mean that conducting inspections necessarily leads to improvements. Inspections are complex interventions, and one can assume that their efficacy depends on the context in which they are conducted and how they are planned and implemented. Policy-makers and inspecting bodies should therefore prioritise assessing the effects of their inspections and pay attention to the mechanisms by which inspections might contribute to improving care for patients.

Author affiliations
[1]Department of Social Science, Western Norway University of Applied Sciences, Sogndal, Norway
[2]Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway
[3]Department of Health and Functioning, Western Norway University of Applied Sciences, Bergen, Norway
[4]Clinic of Anesthesia and Intensive Care, St. Olavs Hospital University Hospital in Trondheim, Trondheim, Norway
[5]Gemini Center for Sepsis Research, Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Norway
[6]Department of Clinical Medicine, University of Bergen, Bergen, Norway
[7]Alliance Manchester Business School, The University of Manchester, Manchester, UK
[8]Institute of Health and Society, University of Oslo, Oslo, Norway
[9]National Centre for Emergency Primary Health Care, NORCE Norwegian Research Centre AS, Bergen, Norway
[10]Department of Research, Stavanger University Hospital, Stavanger, Norway
[11]Department 5, Norwegian Board of Health Supervision, Oslo, Norway
[12]Norwegian Institute of Public Health, Oslo, Norway
[13]Department of Research and Development, Haukeland University Hospital, Bergen, Norway
[14]Department of Clinical Science, University of Bergen, Bergen, Norway

Acknowledgements We thank the inspection teams from the County Governors and the clinical experts for participating in data collection. We also thank the staff at the Norwegian Board of Health Supervision who developed the sepsis inspection, and who participated in the data collection and administration of the project. An earlier draft of the manuscript was reviewed by Trish Reynolds and edited by Jennifer Barrett, both from Edanz Group.

Contributors GH: conception, data collection, data curation, statistical analyses, visualisation, project administration, writing – original draft, writing – critical revision, final approval. RMN: conception, statistical analyses, writing – critical revision, final approval. ES: conception, data collection, writing – critical revision, final approval. HKF: conception, writing – critical revision, final approval. KW: conception, writing – critical revision, final approval. JCF: conception, writing – critical revision, final approval. GTB: conception, writing – critical revision, final approval. GSB: conception, data collection, translation of supplemental material, writing – critical revision, final approval. JH: conception, writing – critical revision, final approval. SH: conception, writing – critical revision, final approval. EH: conception, statistical analyses, project administration, supervision, visualisation, writing – original draft, writing – critical revision, final approval.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Disclaimer Data from the Norwegian Patient Registry have been used in this publication. The interpretation and reporting of these data are the sole responsibility of the authors. No endorsement by the Norwegian Patient Registry is intended nor should be inferred.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Ethics approval was obtained from the Regional Ethics Committee of Norway North (2015/2195/REK nord) and the Norwegian Data Protection Authority (15/01559). The Regional Ethics Committee of Norway Nord ruled that individual patient consent was not required for this study.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. As this study used data from the Norwegian Patient Registry, there are legal restrictions on sharing the data set. The registry is regulated according to the Act relating to Personal Health Data Registries (2014) and specific regulation for the registry under the provision of this act. These laws do not permit us to share these data for secondary use. Third parties can apply to obtain data from the Norwegian Directorate of Health, which is the body granting access to the registry. The Directorate responds to inquiries regarding the Norwegian Patient Registry via email (helseregistre@helsedir.no). Data obtained from the Norwegian Patient Registry can be used to identify electronic health records at the hospitals, pending approval from ethics committee and hospitals.

ORCID iDs
Gunnar Husabø http://orcid.org/0000-0002-8726-1769
Jon Helgeland http://orcid.org/0000-0002-9348-7844
Einar Hovlid http://orcid.org/0000-0003-0300-4405

REFERENCES

1 Shaw CD, Braithwaite J, Moldovan M, *et al*. Profiling health-care accreditation organizations: an international survey. *Int J Qual Health Care* 2013;25:222–31.
2 Shaw C, Groene O, Berger E. External institutional strategies: accreditation, certification, supervision. In: Busse R, Klazinga N, Panteli D, *et al.*, eds. *Improving healthcare quality in Europe: characteristics, effectiveness and implementation of different strategies*. Copenhagen (Denmark): European Observatory on Health Systems and Policies, 2019.
3 Campbell M, Fitzpatrick R, Haines A, *et al*. Framework for design and evaluation of complex interventions to improve health. *BMJ* 2000;321:694–6.
4 Walshe K, Freeman T, Latham L, *et al*. *Clinical governance – from policy to practice*. Birmingham, UK: University of Birmingham, Health Services Management Centre, 2000.
5 Walshe K, Boyd A. *Designing regulation: a review*. Manchester: The University of Manchester – Manchester Business School, 2007.
6 Greenfield D, Braithwaite J. Health sector accreditation research: a systematic review. *Int J Qual Health Care* 2008;20:172–83.
7 Shaw CD, Groene O, Botje 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 1:100–7.
8 Hinchcliff R, Greenfield D, Moldovan M, *et al*. Narrative synthesis of health service accreditation literature. *BMJ Qual Saf* 2012;21:979–91.
9 Shaw C, Groene O, Mora N, *et al*. Accreditation and iso certification: do they explain differences in quality management in European hospitals? *Int J Qual Health Care* 2010;22:445–51.
10 Flodgren G, Gonçalves-Bradley DC, Pomey M-P. External inspection of compliance with standards for improved healthcare outcomes. *Cochrane Database Syst Rev* 2016;12:Cd008992.
11 Oude Wesselink SF, Lingsma HF, Ketelaars CAJ, *et al*. Effects of government supervision on quality of integrated diabetes care: a cluster randomized controlled trial. *Med Care* 2015;53:784–91.
12 Oude Wesselink SF, Lingsma HF, Reulings PGJ, *et al*. Does government supervision improve stop-smoking counseling in midwifery practices? *Nicotine Tob Res* 2015;17:572–9.
13 Salmon J, Heavens J, Lombard CBethesda MD, ed. *The impact of accreditation on the quality of hospital care: KwaZulu-Natal Province, Republic of South Africa*. The U.S. Agency for International Development (USAID) by the Quality Assurance Project, University Research Co., LLC, 2003.
14 Evaluation team OPM. *Evaluation of the Healthcare Commission's Healthcare Associated Infections Inspection Programme*. London: OPM, 2009.
15 Castro-Avila A, Bloor K, Thompson C. The effect of external inspections on safety in acute hospitals in the National health service in England: a controlled interrupted time-series analysis. *J Health Serv Res Policy* 2019;24:1355819619837288:182–90.
16 Allen T, Walshe K, Proudlove N, *et al*. Measurement and improvement of emergency department performance through inspection and rating: an observational study of emergency departments in acute hospitals in England. *Emerg Med J* 2019;36:326–32.
17 Rudd KE, Johnson SC, Agesa KM, *et al*. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the global burden of disease study. *The Lancet* 2020;395:200–11.
18 Levy MM, Rhodes A, Phillips GS, *et al*. Surviving sepsis campaign: association between performance metrics and outcomes in a 7.5-year study. *Crit Care Med* 2015;43:3–12.
19 Seymour CW, Gesten F, Prescott HC, *et al*. Time to treatment and mortality during mandated emergency care for sepsis. *N Engl J Med* 2017;376:2235–44.
20 OECD. *Health at a glance 2017: OECD indicators*. Paris: OECD Publishing, 2017.
21 Braut GS. The requirement to practice in accordance with sound professional practice. In: Molven O, Ferkis J, eds. *Healthcare, welfare and law Oslo*. Gyldendal akademisk, 2011.
22 Blinkenberg J, Pahlavanyali S, Hetlevik Øystein, *et al*. General practitioners' and out-of-hours doctors' role as gatekeeper in emergency admissions to somatic hospitals in Norway: registry-based observational study. *BMC Health Serv Res* 2019;19:568.
23 ISO. *Conformity Assessment – Requirements and Recommendations for Content of a Third-Party Audit Report on Management Systems*. Geneva: International Organization for Standardization, 2012. ISBN: ISO/IEC TS 17022:2012.
24 Hemming K, Lilford R, Girling AJ. Stepped-wedge cluster randomised controlled trials: a generic framework including parallel and multiple-level designs. *Stat Med* 2015;34:181–96.
25 Hovlid E, Frich JC, Walshe K, *et al*. Effects of external inspection on sepsis detection and treatment: a study protocol for a quasiexperimental study with a stepped-wedge design. *BMJ Open* 2017;7:e016213.
26 Dellinger RP, Levy MM, Rhodes A, *et al*. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Crit Care Med* 2013;41:580–637.
27 Knoop ST, Skrede S, Langeland N, *et al*. Epidemiology and impact on all-cause mortality of sepsis in Norwegian hospitals: a national retrospective study. *PLoS One* 2017;12:e0187990.
28 Norwegian Directorate of Health. *Sepsis. Antibiotika i sykehus*. 2, 2018.
29 Charlson ME, Pompei P, Ales KL, *et al*. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373–83.
30 Hemming K, Girling A. A menu-driven facility for power and detectable-difference calculations in stepped-wedge cluster-randomized trials. *Stata J* 2014;14:363–80.
31 Gatewood Medley O'Keefe, Wemple M, Greco S, *et al*. A quality improvement project to improve early sepsis care in the emergency department. *BMJ Qual Saf* 2015;24:787–95.
32 Pruinelli L, Westra BL, Yadav P, *et al*. Delay within the 3-hour surviving sepsis campaign guideline on mortality for patients with severe sepsis and septic shock. *Crit Care Med* 2018;46:500–5.
33 Damiani E, Donati A, Serafini G, *et al*. Effect of performance improvement programs on compliance with sepsis bundles and mortality: a systematic review and meta-analysis of observational studies. *PLoS One* 2015;10:e0125827.
34 National Institute for Health and Care Excellence. Sepsis: recognition, diagnosis and early management. In: *NICE guideline 51*, 2016. https://www.nice.org.uk/guidance/ng51
35 Allison P. Missing data. In: *The SAGE Handbook of quantitative methods in psychology*. SAGE Publications Ltd, 2009. http://methods.sagepub.com/book/sage-hdbk-quantitative-methods-in-psychology
36 Husabø G, Nilsen RM, Flaatten H, *et al*. Early diagnosis of sepsis in emergency departments, time to treatment, and association with mortality: an observational study. *PLoS One* 2020;15:e0227652.
37 Singer M, Deutschman CS, Seymour CW, *et al*. The third International consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016;315:801–10.
38 Nasa P, Juneja D, Singh O. Severe sepsis and septic shock in the elderly: an overview. *World J Crit Care Med* 2012;1:23–30.
39 Smithson R, Richardson E, Roberts J. *Impact of the Care Quality Commission on provider performance: room for improvement?* London: The King's Fund, 2018.
40 Bevan G, Hood C. What's measured is what matters: targets and gaming in the english public health care system. *Public Adm* 2006;84:517–38.

III

Open access Original research

BMJ Open

Promoting leadership and quality improvement through external inspections of management of sepsis in Norwegian hospitals: a focus group study

Gunnar Husabø [1,2] Inger Lise Teig,[2] Jan C Frich,[3] Gunnar Tschudi Bondevik,[2,4] Einar Hovlid [1,2]

To cite: Husabø G, Teig IL, Frich JC, *et al*. Promoting leadership and quality improvement through external inspections of management of sepsis in Norwegian hospitals: a focus group study. *BMJ Open* 2020;**10**:e041997. doi:10.1136/bmjopen-2020-041997

► Prepublication history and additional material for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2020-041997).

Received 22 June 2020
Revised 20 October 2020
Accepted 28 October 2020

[1]Department of Social Science, Høgskulen på Vestlandet, Sogndal, Norway
[2]Department of Global Public Health and Primary Care, Universitetet i Bergen, Bergen, Norway
[3]Institute of Health and Society, Universitetet i Oslo, Oslo, Norway
[4]National Centre for Emergency Primary Health Care, NORCE Norwegian Research Centre AS, Bergen, Norway

Correspondence to
Gunnar Husabø;
gunnar.husabo@hvl.no

ABSTRACT

Objective Inspections and other forms of external assessment may contribute to positive changes in the health services, but the mechanisms of such change remain unclear. We did a study to explore how external inspections may foster clinical improvement in hospitals.

Design Focus group study.

Setting Statutory inspections of sepsis treatment in hospital emergency departments in Norway.

Participants Clinicians, managers and inspection teams involved with the inspections of sepsis treatment in emergency departments at four different hospitals. Twelve focus group interviews were carried out, with a total of 47 participants.

Results Three themes emerged as central for understanding how the inspections could contribute to clinical improvement in the emergency departments: (1) increasing awareness about the need to improve the quality of care by providing data on clinical performance, (2) building acceptance for improvement through professional credibility and focus on clinical practice, and (3) fostering leadership commitment.

Conclusion Our findings suggest that the inspections have the potential to enhance hospital management and staff's understanding of complicated care processes and help strengthen the organisational commitment to bring about systemic quality improvements.

Strength and limitations of this study

- Focus group interviews in hospitals that had achieved improvement in key clinical procedures following an inspection provided information-rich cases of how inspections can contribute to quality improvement.
- The interviews elicited new insights into how inspections can enhance understanding of the clinical system and promote leadership in quality improvement efforts.
- We did not explore change mechanisms related to anticipatory effects resulting from the announcement of upcoming inspections.
- The generalisability of our findings and interpretations are dependent on the organisational and procedural context in which inspections are being held.

INTRODUCTION

External inspection, also referred to as statutory supervision, is an external assessment strategy that is used to evaluate if healthcare providers meet accepted quality standards. Compared with other forms of external assessment, such as certification and accreditation, external inspections differ in that they are run by government bodies and subject to country-specific regulations.[1] While the subject and scope vary greatly from one inspection to another, most inspections have in common the goal of improving the quality of care provided by the organisations subject to the inspection.[2]

The rationale for why external assessment strategies could lead to improved quality is that managers will review the results of assessments and implement changes that are necessary for better and safer healthcare.[1] Such effects might function through directive steps, in which the inspectors guide or force the health organisation to act in a specific way. They can also be a result of 'softer' mechanisms, such as if inspections lead to a shift in focus and organisational objectives at the service provider.[3] In either case, the inspectors themselves cannot directly affect the quality of care being provided. As such, they must find ways to improve the quality of care through influencing the care processes and internal controls at the hospitals. External inspection can thus be seen as a way of boosting the internal quality and patient safety improvement work.[4]

Following the argument above, the effectiveness of inspections would likely depend on the degree to which they support organisational attributes and work processes associated with successful improvement. The literature describes readiness for change as a main dimension influencing the chance of success when implementing improvement efforts in healthcare organisations.[5] This view is rooted in a notion of organisations as communities that contribute to the amplification and development of knowledge rather than merely entities of hierarchical information processing.[6]

Research has shown mixed effects of inspections on improvement in healthcare organisations. Some studies have found care practices to improve following inspections but not been able to fully establish the association between the inspections and the improvements.[7 8] Other studies have not found any improvements following inspections at all.[9 10] Gaining a deeper insight into the mechanisms of change in connection with external inspections is needed in order to understand how and under what circumstances inspections might lead to substantial, long-lasting improvement.[11 12]

Our overall aim was to study how external inspections may foster clinical improvement, using the case of a nationwide inspection of sepsis treatment in emergency departments at Norwegian hospitals. We sought to explore clinicians', managers' and inspection teams' experiences of being involved in the inspection process and to explore their views on how inspections can affect the quality of care.

METHODS

Study design

The study is a part of an ongoing research on the impact of external inspection of sepsis diagnosis and treatment in emergency departments in Norwegian hospitals. The study protocol has been described previously.[13] The inspections were planned and directed by the Norwegian Board of Health Supervision (NBHS) at 24 hospitals with acute care functions.

For this study, we chose a qualitative approach, conducting focus group interviews with clinicians, managers and inspectors. We found this to be a well-suited method of inquiry, as the focus group discussion can provide interpretive insights into the participants experiences and opinions.[14] Our approach is informed by a realist paradigm, its concept of causal mechanisms providing a framework for understanding the conditions under which inspections may foster clinical improvement.[15] The study follows Standards for Reporting Qualitative Research (SRQR) guidelines.[16]

The sepsis inspections

In Norway, health services are publicly funded and based on the principle of universal and equitable access. They are mandated by legislation to be safe, effective and provided in accordance with sound professional standards. NBHS is responsible for ensuring that health services meet these requirements. One of their main supervision approaches is nationwide thematic inspections of services, prioritised on the basis of information about risk and vulnerability. During these inspections, NBHS or the County Governors, who are local representatives of the central government, investigate services and report any identified non-conformities. While NBHS can impose its authority on healthcare organisations and individual healthcare workers through a wide range of responses and sanctions, the reactions issued after nationwide inspections are normally limited to instructing the organisations to correct the situation. The inspectors will then follow-up the organisation until the non-conformity is considered satisfactorily corrected.[17]

NBHS chose diagnosis and treatment of sepsis in hospital emergency departments as a subject of a thematic inspection starting in 2016 because patients presenting to emergency departments with sepsis often receive substandard care.[18] Delayed treatment is a major challenge, as time is of paramount importance in treatment of sepsis.[19 20] Because early treatment depends on early diagnosis and recognition,[21 22] the failures in expediting the treatment often come down to failures in recognising the diagnosis at an early stage.[18]

There were six regional inspection teams. Each team included three to four inspectors from the County Governors with prior training and experience from either healthcare or law. Additionally, each team had an external medical specialist who had extensive clinical experience from working with sepsis diagnosis and treatment.

Methodologically, the inspections were system audits.[23] NBHS used existing guidelines and conferred with experts to formulate a set of quantitative criteria for recommended diagnosis and treatment of sepsis.[24 25] At inspection, which typically lasted for 2 days, the team gathered data from the electronic health records of a set of 66 patients with sepsis and evaluated the care given against the criteria. As is customarily done in system audits, the inspection teams also reviewed documentation of relevant procedures and interviewed clinicians and managers responsible for the care of patients with sepsis. At the final day of inspection, the main findings were presented to the hospital management and staff in a closing meeting. Afterwards, the inspection team wrote up a report that included findings and a list of non-conformities. The report was sent as a draft to the hospital's executive management for comments and eventually finalised and released to the public via the internet. A translated version of the report from one of the inspections is provided as online supplemental file 1, and an overview of the findings from the four inspections included in this study is provided as online supplemental file 2.

Participants and data collection

This study draws on data from 12 focus group interviews with clinicians, managers and inspection teams involved in the inspection of four of the hospitals (designated A, B,

C and D). The interviews were conducted after the initial inspection, in the period from March 2017 to November 2018. Analyses that included all inspected hospitals found that, on average, the inspection had a positive effect on several care process measures.[26] We chose to include these four hospitals in the present study because they were among the hospitals that showed substantial improvements following the inspection. An overview of the improvements in a key indicator, time to antibiotic treatment, is provided in online supplemental file 2.

We conducted separate focus group interviews with clinicians, managers and the inspection teams at each hospital. The focus groups were sized from three to five participants and included in total 47 interviewees: 15 clinicians, 16 managers and 16 inspection team members.

The groups of clinicians consisted of physicians and nurses who had diagnosis and treatment of patients with sepsis in the emergency department as a part of their daily tasks. The managers were either head nurses at emergency departments, chief physicians or heads of clinics. As such, the manager focus groups had a mix of interviewees in managerial roles and interviewees with combined responsibility for management and patient care. Clinicians and managers were recruited to the focus groups via contact persons with responsibility for quality management in the hospitals. We recruited all members currently on the inspection team who were available to attend the interview. As the members of the inspection teams changed over time, some inspection team interviewees had not participated in the inspections at the specific hospitals included in our study. The participants were informed beforehand about the purpose of the interviews and they signed a form agreeing to participate in the study. No compensation was given for participation in the study.

The interviews were conducted by GH (male, MSc), except for two interviews that were conducted in collaboration with EH (male, MD/PhD). GH had no previous affiliation with NBHS but had experience from performance audit work in healthcare organisations. EH had a part-time position as a researcher in NBHS and had previously participated in NBHS inspections. He was acquainted with some of the interviewees from his work in NBHS.

For hospitals A, B and C, the interviews with clinicians and managers were conducted at the respective hospitals. The interviews with the inspection teams were conducted at County Governors' offices. For hospital D, all interviews were conducted by conference call due to vast travel distances and logistical challenges with convening the inspection team to a physical meeting. The interviewers and the participants were the only ones attending the interviews.

We used three different interview guides, one for each of the three types of groups. The interview guides focused on the impact of the inspections on the quality of care, and the interviews were centred on the experiences from the sepsis inspections (see table 1). Additionally, time was devoted to discussing sepsis care in general and specific issues surrounding the organisation of work in emergency departments.

Table 1 Interview topics

Topic	Probes (sample items)
General experience of the inspection process	
Relevance	► What was the focus of the inspection? ► Are the themes covered in the inspection relevant for clinical practice?
Dialogue between inspection team and hospital	► How were findings conveyed to the hospital? How did the management/staff react to the findings?
Process for following up	► What has the hospital done in response to the identified non-conformities? ► Who were involved in following up the findings from the inspection?
The role of management	► What are important management tasks related to the inspection?
Contribution to change	► How did the inspection impact the internal quality improvement work? ► What factors other than the inspection have had an impact on quality improvement work? ► How is the quality of care now, compared with before the inspections?

The focus group interviews lasted from 35 to 105 min. After each session, field notes were recorded describing how the interview went and whether there were important contextual factors that should be taken into account in the analysis.

Transcription and analysis

Interviews were digitally recorded and subsequently transcribed and imported to NVivo Qualitative Data Analysis Software V.12 (QSR International Pty). Participants did not receive copies of transcripts.

We analysed the data using a thematic analytic approach.[27] After the first interview, before analysing the transcript, EH and GH introduced some preliminary codes (awareness of current and desired practice, leader commitment, use of performance metrics, communication and network, staff engagement and systems thinking). Other codes were added throughout the interviews and the subsequent coding of the material.

Once GH had done the initial coding of the interview transcriptions, EH and GH identified potential themes from the data material. We grouped the codes we

considered relevant for understanding the relationship between inspections and improvement work into these themes. Next, we analysed the interviews, first within each hospital, and then cross-case including all interviews, using the themes as an analytical framework.

As the focus groups were made up of three distinct roles, clinicians, managers and inspection team, we took extra care to compare and contrast the analyses between these roles. The interviews with clinicians and managers were more specific to the inspection in their hospital compared with the interviews with the inspection teams because the inspection teams could draw on experiences from all inspected hospitals in their region.

We read the transcripts and listened to the recorded interviews numerous times to ensure immersion, and we refined, synthesised and reorganised the identified themes according to our developing understanding of the material. We also extracted quotations from the material to illustrate themes and analytical points.

GH translated the quotes into English, and the translations were checked by all co-authors.

Patient and public involvement

Patient organisations participated in a reference advisory group for the overall research programme that this study is a part of. They were involved from the planning stage on, but they did not directly participate in developing or framing this specific article. We used their inputs to inform the overall study design. Patient organisations strongly advocated the importance of disseminating the study findings to relevant parties. NBHS has held a national, public conference for hospitals, government agencies and patient representatives, where we presented preliminary study findings.

RESULTS

We identified three themes as central for understanding how the inspections could contribute to clinical improvement in the emergency departments: (1) increasing awareness about the need to improve the quality of care by providing data on clinical performance, (2) building acceptance for improvement through professional credibility and focus on clinical practice and (3) fostering leadership commitment.

Increasing awareness about the need to improve the quality of care by providing data on clinical performance

According to the clinicians, managers and inspection teams, the discrepancy between guidelines and clinical practice was in part caused by the heterogeneous nature of the group of patients with sepsis and by how sepsis can manifest itself through various symptoms. They explained that deciding the course of the patient care is challenging, that the clinical processes of diagnosing and treating sepsis is complex and that judgements often are being made under quite stressful conditions.

A point that was clearly made during the interviews was that the hospitals lacked systems to monitor the extent to which diagnosis and treatment complied with desired practice and procedures. Though data is entered into patients' electronic health records from the time the patients are admitted to the hospitals, the information is not structured in a way that is easily aggregated so that the hospital can track the performance statistically over time.

One of the members of the inspection team at hospital C, who had long experience from leading system audits, told that this was the first time she had dared to state that an inspection had saved lives. She pointed to the systematic collection and analysis of patient data as the main reason for why the inspection had made a difference:

> I think what makes a difference, and impacts very strongly, is simply that we have measured, that we have systematised the findings from the electronic health records, (and) presented this using bar charts. The hospital employees were deeply affected by seeing these data. Across-the-board everyone thought they were very good and (in reality) no one were up to the mark.

Some clinicians found that, while they were not exceedingly surprised by the results, the data presented by the inspection team helped frame the challenges they experienced in their day-to-day activities. Describing how the efforts of improving the patient care had changed after the inspection, a clinician from hospital A referred to how the attention to completing diagnostic procedures quickly increased after the inspection results were presented. It made them 'see through other's eyes' what they already knew:

> After the inspection, and after (one of the managers) presented the findings in the auditorium, (the diagnostic work) got a lot more focused. It was nice because in a way… we saw through others eyes what we in reality knew, and then we focused on that work in a whole other way. So these patients have been given much better treatment after the inspection, compared to before.

Having performance data presented by the inspection team can help managers and clinicians re-evaluate their own experiences and assessment of clinical performance. The inspection team of hospital B described how their presentation of data in a closing meeting at one of the hospitals had encouraged the participants at the meeting to share and discuss recent experiences of challenges in the emergency department:

> We just displayed our own data, but (the managers and clinicians) brought it up on the agenda. And then someone just pointed out: "We heard that there was a surge of patients yesterday as well". We overheard that a discussion and a dynamic emerged that we could pitch into.

Building acceptance for improvement through professional credibility and focus on clinical practice

Professional credibility was a topic that was underscored by inspection teams, clinicians and managers. The clinicians and managers expected the inspection teams to include professionals with medical background, and they expected the inspection team to have insight into the requirements and practices of acute functions in hospitals. A manager at hospital A argued that the inclusion of medical experts was important for the legitimacy of the findings from the inspections:

> It is crucial that there is someone (on the inspection team) who comes from clinical practice, and possibly also from clinical research, and sort of knows the details of the issues that they enquire into; and who also is going to have an understanding of what the management component of these issues might be. So I think this is crucial for the legitimacy of this inspection.

The inspection teams also shared this view, that the medical experts' knowledge of sepsis care and experience with the day-to-day operations of emergency departments enhanced the legitimacy of the inspections.

Clinicians and managers stressed the need for the inspection teams to have a clear understanding of the work processes in emergency departments. By focusing on how the different processes were interconnected, the inspections identified system-level weaknesses that could produce barriers to timely diagnosis and treatment. One of the managers from hospital D pointed out that one of the strengths of the inspection had been how these findings were related to issues critical to patient care:

> The direct effect of the inspection is obvious. In this case one can relate it directly to the patient, even though much is related to systems and how systems are in place to take care of patients presenting with sepsis. But (the inspection) is very efficient, benefiting the patient directly.

A factor that both clinicians and managers pointed out across interviews is that diagnosing and treating sepsis patients involve several different organisational subunits within the hospitals. As such, there are very real organisational hurdles that need to be overcome in order to achieve the desired improvement in clinical performance. The inspection teams' understanding of complicated care processes was especially important because it enabled them to direct the inspection on how different groups of clinicians worked together. This forced the different organisational subunits to take a more birds-eye view of the patient care processes as a whole. A manager from hospital B explained:

> I believe that it is positive that someone comes from the outside and then points out that you have to have these things up and running. Because […] the workday is so hectic that every department is preoccupied with themselves and their work […] And I think that (the inspection) is a good pry tool, because then we have to cooperate between departments. And you could say that as a hospital we should be able to do this of our own volition, but this has turned out to be difficult.

Fostering leadership commitment

Because of the challenges of making improvements across different subunits within the hospital, hospital management had an important role in the improvement efforts. In this context, leadership commitment refers to the whole chain of command from the executive director on top to the senior nurses in the emergency department.

Both clinicians, managers and the inspection teams argued that without bringing the clinical managers and leaders on board and making sure that they were invested in this work, it would be exceedingly difficult to achieve successful improvement of the patient care. When discussing experiences with the improvement initiatives that started up after the inspections, a clinician at hospital D commented on the role of managers:

> Of course they nag a bit, but often because they want to get better. They are genuinely concerned with the medical issues, and that makes one want to join in.

Similarly, one of the clinicians at hospital C pointed out that it was important that clinical managers were genuinely interested in the improvement efforts:

> The clinical managers are actually interested in putting much effort into it, ensuring that one has resources, and that time is allocated to this. And in a way … they join in and look at the results of what is being presented. […] And this holds true both for nurses and for doctors; that one gets motivated to continue working (with improvements) and feel a bit acknowledged for the work one does.

An important function of the inspections was how they precipitated communication between different leadership levels on matters related to patient care. A clinician from hospital B described how the inspection report affected the hierarchy from clinic to department, and how this caused ripple effects throughout the organisation:

> An inspection makes an impact on the management. The head of clinic just said: "This is not good, this is not good enough. Now; who takes care of what? Now we have to do something different." And the head of department joins in. The heads of departments talk together and in a way you get a whole organization joining … This is clearly an effect of the inspection; from the top management and downwards. It feels more momentous: Here we need to do something, to close the nonconformities, we need to … And this has yet more ripple effects. So in that sense, (the inspection) has major consequences, in my opinion.

Facilitating communication networks that also included the managerial level was reported to be an important part of achieving organisational commitment to the issues of the inspection. The inspection facilitated that a large group of decision makers came together to discuss issues related to patient care.

In the period following the initial report from the inspection, hospitals are expected to develop a response and action plan to the NBHS. Many interviewees explained that this was an occasion for mutual learning between different disciplines and different hierarchies of management. A manager from hospital A argued:

> Almost nothing happens one-to-one, right? It happens across supporting professions or laboratory professions and radiology and shift teams and positions. So to get some of this reciprocity in the learning process we have tried bringing together these groups and develop a common response (to the NBHS inspection report).

DISCUSSION

In this study, we set out to explore how inspections may foster clinical improvements in hospitals. The first theme we identified was related to how the inspections provided data on the quality of care for patients with sepsis. Our findings suggest that by providing these data, the inspection promoted increasing awareness of clinical performance.

Second, we found that there was a need for inspection teams to have a clear understanding of the clinical work and of work processes in the emergency department. Without such knowledge, the legitimacy of the inspection would suffer, and the inspection would be rendered ineffective as a tool for systemic improvements. By directing attention to the interdependencies of the care processes, the inspection could help the hospital to target their efforts on improving the clinical system as a whole.

Lastly, the hospital management seems to be the main conduit through which the inspection team can affect the hospital's work on improving a clinically complex task such as sepsis management. Not only do inspection teams engage managers directly they also play a role in opening up channels of communication between clinical and top-level management and leadership. External inspections could therefore create arenas for discussion and interprofessional reflection between different levels of management on how the hospital as a whole could improve their services to patients with sepsis.

Strengths and limitations

The findings and interpretations of this study are intrinsically linked to the organisational and procedural context in which they are being held. Inspections are complex interventions. Reviewing their effects, we need explanatory analyses that bring to bear both theoretical and practical understanding of the intervention and the contexts within which it is being implemented.[28] The generalisability of the findings should be judged accordingly. We have purposively chosen to study the experiences of actors involved in presumptively successful inspections within a clinically demanding field of patient care. If we had selected less successful cases or studied inspections of another type of theme, for instance administrative tasks, one could expect our findings to diverge substantially. It is also worth noting that the selection of successful inspections was based on disease-specific indicators. Therefore, we do not know whether the inspections had any significant effect on hospital-level performance.[29]

Our focus on change mechanisms related to improvements in quality of care also implies that we have not explored potential costs and adverse side-effects of the inspections. Inspections may impose compliance costs on regulated organisations, including costs related to handling requests for information, consulting the inspection team and acting as guides on site-visits.[30] If the organisation frequently receives inspections, inquiries or instructions from different regulatory bodies, such costs might add up to a substantial strain, especially on the management and administrative staff. This study should therefore not be considered an exhaustive evaluation of the benefits and disadvantages of the sepsis inspections or inspections in general.

Furthermore, we do not argue that the aspects highlighted in this study are the only mechanisms that might be set in motion during an inspection process. One line of argument worth mentioning in this respect is that the prospect of being inspected in itself can initiate improvement efforts.[3 31] Though the search for such anticipatory effects is an important avenue of research, the focus of this study has been on how the findings and recommendations from the inspections, and the interaction with the inspection teams, might influence the hospitals' improvement efforts.

Interpretation in relation to previous studies

Our analyses echo previous research regarding how inspections with a patient-centred focus might promote awareness among clinicians and managers.[32] Furthermore, our analyses lend support to studies highlighting how using data in external assessments of quality of care can help hospitals track improvement.[33] Providing measurable data seems especially pertinent in the case of the sepsis inspections, as previous studies have shown the importance of performance metrics in fostering change in clinical behaviour in care for patients with sepsis.[34]

Some authors have argued that if external assessment schemes lead to increased use of data, they do so primarily through a strengthening of the bureaucratic control in the organisation.[35] We, however, found that the quality metrics were not considered as being solely within the purview of bureaucratic control; the professionals in the organisation viewed the use of data as a necessity for improving quality.

Our analyses nonetheless show that clinical leads played a key role in any improvement effort. Making leaders commit to improving patient care was seen as a *sine qua non* for the inspections to succeed. While this supports an argument for seeing external assessments as a platform from which clinicians can negotiate with senior management,[36] we would add that inspections might empower leaders and managers as well as clinicians.[37] Some important ways in which leaders wield power within organisations are by calling on shared organisational values and by leveraging facts and reasoning.[38] Clinical leaders can facilitate change processes and organisational learning by providing front-line clinicians with an arena for sharing information and a context for reflecting on shared information.[39] The effectiveness of such leadership approaches can be bolstered by the inspections. The sepsis inspections highlighted patient safety, which is a laudable and legitimate shared value goal in the emergency departments, and they did so by providing tangible facts for the leaders to leverage vis-a-vis their subordinates and team members.

Recent research has found that educative approaches to regulation can succeed when regulators are able to leverage existing norms and accountability structures in the regulated community.[40] This seems to be the case for the sepsis inspection. They have resulted in an improved understanding of the inherent complexities in the care of patients with sepsis, and the improved understanding brings forth organisational commitment and readiness for change, which are pivotal for improvement to take place. These processes also parallel findings from a study of professionals' motivation in hospital accreditation, which showed that external assessment opened up opportunities for collaborative learning and promoted understanding of the whole organisation across organisational boundaries.[41] Similarly, the importance of the system perspective runs like a red thread through our interviews, both in terms of the inspection teams' competencies and in terms of how clinicians and managers address quality challenges in their own organisations.

It should be noted that this argument presupposes the existence of norms and accountability structures in the inspected organisation that can be harnessed for quality improvement. If the management and staff are not amenable to the inspection team's suggestions, the learning process will likely flounder. Whether the organisation responds to the inspection with organisational commitment is not only dependent on which organisation is being inspected but also on the theme of the inspection. The way the clinical, patient-centred focus provided a legitimisation for the sepsis inspections is a case in point.

Other contextual factors are also important. If the healthcare organisation already performs at a high level, the inspection might not be able to contribute significantly to further improvement.[29] Furthermore, healthcare organisations often require financial resources to initiate improvement efforts, and in some cases they also need external improvement support.[3 29] Consequently, our findings cannot be extrapolated as universally applicable for all types of inspections within all types of organisations.

Policy implications

Even if performance data is key, focusing exclusively on performance data and quantifiable targets might pose a risk by underestimating the measurement problems or risks of health organisations gaming the system.[42] There is a risk that externally imposed standards in external assessment schemes may end up being perceived as a 'tick-box' exercise for the clinicians involved.[43]

When assessing performance within a specific area of patient care, the inspection authorities should use indicators that carry a clinical relevance for those working in the inspected organisations. To achieve this, they need to operationalise clinical standards into indicators that are well-suited for identifying subpar services and sensitive for improvement. It is also necessary to combine the evaluation of the indicators with a thorough understanding of the clinical processes at work. The task of the inspectors is to review the numbers and bring to the table an assessment of why the hospital might fail to meet the standards. This might necessitate prioritising regulatory resources so that external clinical experts are extensively involved both in the preparation stages, when relevant indicators are identified, and during the on-site inspections.

Organisations do of course review their own performance data and make efforts to improve without the help of external inspections. When it is feasible to make improvements through smaller adjustments, it is likely that the hospitals will do so. Addressing the underlying challenges inherent in tasks like sepsis diagnosis and treatment, on the contrary, entails both deeper analysis and more profound systemic changes. Here, the clinical data and assessments provided by the inspection team can be of great value for the management and staff in their search for flexible solutions for quality improvement. Here, however, we also see the limits of this approach to inspections: For the inspection to succeed, the organisation must have sufficient personnel and resources that can be mobilised for a sustained commitment to quality improvement.

Acknowledgements We thank the focus group participants for devoting their time to this study. We also thank the patient organisations for their involvement in the research project, the inspection teams for participating in data collection and the staff at the Norwegian Board of Health Supervision for developing the sepsis inspection. Professor Geir Sverre Braut translated the inspection report, for which we are most grateful.

Contributors GH contributed to conception and design, planning and conducting interviews, initial analysis, interpretation of findings, authoring original draft, making critical revisions and approving the final version of the manuscript. ILT contributed to conception and design, interpretation of findings, making critical revisions and approving the final version of the manuscript. JCF contributed to conception and design, interpretation of findings, making critical revisions and approving the final version of the manuscript. GTB contributed to conception and design, interpretation of findings, making critical revisions and approving the final version of the manuscript. EH contributed to conception and design, planning and conducting

interviews, initial analysis, interpretation of findings, making critical revisions and approving the final version of the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The research project was reviewed and approved by the Regional Ethics Committee of Norway North (Identifier: 2015/2195) and the Norwegian Data Protection Authority (Identifier: 15/01559).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. This is a qualitative study and therefore the data generated is not suitable for sharing beyond that contained within the report. Further information can be obtained from the corresponding author.

ORCID iDs
Gunnar Husabø http://orcid.org/0000-0002-8726-1769
Einar Hovlid http://orcid.org/0000-0003-0300-4405

REFERENCES

1 Shaw CD, Groene O, Berger E. External institutional strategies: accreditation, certification, supervision. In: Busse R, Klazinga N, Panteli D, eds. *Improving healthcare quality in Europe*, 2019: 203–31.
2 Sutherland K, Leatherman S. *Regulation and quality improvement - A review of the evidence*. London: The Health Foundation, 2006.
3 Smithson R, Richardson E, Roberts J, *et al*. *Impact of the care quality Commission on provider performance: room for improvement?* 2018.
4 Schaefer C, Wiig S. Strategy and practise of external inspection in healthcare services – a Norwegian comparative case study. *Safety in Health* 2017;3:1–9.
5 Weiner BJ, Amick H, Lee S-YD. Conceptualization and measurement of organizational readiness for change: a review of the literature in health services research and other fields. *Med Care Res Rev* 2008;65:379–436.
6 Nonaka I. A dynamic theory of organizational knowledge creation. *Organization Science* 1994;5:14–37.
7 Oude Wesselink SF, Lingsma HF, Reulings PGJ, *et al*. Does government supervision improve stop-smoking counseling in midwifery practices? *Nicotine Tob Res* 2015;17:572–9.
8 Oude Wesselink SF, Lingsma HF, Ketelaars CAJ, *et al*. Effects of government supervision on quality of integrated diabetes care: a cluster randomized controlled trial. *Med Care* 2015;53:784–91.
9 Castro-Avila A, Bloor K, Thompson C. The effect of external inspections on safety in acute hospitals in the National health service in England: a controlled interrupted time-series analysis. *J Health Serv Res Policy* 2019;24:182–90.
10 Allen T, Walshe K, Proudlove N, *et al*. Measurement and improvement of emergency department performance through inspection and rating: an observational study of emergency departments in acute hospitals in England. *Emerg Med J* 2019;36:326–32.
11 Flodgren G, Gonçalves-Bradley DC, Pomey M-P. External inspection of compliance with standards for improved healthcare outcomes. *Cochrane Database Syst Rev* 2016;12:Cd008992.
12 Hovlid E, Braut GS, Hannisdal E, *et al*. Mediators of change in healthcare organisations subject to external assessment: a systematic review with narrative synthesis. *BMJ Open* 2020;10:e038850.
13 Hovlid E, Frich JC, Walshe K, *et al*. Effects of external inspection on sepsis detection and treatment: a study protocol for a quasiexperimental study with a stepped-wedge design. *BMJ Open* 2017;7:e016213.
14 Morgan DL. *The focus group Guidebook*. Thousand Oaks, CA: SAGE Publications, 1997.
15 Sayer A. *Realism and social science*. Thousand Oaks: Sage, 2000.
16 O'Brien BC, Harris IB, Beckman TJ, *et al*. Standards for reporting qualitative research: a synthesis of recommendations. *Acad Med* 2014;89:1245–51.
17 Norwegian Board of Health Supervision. Introduction to the supervisory authorities and the supervision of child welfare services, social services and health and care services in Norway, 2019. Available: https://www.helsetilsynet.no/en/introduction-to-the-supervisory-authorities-and-the-supervision-of-child-welfare-services-social-services-and-health-and-care-services-in-norway
18 Goodwin A, Srivastava V, Shotton H, *et al*. *Just say sepsis. A review of the process of care received by patients with sepsis*. London: National Confidential Enquiry into Patient Outcome and Death, 2015.
19 Rhodes A, Evans LE, Alhazzani W, *et al*. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. *Intensive Care Med* 2017;43:304–77.
20 Plata-Menchaca EP, Ferrer R, Ruiz Rodríguez JC, *et al*. Antibiotic treatment in patients with sepsis: a narrative review. *Hosp Pract* 2020;319:1–11.
21 Gatewood Medley O'Keefe, Wemple M, Greco S, *et al*. A quality improvement project to improve early sepsis care in the emergency department. *BMJ Qual Saf* 2015;24:787–95.
22 Husabø G, Nilsen RM, Flaatten H, *et al*. Early diagnosis of sepsis in emergency departments, time to treatment, and association with mortality: an observational study. *PLoS One* 2020;15:e0227652.
23 ISO. *ISO/IEC TS 17022:2012. Conformity Assessment - Requirements and Recommendations for Content of a Third-Party Audit Report on Management Systems*. Geneva: International Organization for Standardization, 2012.
24 National Institute for Health and Care Excellence. *Sepsis. quality standard QS161*, 2017.
25 Dellinger RP, Levy MM, Rhodes A, *et al*. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock, 2012. *Intensive Care Med* 2013;39:165–228.
26 Husabø G, Nilsen RM, Solligård E, *et al*. Effects of external inspections on sepsis detection and treatment: a stepped-wedge study with cluster-level randomisation. *BMJ Open* 2020;10:e037715.
27 Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006;3:77–101.
28 Pawson R, Greenhalgh T, Harvey G, *et al*. Realist review--a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy* 2005;10 Suppl 1:21–34.
29 Castro AC. *Evaluating the effect of external inspections of the care quality. Commission of acute NHS hospitals in England*. University of York, 2018.
30 Hood C, Scott C, James O, *et al*. *Regulation inside government: waste watchers, quality police, and sleaze-busters*. Oxford: Oxford University Press, 1999.
31 Walshe K, Phipps D. *Developing a strategic framework to guide the Care Quality Commission's programme of evaluation*. London: CQC, 2013.
32 Åsprang AF, Frich JC, Braut GS. Organizational impact of governmental audit of blood transfusion services in Norway: a qualitative study. *Transfus Apher Sci* 2015;53:228–32.
33 El-Jardali F, Jamal D, Dimassi H, *et al*. The impact of hospital accreditation on quality of care: perception of Lebanese nurses. *Int J Qual Health Care* 2008;20:363–71.
34 Levy MM, Rhodes A, Phillips GS, *et al*. Surviving sepsis campaign: association between performance metrics and outcomes in a 7.5-year study. *Crit Care Med* 2015;43:3–12.
35 Paccioni A, Sicotte C, Champagne F. Accreditation: a cultural control strategy. *Int J Health Care Qual Assur* 2008;21:146–58.
36 Baskind R, Kordowicz M, Chaplin R. How does an accreditation programme drive improvement on acute inpatient mental health wards? an exploration of members' views. *J Ment Health* 2010;19:405–11.
37 Walshe K. *Regulating healthcare: a prescription for improvement?* Berkshire: McGrawHill Education, 2003.
38 Van Wart M, Suino PArmonk NY, ed. *Leadership in public organizations : an introduction*. 2nd ed. London: M.E. Sharpe, 2012.
39 Hovlid E, Bukve O, Haug K, *et al*. Sustainability of healthcare improvement: what can we learn from learning theory? *BMC Health Serv Res* 2012;12:235.

40 Erp J, Wallenburg I, Bal R. Performance regulation in a networked healthcare system: from cosmetic to institutionalized compliance. *Public Adm* 2020;98:46–61.
41 Greenfield D, Pawsey M, Braithwaite J. What motivates professionals to engage in the accreditation of healthcare organizations? *Int J Qual Health Care* 2011;23:8–14.
42 Bevan G, Hood C. What's measured is what matters: targets and gaming in the english public health care system. *Public Adm* 2006;84:517–38.
43 Campbell SM, Chauhan U, Lester H. Primary medical care provider accreditation (PMCPA): pilot evaluation. *Br J Gen Pract* 2010;60:e295–304.

Appendices

Study 1

Study 2

Study 3

ICD 10 codes used to search the National Patient Register

A02.1 Salmonella sepsis

A20.7 Septicemic plague

A21.7 Generalized tularemia

A22.7 Anthrax sepsis

A24.1 Acute and fulminating melioidosis

A26.7 Erysipelothrix sepsis

A32.7 Listerial sepsis

A39.2 Acute meningococcemia

A39.3 Chronic meningococcemia

A39.4 Meningococcemia, unspecified

A40 Streptococcal sepsis (0.,1.,2.,3.,8.,9.)

A41 Other sepsis (0.,1.,2.,3.,4.,5.,8.,9.)

A42.7 Actinomycotic sepsis

A46 Erysipelas

A48.3 Toxic shock syndrome

A54.8 Other gonococcal infections

O85 Puerperal sepsis

B00.7 Disseminated herpesviral disease

J09 Influenza due to certain identified influenza viruses

J10 Influenza due to other identified influenza virus

J13 Pneumonia due to Streptococcus pneumoniae

J14 Pneumonia due to Hemophilus influenzae

J15 Bacterial pneumonia, not elsewhere classified

J36 Peritonsillar abscess

J39 Other diseases of upper respiratory tract

J85 Abscess of lung and mediastinum

J86 Pyothorax

K65 Peritonitis

K81 Cholecystitis

M72.6 Necrotizing fasciitis

N10 Acute pyelonephritis

R57 Shock, not elsewhere classified

R65 Symptoms and signs specifically associated with systemic inflammation and infection

Missing data and imputation

Several of the variables in our data set had missing values. The teams that collected data were meticulous in their recording of data from the original patient records, and they would routinely verify that they had included all available data from the patient records. Hence, missing data were indicative of the information never being included in the patient record, rather than failure to collect a complete set of data from the records.

Time to administration of antibiotics had been registered in two separate variables, both in minutes and in hours (measured as prior to admission, within 1 hour, between 1 and 2, between 2 and 3, between 3 and 4, and after more than 4 hours). Time to antibiotic administration measured in hours had fewer missing observations (11%) than time to antibiotic administration measured in minutes (42%), and no observations that had information on minutes to antibiotic treatment lacked information on hours to antibiotics. For the regression analyses, we chose to exclude the observations that had no information on antibiotic treatment. We also excluded observations where antibiotics were administered prior to admittance (5%), or where the patients' records indicated that they should not receive antibiotic treatment (2%). In table 2 in the article, these two last groups were subtracted from total n when reporting the proportion of patients receiving antibiotic treatment. The total n included, however, patients with missing data on time to antibiotic treatment.

The missing data on the other variables used in the regression analyses could potentially lead to bias and loss of power in the analyses. To alleviate these problems, we decided on a strategy of combining recoding and multiple imputation of missing data.

For two of the diagnostic process variables that had missing data in the original data material, blood lactate and adequate observation regimen, we coded missing values as 0, i.e. incomplete performance of the process. We did this because information on these procedures

should be documented in the patient record. Lack of documentation therefore indicated that the procedures were not completed according to guidelines.

For four variables, we imputed missing data: Time to antibiotics in minutes (42% missing), time to examination by a physician (35% missing), time to triage (12% missing), and organ failure (3% missing). Imputation is recommended in cases where data is *missing at random* (MAR). Data are MAR when missingness in the relevant variable only depends on data from variables that we have observed and can control for. Assuming missing data to be MAR, we created 100 imputed datasets by using the fully conditional specification algorithm [1], a method for running a series of regression analyses, one for each variable to be imputed. To obtain pooled estimates with 95% confidence intervals across the imputed datasets, we used Rubin's combination rules as implemented in the *mi suite of commands* in Stata.

We chose the imputation model according to the distribution of the variable to be imputed. For our primary variable of interest for the regression analyses, time to antibiotic administration in minutes, we used interval regression in the imputation model. Upper and lower bounds for imputed values were set based on information about time to antibiotic administration in hours. We imputed the other three variables using logit models. Each imputation model included the other imputation variables as predictor variables, along with the rest of the variables which were to be used in the regression analyses: age, admission year, Charlson comorbidity index, blood lactate measured within one hour, adequate observation regimen, all-cause mortality after 30 days. In addition, we added hours to antibiotic treatment as an auxiliary variable.

Fitting the linear model

We performed sensitivity tests by omitting some of the most extreme values for time to treatment with antibiotics; these were six observations which had more than 1000 minutes from admission to administration of antibiotics. This adjustment proved to have an impact on model fit and individual coefficient estimates, so we chose to omit these six outliers in the final models.

In addition to testing the models with and without the outlier variables, we also tested modeling the analyses using standard Poisson and Negative binomial regression, as well as fitting an identical linear regression model using a log transformed outcome variable. The log transformed model had slightly better fit than the linear model without the six extreme values, but the two models showed similar effect sizes and p-values. We chose to report the results of the linear analysis in the paper because the linear model provides predictions of delay in minutes for each predictor variable, rather than coefficients that need to be exponentiated to provide information on delay in relative terms.

In all analyses, we first analyzed the crude association between outcome and exposure variables, and them we ran the analyzes again, controlling for age, organ failure, comorbidity, and time to admission. We also performed sensitivity analyses by running the regression analyses for the subgroup of patients who were triaged to code red or orange, i.e., a subgroup of patients who needed expedient attention. This analysis found that, when controlling for the other potential confounders, the associations between timely diagnostic tests and time to antibiotic treatment in the subgroup of patients triaged as orange or red were similar to the associations we found in the patient group as a whole.

Estimates from logistic regression

After fitting the logistic regression model, we obtained predictive margins with confidence intervals using the user written Stata command *mimrgns* [2], requesting conditional predictions for each 15 minutes from zero to nine hours. We re-calculated these margins into predictions of probabilities for outcome 1 (death within 30 days) using logit transformation. Subsequently, we imported the recalculated predictive margins into Stata, along with mean mortality rates by hour to antibiotics. We then combined a bar graph for the empirical mortality rates and a line graph for the predicted probabilities into one figure, using Stata's *twoway* command.

References

1. Buuren S. Flexible Imputation of Missing Data, Second Edition: Chapman & Hall/CRC; 2018.
2. Daniel K. MIMRGNS: Stata module to run margins after mi estimate. S457795 ed: Boston College Department of Economics; 2014.

Sub-analyses – patients with organ failure

We replicated the analyses of association between diagnostic procedures and time to treatment and between time to treatment and mortality on the sub-group of patients with organ dysfunction (Table S3.1 and Fig S3.1, below).

Table S3.1. Linear regression for factors associated with delay in antibiotic treatment. Patients with organ failure.

	Unadjusted	**Model 1***	**Model 2†**
	b (95% CI)	*b (95% CI)*	*b (95% CI)*
Not triaged within 15 minutes	53.5 (21.4 to 85.7)	56.1 (26.2 to 86.0)	17.1 (-11.1 to 45.3)
Examination by physician not in accordance with priority	64.2 (34.8 to 93.5)	67.4 (39.1 to 95.7)	47.2 (18.5 to 75.9)
Lactate not measured within 1 hour	78.5 (56.9 to 100.2)	84.3 (65.2 to 103.3)	25.3 (8.6 to 42.0)
Inadequate observation regimen	42.8 (21.4 to 64.2)	42.3 (22.0 to 62.6)	69.1 (48.6 to 89.5)

Outcome variable: Time to antibiotics measured in minutes. n=487

* Adjusted for patient age, comorbidity, and time to admission

† Adjusted for the other variables in this table, and age, comorbidity, and time to admission

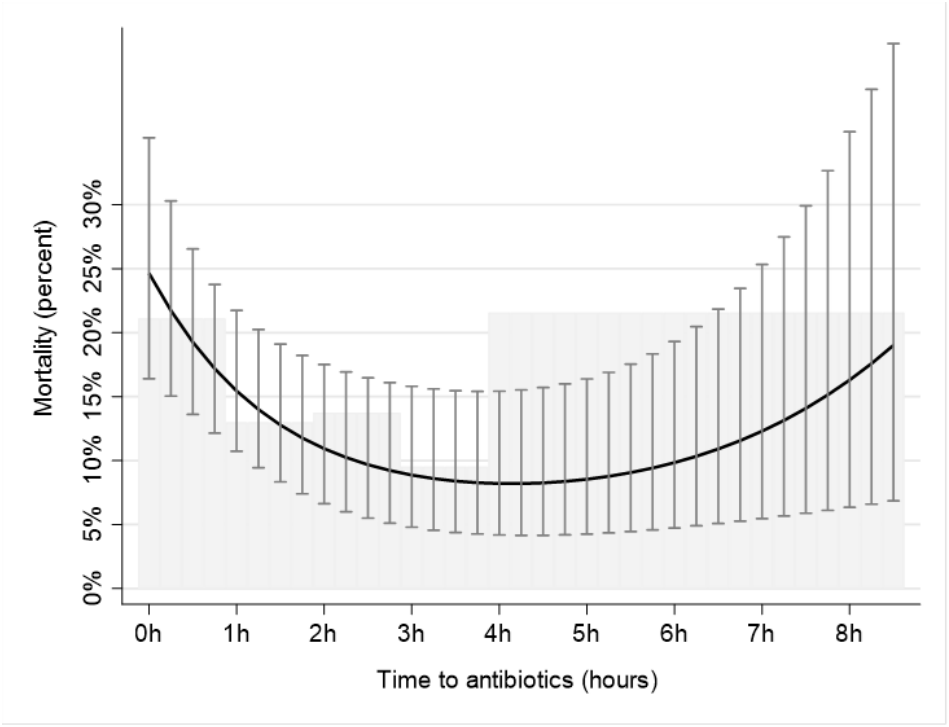
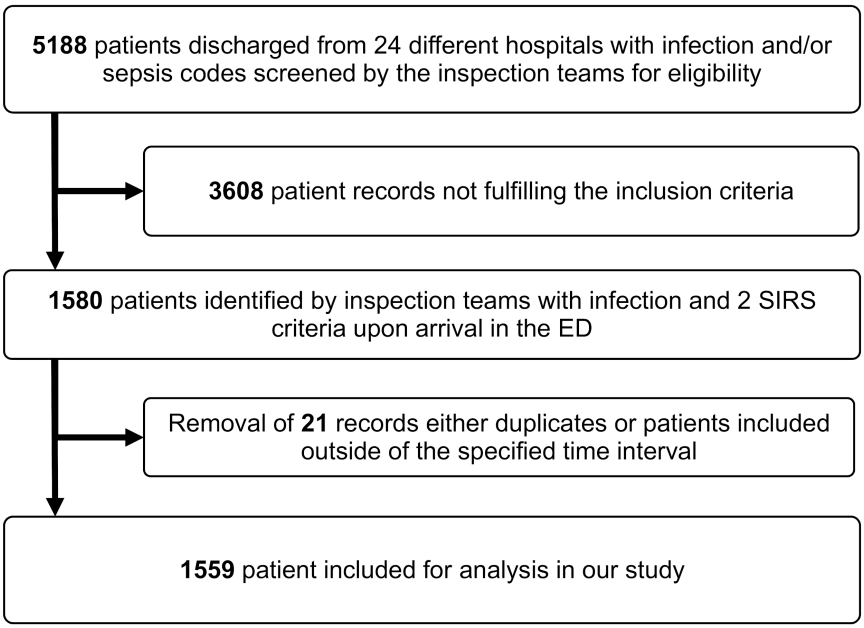


Fig S3.1. All-cause 30-day mortality by time to antibiotic treatment. Patients with organ failure.

Gray shaded histogram represents mortality rates according to time to antibiotic treatment in hours. Solid black curve with bars represents model-predicted mortality rates with 95% confidence intervals according to time to antibiotic treatment in minutes using logistic regression models, adjusted for patient's age, date of admission, comorbidity, and presence of organ failure. Date of admission was measured using calendar days since study start, entered as a polynomial function with first (b -.0113 $p < .01$), second (b 2.8e-5 $p < .05$) and third degree (b -1.37e-8 $p < 0.1$) variables. The model prediction uses average values for adjustment values. N=488.


5188 patients discharged from 24 different hospitals with infection and/or sepsis codes screened by the inspection teams for eligibility
3608 patient records not fulfilling the inclusion criteria
1580 patients identified by inspection teams with infection and 2 SIRS criteria upon arrival in the ED
Removal of **21** records either duplicates or patients included outside of the specified time interval
1559 patient included for analysis in our study

Correction

After publication of study 1, we found an error in a caption text. We have submitted a request for corrections to the journal.

There is an error in the caption of figure 2.

The caption says:

“Gray shaded histogram represents mortality rates according to time to antibiotic treatment in hours. Solid black curve with bars represents model-predicted mortality rates with 95% confidence intervals according to time to antibiotic treatment in minutes using logistic regression models, adjusted for patient’s age, date of admission, comorbidity, and presence of organ failure. Date of admission was measured using calendar days since study start, entered as a polynomial function with first (b -0.011 p<0.001), second (b 2.5e-5 p<0.001) and third degree (b -1.2e-8 p<0.01) variables. The model prediction uses average values for adjustment values.”

The error is in the next to last sentence. Date of admission was not measured using calendar days and it was not entered as a polynomial function. Date of admission was entered as year, and it was time to antibiotics (measured in minutes), that was entered as a polynomial function.

The caption should read:

“Gray shaded histogram represents mortality rates according to time to antibiotic treatment in hours. Solid black curve with bars represents model-predicted mortality rates with 95% confidence intervals according to time to antibiotic treatment in minutes using logistic regression models, adjusted for patient’s age, year of admission, comorbidity, and presence of organ failure. Time to antibiotics was measured in minutes, entered as a polynomial function with first (b -0.011 p<0.001), second (b 2.5e-5 p<0.001) and third degree (b -1.2e-8 p<0.01) variables. The model prediction uses average values for adjustment values.”

The same error is repeated in Fig S3.1. in the supplemental file “S3 File”, which uses the same type of figure for a sub analysis.

This caption should read:

“Gray shaded histogram represents mortality rates according to time to antibiotic treatment in hours. Solid black curve with bars represents model-predicted mortality rates with 95% confidence intervals according to time to antibiotic treatment in minutes using logistic regression models, adjusted for patient’s age, year of admission, comorbidity, and presence of organ failure. Time to antibiotics was measured in minutes, entered as a polynomial function with first (b -.0113 p < .01), second (b 2.8e-5 p < .05) and third degree (b -1.37e-8 p < 0.1) variables. The model prediction uses average values for adjustment values. N=488.”

The errors do not affect the overall conclusion or scientific understanding of the article.

The method section correctly describes the analysis model and the figure itself is correct. The effect sizes are also correct, but the text refers to the wrong variable.

The county governor of Troms

Report from inspection of sepsis treatment

in the emergency department at

University Hospital of Northern Norway, Tromsø

UNOFFICIAL TRANSLATION[1]

Address of the enterprise: 9030 Tromsø

Time span for the inspection: 6. September 2016 – 9. March 2017

Summary

Norwegian Board of Health Supervision (NBHS) has decided that in the period 2016-2017, there will be performed nationwide inspections of the hospitals' emergency departments and their work with recognition and treatment of patients with sepsis.

The county governor of Troms has performed a inspection designed as a system audit at the University Hospital of Northern Norway, Tromsø. This report describes the nonconformities identified within the audited areas. The system audit comprised the following themes:

Identification and initiation of treatment in the emergency department of patients with sepsis or suspected sepsis.

During the inspection we would investigate if the hospital ensures:

- adequate admission, registration and prioritisation (triage) of patients with sepsis or suspected sepsis at the time of admission to the emergency department
- adequate assessment and diagnosis of the patients during their stay in the emergency department
- adequate initiation of treatment of the patients in the emergency department
- adequate observation of the patients in the emergency department
- adequate preparation and discharge of the patients to other departments, supplemented by ordinations/plans for further observation and treatment

The inspection team has 66 health records of patients presenting to the emergency department with sepsis or suspected sepsis.

[1] This report is an unofficial translation of the original report from Norwegian Board of Health Supervision. The original report, along with the reports from the other sepsis inspections, is available on the NBHS website: https://www.helsetilsynet.no/tilsyn/tilsynsrapporter/?w=2016+Sepsis+i+somatiske+akuttmottak

At the inspection, three nonconformities were identified:

Nonconformity 1:

The majority of the patients with sepsis did not receive treatment with antibiotics within the time limits prescribed in nationwide guidelines and in the hospital's own goal statements. Patients with severe sepsis who had to wait more than one hour, did not receive adequate treatment.

Nonconformity 2:

The management has not ensured that there is sufficient medical competence available in the emergency department so that assessments and initiation of treatment of patients with sepsis can be performed within the time limits prescribed in nationwide guidelines and in the hospital's own goal statements.

Nonconformity 3:

The hospital management has been aware that patients with sepsis receive delayed treatment with antibiotics in the emergency department but has not implemented sufficient corrective actions.

Date: 9. March 2017

xxxxx
Lead Auditor

yyyyy
Auditor

1. Introduction

This report is written after a system audit at University Hospital of Northern Norway, Tromsø in the period 6. September 2016 – 9. March 2017. It is a part of a nationwide inspection performed in 2016-2017, and one of the planned inspections to be performed by the County governor of Troms this year. The county governors of Finnmark, Troms and Nordland have appointed a joint inspection team to perform the inspections in these counties.

The county governor is through section 2 of the act on governmental supervision of the health and care services given authority to perform inspections with the provision of health and care services.

The aim of a system audit is to evaluate if the enterprise by means of internal control meets the legal requirements. The audit encompassed the following themes:

- which actions were taken by the enterprise to disclose, correct and prevent infringement of the legal requirements relevant for the analysed issues
- if the prescribed actions were performed in practice and, if necessary, corrected
- if the prescribed actions are sufficient to ensure adherence to the legal requirements

A system audit is performed by analysis of documents, through interviews and by other investigations.

This report deals with the nonconformities identified at the system audit, and thus does not present a complete evaluation of the work of the enterprise relevant for the themes covered by the inspection.

- **Nonconformity** is lack of fulfilment of requirements given by or on basis of acts and regulations

The background for the decision to perform inspection of the sepsis treatment, is, i.a. that NBHS has received several reports according to the requirement [on reporting adverse events] in section 3-3 of the act on specialised health care about serious infections and sepsis, where detection of infection has been too late, and where there has been delayed initiation of treatment with antibiotics.

NBHS has established a research project to gain knowledge on how planned inspection can contribute to improving quality on health services. Data collected from patient files in this inspection will be used to evaluate the effect of inspection on the quality of the service. As part of the inspection and this project, we will perform sampling from relevant health records in 8 months and 14 months from now.

2. Description of the enterprise – particular conditions

The University Hospital of Northern Norway (UNN HF) serves a population of about 190.000 inhabitants and consists of three hospitals, respectively in Tromsø, Harstad and Narvik, in addition to Longyearbyen hospital on Svalbard. The main administrative centre of the hospital is located to Tromsø, and is led by the chief executive director.

The health enterprise is divided into nine clinics, among them the *clinic for acute medicine* and the *clinic for medicine*. Each clinic is led by a director who reports to the chief executive director.

The emergency department at UNN HF Tromsø is a department in the clinic for acute medicine. Head of department reports to the director of the clinic. Head of department is at the moment also acting director of clinic for the clinic for acute medicine. Head of the unit for acute somatic admissions is responsible for the nursing services in this unit and reports to the head of the department. There is a medical consultant, 60% of a full position, adhered to the unit for acute somatic admissions as a medical advisor.

The medical on-duty teams consist of an intern, first line and second line registrars, first line registrar for heart and pulmonary diseases and subspecialised consultants in the different parts of internal medicine. The first line registrar is available 24hrs, the second line registrar is available 8hrs-22hrs on week days and 9hrs-15hrs in the weekends. The intern is not available at night time. The intern shall confer with the second line registrar (or first line registrar) related to all investigated patients.

The physicians working in the unit for acute somatic admissions are employed at different parts of the clinic for medicine or the clinic for heart and lung diseases. All physicians in first line or second line duty are undergoing training as a specialist. Head of department/chief consultant of the department of gastrology and nephrology is responsible for planning the on duty scheme and for arranging regular meetings with the physicians on both levels.

RETTS (Rapid Emergency Triage and Treatment System) is used in the unit for acute somatic admissions. According to activity under algorithm 47 treatment with antibiotics shall be initiated within 1 hour after arrival of the patient.

3. Execution

The system audit consisted of the following activities:

Notice/information regarding the inspection was sent 6. September 2016.

Overview over documents presented by the enterprise is to be found in the chapter on Documents.

Analysis of patient files were performed 7. November 2016 and 5. January 2017.

Opening meeting was arranged 25. January 2017.

Interviews

15 persons were interviewed.

On site visit in the unit for acute somatic admissions was performed 25. January 2017.

Closing meeting was arranged 26. January 2017.

4. What the inspection comprised

In the inspection, we have investigated if the health enterprise governs and controls that patients admitted with sepsis or suspected sepsis are identified and treated according to the requirements laid down in the legislation related to health care.

The inspection was limited to the unit for acute somatic admissions, and activities that are planned and ordered from the unit for acute somatic admissions.

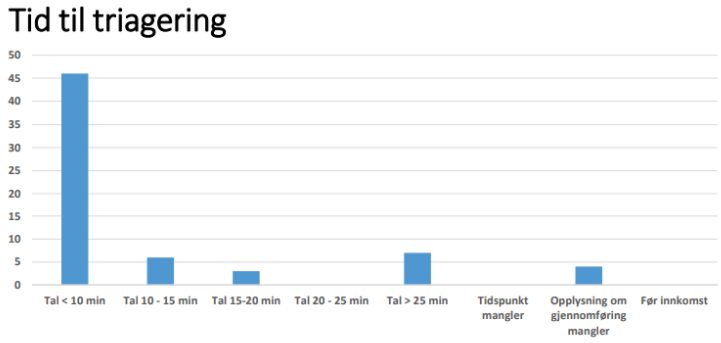
In particular we investigated if the University Hospital of Northern Norway had:

- prudent admission, registration and prioritisation (triage) of patients with sepsis or suspected sepsis at the time of admission to the emergency department
- prudent investigation and diagnosis of the patients during their stay in the emergency department
- prudent initiation of treatment of the patients in the emergency department
- prudent observation of the patients in the emergency department
- prudent preparation and transferral of the patients to other departments, supplemented by ordinations/plans for further observation and treatment

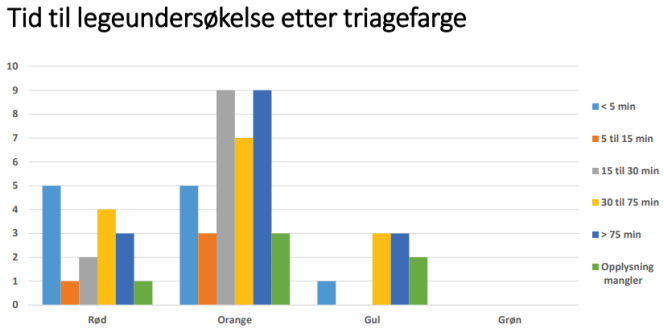
5. Findings

The inspection team has analysed patient files from patients admitted to the unit for acute somatic admissions with sepsis or suspected sepsis. The 66 patients included had an infection and fulfilled at least two of four SIRS-criteria. 33 patient files were from 1. October 2015 and immediately before (called P0), and 33 from 1. December 2016 and immediately before (called P1).

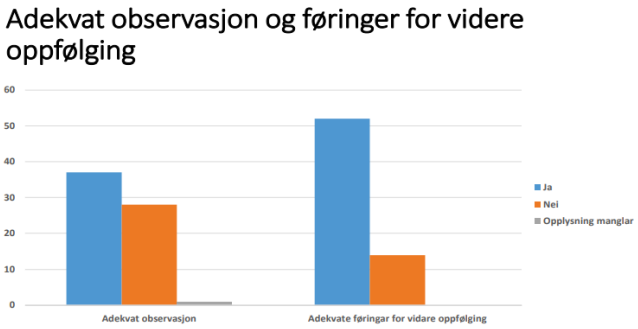
In the graphics below P0 and P1 are combined. The analysis showed:



(Time till triage, in minutes)



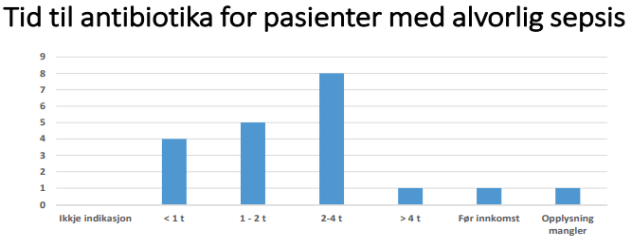
(Time until investigation by physician in minutes, according to triage colour)



(Adequate observation and instructions for further treatment, Yes (ja), No (nei), Lacking information (grey))

(Time till treatment with antibiotics in hours, all patients.

No indication, < 1 hr ….. > 4 hrs, Before admission, Lacking information)



(Time till treatment with antibiotics in hours, patients with severe sepsis.

No indication, < 1 hr ….. > 4 hrs, Before admission, Lacking information)

Three nonconformities were indicated.

Nonconformity 1:

The majority of the patients with sepsis did not receive treatment with antibiotics within the time limits prescribed in nationwide guidelines and in the hospital's own goal statements. Patients with severe sepsis who had to wait more than one hour, did not receive adequate treatment.

This is a deviation from the requirement in section 2-2 of the act on specialised health care and sections 6 to 9 in the regulation on governance and quality improvement in the health and care services.

Justification of this claim:

- The analysis of 66 patient files shoved that:
    - 9 of 16 patients with triage colour red were investigated by a physician more than 15 minutes after admission to the hospital
    - 24 of 49 patients with sepsis got their first treatment with antibiotics more than two hours after admission to the hospital
    - 9 of 18 patients with severe sepsis had to wait over two hours before treatment with antibiotics was initiated, 14 of 18 had to wait over one hour. One patient waited more than four hours
- None of the directors of the clinics (clinic for medicine and clinic for acute medicine) have determined specific routines or practice for treatment of sepsis in the unit for acute somatic admissions. Instead, there are several different, older versions of written procedures in Docmap. These are not known for the health personnel, and their status remains unclear. There is also a non-dated flow chart with unclear status. This is presented as wall charts in the unit for acute somatic admissions.
- The health personnel is unsure about which procedures that are currently valid and they have different opinions about if and when treatment with antibiotics shall be initiated.
- Inexperienced physicians use much time for investigating the patients and decide upon treatment with antibiotics. Front line physicians do not always get a go-signal to initiate treatment when searching for support on decisions, even when related to patients with sepsis that according to national guidelines should get treatment.
- The management of the hospital and the directors of the clinics (clinic for medicine and clinic for acute medicine) do not follow up if the hospital achieves the goal specifying that patients with sepsis should get treatment with antibiotics within one hour.
- Conflicts of simultaneity and problems with vacant beds in the unit for acute somatic admissions arise several times every week and this is leading to delayed initiation of treatment with antibiotics.
- Observation of vital parameters of patients with sepsis are not always documented after triage when the patient still is in the unit for acute somatic admissions.
- Physicians and nurses work to a low degree in teams related to the sepsis patients.
- The bed wards often have low capacity and need a long time before being able to accept new patients, and the intensive care unit for internal medicine is often full. This leads to congestion in the unit for acute somatic admissions of patients that are ready for transferral to a bed ward. The capacity of rooms thus is reduced, and leads to new patients with sepsis not always are investigated by a physician when the physician is available. This in turn leads to delayed initiation of treatment with antibiotics.
- The day of the on-site visit we were informed that a patient with severe sepsis had to wait three hours before initiation of treatment with antibiotics, and had to wait more than nine hours before transferral to a bed ward.

Nonconformity 2:

The management has not ensured that there is sufficient medical competence available in the emergency department so that assessments and initiation of treatment of patients with sepsis can be performed within the time limits prescribed in nationwide guidelines and in the hospital's own goal statements.

This is a deviation from sections 6 to 9 in the regulation on governance and quality improvement in the health and care services.

Justification of this claim:

- It is not planned for the physicians in the unit for acute somatic admissions to investigate and treat all patients in accordance with the national guidelines and the hospital's own goals, cfr. nonconformity 1.
- Interns in some occasions are left alone with a higher degree of responsibility than planned due to first line registrars are occupied with telephone calls from physicians outside the hospital and for distributing patients from the unit for acute somatic admissions to the bed wards of the hospital. The second line registrar often is occupied at the observation unit.
- Training of subordinate physicians in treatment of sepsis is failing, and characterised of lacking procedures for this activity.

Nonconformity 3:

The hospital management has been aware that patients with sepsis receive delayed treatment with antibiotics in the emergency department but has not implemented sufficient corrective actions.

This is a deviation from sections 8 and 9 in the regulation on governance and quality improvement in the health and care services.

Justification of this claim:

- Statistics and other instruments are scarcely used to follow up results and objectives.
- The management demands few data on results from the unit of acute somatic admissions, e.g. on waiting time for investigation by a physician and time till initiation of treatment with antibiotics.
- The health personnel has reported nonconformities related to delayed treatment of sepsis in the unit for acute somatic admissions but sufficient actions have not been taken.
- The chief executive officer as well as the directors of the clinics have been aware of the long waiting times for the patients in the unit for acute somatic admissions.
- It remains unclear who is responsible for developing av implementation of joint procedures for nurses and physicians in the unit for acute somatic admissions. The management scarcely has an overview of which procedures that are currently valid.

6. Evaluation of the system of governance of the enterprise

The management scarcely has an overview of which goals that are established for the treatment of sepsis in the unit for acute somatic admissions and if these goals are achieved. It remains unclear who is responsible for ensuring unambiguous procedures for treatment of sepsis unit for acute somatic admissions that is known for everyone. It is known for the management that patients risk to be waiting in the unit for acute somatic admissions to be transferred to a bed ward, but efficient actions have not been taken. The health enterprise thus has not arranged for the health personnel enabling them to take care of their duties in a way that ensures that patients with sepsis at the unit for acute somatic admissions are treated according to national guidelines and the hospital's own goals.

7. Legislation

- Act of 2. July 1999 no. 61 relating to specialised health care.
- Act of 2. July 1999 no. 64 relating to health personnel.
- Regulation of 21. December 2000 no. 1385 relating to patient files.
- Regulation of 28.October 2016 no 1250 relating to on governance and quality improvement in the health and care services.

8. Documentation

Documentation from the enterprise related to management of the services, provided by the enterprise during the preparation of the audit:

- Information in letter from the head of the unit dated 22. September 2016
- Organisational mapping for the health enterprise and the unit for acute somatic admissions
- Overview of physicians taking part in the on-duty scheme in the unit for acute somatic admissions
- Overview of first line and second line registrars, with information on length of service
- Overview of anaesthesiologists
- Overview of nurses in the unit for acute somatic admissions
- Overview of nurses functioning as coordinators in the unit for acute somatic admissions
- Work tasks for coordinator at the unit for acute somatic admissions in Tromsø
- Work tasks for responsible for the waiting room in Tromsø
- Work tasks for the triaging nurse at the unit for acute somatic admissions in Tromsø
- On-duty-order intern (FB1485)
- On-duty-order first line registrar (FB1484)
- On-duty-order second line registrar (FB1483)
- Admission of patients from the ambulance service.
- Algorithm 47 from the RETTS-manual
- Blood sampling routine sepsis

- Joint patient file for acute admissions UNN HF
- Flow chart treatment and monitoring at intermediary and/or intensive care units
- Transferral of patients with internal medical conditions from the unit for acute somatic admissions when lacking places at medical bed wards
- Procedure for handling of deviations UNN
- Copy of reports of deviations
- Minutes of meeting. Sepsis 1 – patient flow 11. April 2013
- Terms of reference, follow up of Sepsis 1 – 29. May 2013
- Minutes of meeting, Quality Commission UNN HF 3. June 2014
- Minutes of meeting, Quality Commission UNN HF 11. May 2016
- Plan for training for newly engaged health personnel in the units for acute somatic admissions and observations
- "Welcome to the physicians department, Clinic of medicine" (Valid from 9. December 2011)
- Check list newly engaged physicians (valid from 21. January 2013)
- Check list – joint plan for training for newly engaged employees in the units for acute somatic admissions and observations
- Agenda internal education internal medicine spring term 2016
- Agenda internal education internal medicine autumn term 2016

Documentation analysed during the inspection:

- Admission of adult patients with infection and suspected sepsis and serious sepsis/septic shock, common part (elaborated 8. February 2010)
- Admission of the patient with serious sepsis and septic shock (elaborated 11. January 2010)
- Admission of the patient with sepsis (SIRS score 2 or above and no symptoms of organic failure) (elaborated 4. March 2010)
- Placing [in bed wards] of patients with sepsis (elaborated 2. February 2010)
- Flow chart admission of adult patients with infection and suspected sepsis (19. February 2010)
- Sepsis-algorithm for physicians in in the unit for acute somatic admissions (valid from 28. October 2011)

Correspondence between the enterprise and the county governor:

- Notification of the inspection in letter dated 6. September 2016
- Documentation from the enterprise dated 22. September 2016
- Additional information/documentation from the enterprise in e-mail 31. October 2016, 4. November 2016 and 13. December 2016
- Agenda sent in letter dated 2. January 2017, revised 10. January 2017

9. Participants at the inspection

[In the original report participants are presented by name and position. Here only position is presented.]

In this table the participants from the enterprise and their type of participation is presented.

Function/position	*Opening meeting*	*Interview*	*Closing meeting*
Nurse, responsible for nuring development, unit for acute somatic admissions	x	x	x
Registrar, internal medicine	x	x	x
Specialist nurse, unit for acute somatic admissions		x	x
Nurse, unit for acute somatic admissions		x	x
Registrar, internal medicine	x	x	
Nurse, unit for acute somatic admissions	x	x	
Registrar, internal medicine		x	
Leading nurse, unit for acute somatic admissions	x	x	x
Consultant, infection medicine	x	x	
Consultant, unit for acute somatic admissions	x	x	x
Head of department, gastrology & nephrology	x	x	x
Director of clinic, medical clinic	x	x	x
Head of department & acting director of clinic (acutemedicine)	x	x	x
Deputy chief executive officer	x	x	x
Chief executive officer	x	x	
Director for quality and development	x		x
Deputy head of department, unit for acute somatic medicine			x

From the inspection authority these took part:

Chief county medical officer, lead auditor

Dep. chief county medical officer, auditor

Senior advisor, auditor

Advisor, auditor

Consultant (anaesthesiologist), medical auditor

Senior advisor, observer

Supplementary file: statistical analyses

Missing data

Table 1 shows the association between time to antibiotics (within 1 hour or after 1 hour) and the status of documentation (non-missing or missing) of the other four variables with missing data, blood culture taken prior to antibiotic treatment (5% missing), timely assessment by physician (24% missing), time to triage (11% missing), and time to fluid administration (29% missing). The table shows that for all the four variables, the proportions of patients with timely antibiotic treatment is larger within the group who has documented treatment than within the group of patients with missing documentation.

Table 1 Association between time to antibiotic treatment and status of documentation for variables with missing observations

		Antibiotics administered		
Diagnostic/treatment measure	**Documentation**	**after 1 hour**	**within 1 hour**	**p-value[a]**
N		1888	981	
Triage within 15 min	non-missing	1670 (65.1%)	894 (34.9%)	0.027
	missing	218 (71.5%)	87 (28.5%)	
Timely exam. by physician	non-missing	1386 (63.8%)	786 (36.2%)	<0.001
	missing	502 (72.0%)	195 (28.0%)	
Blood culture prior to antibiotics	non-missing	1779 (64.9%)	961 (35.1%)	<0.001
	missing	109 (84.5%)	20 (15.5%)	
Fluids administered within 1 hour	non-missing	1277 (62.2%)	775 (37.8%)	<0.001
	missing	611 (74.8%)	206 (34.2%)	

[a] Pearson's chi square test

Regression analyses

We modeled the regression analyses of process and outcome measures using mixed effect models with hospital entered as a random effect. The process measurements and 30-day mortality were modeled using a logistic model (*melogit* in Stata). Length of stay was measured in days by subtracting admission date from discharge date. As this produced a left-skewed distribution, we chose to use a negative binomial regression model (*nbreg* in Stata).

In the adjusted models we first entered the following control variables: age (measured in years), organ failure, Charlson's comorbidity index, sex, and secular trends (entered as categorical variable denoting year of admission). We then tested alternative models by excluding control variables one at a time. At each step we obtained Akaike Information Criteria (AIC) and used this as a basis of comparison between the different models. For each adjusted analysis we chose the model that gave the lowest AIC. The estimates from the final adjusted models of process measurements are

presented in Table 2 below. The estimates from adjusted models of outcome variables 30-day mortality and length of stay are presented in Table 3 and Table 4, respectively.

Table 2 Adjusted regression models – process measures.

Predictors[a]	Outcome variables[bc]										
	Triage15m	Physic	CompVital1h	Blood1h	Lactate1h	SuppleDiag	BloodCult	Observ	Followup	Antibio1h	Fluid1h
After inspection	1.157	1.276**	1.673*	1.655	2.757***	1.151	1.331	2.204***	0.657	2.157***	1.152
Age	1.008**			1.006*	1.021***		0.993			0.994*	
Organ failure	1.542***	1.353**	0.999	1.263	1.981***	0.883	1.122	1.487***	1.426***	2.056***	1.771***
Admission year[d]											
2015			0.223**	0.444	0.913	0.937		0.518*	0.549*		
2016			0.348*	0.639	0.997	0.962		0.501**	0.686		
2017			0.302**	0.711	0.645*	2.187*		0.594**	1.328		
Sex (female=1, male=0)						0.757	0.77	0.849*	0.842*	0.787**	0.822
Charlson										0.968	
N	2724	2309	3047	3047	3047	3047	2697	3047	3047	2638	2179

[a] Constant term left out from table

[b] Numbers showing odds ratio. * p<0.05, ** p<0.01, *** p<0.001

[c] outcome variables

Triage15m	Triage within 15 minutes
Physic	Assessment by physician within time frame set in triage
CompVital1h	Complete vital measurements within 1 hour
Blood1h	Blood samples within 1 hour
Lactate1h	Lactate sampled within 1 hour
SuppleDiag	Adequate supplementary diagnostics
BloodCult	Blood culture sampled before antibiotic administration
Observ	Adequate observation in emergency department
Followup	Adequately discharged from emergency department
Antibio1h	Antibiotic administration within 1 hour
Fluid1h	Fluid administered within 1 hour

[d] reference year: 2018

Table 3 Adjusted model - 30-day mortality. N= 7360.

	Adjusted model 1			Adjusted model 2		
		95% Confidence interval			95% Confidence interval	
Predictors[a]	**Odds Ratio**	**Lower limit**	**Upper limit**	**Odds Ratio**	**Lower limit**	**Upper limit**
Inspection completed	0.81	0.69	0.95	1.25	0.86	1.80
Age in years	1.05	1.04	1.06	1.05	1.04	1.06
Organ failure	3.59	3.04	4.23	3.59	3.04	4.24
Charlson's comorbidity index	1.09	1.06	1.13	1.09	1.06	1.12
Sex (female=1, male=0)	0.93	0.79	1.09	0.92	0.79	1.08
Admission Year[b]						
2015				1.84	1.09	3.09
2016				1.29	0.80	2.06
2017				1.01	0.71	1.42

[a] Constant term left out from table

[b] Reference year: 2018

Table 4 Adjusted model - length of stay in days. N=7371.

	Adjusted model 1			Adjusted model 2		
		95% Confidence interval			95% Confidence interval	
Predictors[a]	**Incident rate ratio**	**Lower limit**	**Upper limit**	**Incident rate ratio**	**Lower limit**	**Upper limit**
Inspection completed	0.87	0.84	0.90	0.94	0.86	1.03
Age in years	1.00	1.00	1.00	1.00	1.00	1.00
Organ failure	1.33	1.29	1.38	1.33	1.28	1.38
Charlson's comorbidity index	1.05	1.04	1.05	1.04	1.04	1.05
Sex (female=1, male=0)	0.97	0.94	1.00	0.97	0.94	1.00
Admission Year[b]						
2015				1.20	0.76	1.88
2016				1.06	0.67	1.66
2017				1.06	0.67	1.68
2018				1.02	0.64	1.63

[a] Constant term left out from table

[b] Reference year: 2014

Inspection findings

Reported in the table below are the main findings from the inspections at the three hospitals, a description of key measures implemented by the hospitals after the inspections, and the percentages before and after the inspection of patients with sepsis who had antibiotic administration within one hour. Time to antibiotics was an important performance measurement included in the inspections' review of electronic health records (EHR). A previous study from this project lists all indicators that were included in the EHR review.[1]

The data for the main findings are based on the focus group interviews and the publicly available inspection reports.

The data on the percentages of patients with antibiotic administration within one hour were collected by the inspection teams. Patients presenting to the emergency department with an International Classification of Diseases, 10th Revision (ICD-10) diagnostic code classifying sepsis or infection were identified through the Norwegian Patient Registry. The EHR and included patients with clinically suspected infection and two systemic inflammatory response syndrome signs (not including high leukocyte count) were included.[2] Patients were sampled from four time periods specific to each hospital: two before the inspection and two after. Records from the two pre-inspection time periods were reviewed during the inspection, and records from the post-inspection periods were reviewed at 8 and 14 months after the inspection, using records from the most recent patients. For each time period, 33 patients were sampled, though the number of patients included in the analyses in some cases ended up being slightly smaller due to duplicate records.

References

1. Husabø G, Nilsen RM, Flaatten H, et al. Early diagnosis of sepsis in emergency departments, time to treatment, and association with mortality: An observational study. PLoS One 2020;15(1):e0227652 doi: 10.1371/journal.pone.0227652.
2. Dellinger RP, Levy MM, Rhodes A, et al. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock, 2012. Intensive Care Med 2013;39(2):165-228 doi: 10.1007/s00134-012-2769-8.

Supplementary table 1 Main findings from the inspections

Hospital	Population*	Main findings from the inspection	Follow-up by hospital	Percent of patients with antibiotic administration within one hour		
				Before insp.	After insp.	n
Hospital A	350 000	The inspection found that for a substantial proportion of patients, time from presentation to examination by physician and administration of antibiotics was delayed.	In response to the inspection, the hospital evaluated their procedures in inter-professional meetings and implemented changes in procedure and training initiatives.	22%	49%†	123
Hospital B	100 000	Some of the main findings from the inspection were delays in examination by physician and antibiotic administration. There were also inadequacies in documentation of responsibility and medical procedures. The emergency department in Hospital B had already started an improvement project for sepsis care prior to the inspection. The inspection nevertheless found deficiencies that the hospital had not been aware of.	The inspection led to a deepened commitment by the top-level management for the ongoing improvement project.	35%	59%†	122
Hospital C	50 000	The inspection found that for many patients, antibiotic treatment started too late. Furthermore, there were at times not enough available physicians to attend to patients in emergency department and not clear designation of responsibility for treatment between interns and resident physicians.	Following the inspection, the hospital started measuring indicators related to treatment in the emergency department, and clinicians and managers used these measurements for quality improvement purposes. In addition, there was a change in prehospital practice where more patients were administered antibiotics before being sent to the hospital.	18%	41%‡	77
Hospital D	300 000	The inspection found delays in antibiotic treatment and inadequate triage and observation of patients in emergency department.	After the inspection the hospital has implemented several initiatives, including training, revised procedures, and stand-up improvement board meetings.	15%	43%†	121
All hospitals§				25%	43%†	2869

* The hospitals are publicly owned and run institutions with responsibilities for specialized acute somatic care for all inhabitants in their local area. “Population” figures reported here are (rounded off and) based on information from the governments National plan for hospitals Meld. St. 11 (2015–2016).

† P-value < 0.01 (chi square test for difference between before and after inspection)

‡ P-value < 0.05 (chi square test for difference between before and after inspection)

§ All hospitals = all 24 hospitals included in the nation-wide inspection, including hospitals A - D.

Graphic design: Communication Division, UiB / Print: Skipnes Kommunikasjon AS

uib.no

ISBN: 9788230849613 (print)
9788230856512 (PDF)