PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Effects of external inspection on sepsis detection and treatment: a study protocol for a quasi-experimental study with a stepped wedge design
AUTHORS	Hovlid, Einar; Frich, Jan; Walshe, Kieran; Nilsen, Roy; Flaatten, Hans; Braut, Geir; Helgeland, Jon; Teig, Inger Lise; Harthug, Stig

VERSION 1 - REVIEW

REVIEWER	Renée Bouwman NIVEL, Netherlands Institute for Health Services Research
REVIEW RETURNED	08-Feb-2017

GENERAL COMMENTS	Very interesting and innovative research, very curious to know the results! I still have some concerns: Throughout the article, it is sometimes confusing which is done regularly by the Norwegian supervisor as a part of its work, or which is going to be carried out 'as a new activity' by the research team. Especially the text under the header 'interventions'. It is the intervention of your study, but normally, this is what the supervisor anyway. Am I right? I have the same problem with the process and outcome indicators you are going to collect: are'nt these normally also collected by the supervisor? Why not? Why does the supervisor not monitor these indicators to measure quality of care? In this light: it may help to give some more information about the other methods the supervisor uses to supervise quality of care. Occasionally, I am known with a stepped wedge design. However, it is still not totally clear to me why this design is working out well for this study, and I think other people would not understand it at all. At page 10, the sentence before reference no. 51: could you explain more in detail why this design is recommended for evaluating intervention effects? A very extensive list of process and outcome indicators is going to be collected. However, patient characteristics such as age, gender, educational level, etc. are not listed. These are all variables that could bias the relationship you study. For instance, higher educated people often have a better health status, people living in urban areas may have less health statuses, older people etc etc. Are you going to collect these data and are you going to correct for these characteristics? And how, in your statistical analysis? I have not enough expertise to assess how this must be analysed, and whether your analysis, the way it is described now, is adequate. I understand the relevance of the study, but in the introduction and discussion, you may want to have a closer look at what already has been done in other countries, concerning the effects of supervision

	and regulation on quality of care. In the Netherlands we are also attempting to measure this. Also see for instance the dissertation of Annemiek Huisman on supervision on suicides reports. This may also be interesting: Stoopendaal AMV, De Bree MA, Robben PBM. Reconceptualizing regulation: Formative evaluation of an experiment with system-based regulation in Dutch healthcare. Evaluation, okt. 2016. http://journals.sagepub.com/doi/abs/10.1177/1356389016667889 Abstract: I miss a reference to 'regulation' or regulatory regimes, to me that would be a word through which I recognize that this is my study field.

REVIEWER	Prof. Dr. Paul Robben Institute of Health Policy & Management, Erasmus University Rotterdam, the Netherlands
REVIEW RETURNED	13-Feb-2017

GENERAL COMMENTS	1. What is incomplete in the stepped wedge design? 2. Why not use the unexpected length of stay as a process indicator instead of length of stay. Reference: Reducing hospital length of stay by improving quality and safety of care? Thesis Radboud University Nijmegen, H, J. Borghans, 2012 3. "we do not intend to evaluate the effect of the different components of inspection; rather, we will analyze the effects of inspection as a whole". The qualitative data (focus groups) make it possible to give some insight in the effect of the different components of the inspection process as a complex intervention. 4. Will it be possible to generalize the outcome of this study (sepsis) to inspection in general? 5. the selection of hospitals needs more explanation. This doesn't seem to be a well controlled process. 6. there is not enough information about the inspection teams (seize, experience, variation in and between the teams). Is the report a consensus product of the inspection team? Data collection is done by the inspectors and not independently , this seems to be a major limitation of the study but is not mentioned as such.

REVIEWER	Anne-Marie Hill Curtin University Australia
REVIEW RETURNED	09-Mar-2017

GENERAL COMMENTS	Overall An important subject but a complex study protocol which needs to be more clearly explained to the reader. A major query that needs addressing is that this is not an observational trial, it has an intervention. The methods and procedures do not reflect this clearly and I wonder why it is registered as an observational trial. Authors should clarify their description at various points and also justify why they are talking observation but mention intervention. Authors themselves make the point that inspections can be considered interventions. Title – the manuscript needs to have the words study protocol in the trial

	Abstract Clarify the aim – there is no primary aim presented for the study Quality of care for who and where? But you have an Intervention clearly described in the procedure which you don't mention in the abstract? Your audit is an intervention and is what you are measuring – this is not an observational study. Why effect measures and not outcome measures? Change in process indicators is very vague to have in an abstract this needs to be clarified A key limitation of the study is the lack of randomisation and blinding this needs to be mentioned as it will clearly affect what sites to prior to and after such external audits. Introduction Page 4 line 58 – quality of care is vague and gives no background to the scope of the problem please explain what you mean using relevant studies. Line 23 page 5 should be "have reported.." Line 8 page 6 – again need to be specific about what you mean by process indicators at this point you should discuss the term briefly in the introduction so the reader knows what you are meaning there is no context for why this becomes an aim. Methods' Overall I think the methods section needs to be clarified and presented in a standardised research framework. This study uses an audit cycle and the measurements need to be presented more clearly around that. Page 6 line 39 – you need to supply a figure of your model you state here that the whole framework of the proposed study is based on it but a short text description doesn't demonstrate how this will be applied to measuring robust outcomes. Page 7 lines 10 to 26 – I don't understand why this is here – you seem to be discussing your outcome measures and concurrently mentioning in part a description how your statistical analysis will work. I think you need to re-sort the methods section including this section Your procedure should have a table that summarises the whole process as the text is very wordy. This intervention is not a single entity. Therefore it needs to be made much clearer how the outcomes will be robustly measured in relation to when the reports and feedback are provided during the study period. You state that (page 16 lines 44 on) that you measure after the inspection (the intervention) but you do this by conducting another audit – so surely this is not just measurement this is further intervention? Data analysis I think this needs to be much more clearly presented and in much more depth – the aims don't seem to be matched by the analysis and which outcomes exactly are going to be analysed using which methods – your description is so broad it's not possible to know what you are evaluating and how. The boxed texts contain multiple outcomes. Importantly you don't clearly state what you are measuring in terms of the audit – is it baseline with the final data collected? Is it each step in the wedge, you need to present this in a coherent manner for the reader Line 32 page 13 where did you get the ICC of 0.05 from? Figure is difficult to grasp- the inspections don't seem to match the observations (how many inspections for each hospital) and then hard to work out the timeline the reader should not have to add squares manually to work out how the stepped wedge is working

	I think the figure needs to be altered to show the inspection (it is not an intervention) and give the totals for the reader in an overall manner showing how the procedure works so a highlighted example in there of one hospital with more text Clarify intervention vs inspection clearly in the figure It's also not clear why the baseline is so far before the beginning of the stepped wedge and that it occurs retrospectively and this does not match your description of how it is collected in text? Mention in figure what is occurring at that point? IF baseline done and no blinding are sites altering procedures before you start the inspections? Additionally its not clear where the report and feedback occurs and hospital action occurs in the figure Minor – would be better practice to describe "patients with a diagnosis of sepsis" or patients with sepsis not "sepsis patients) The SPIRIT checklist demonstrates that the study does not fit the criteria for an observational trial – I think authors need to clarify this - perhaps register it as a Quality improvement project

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Renée Bouwman
Institution and Country: NIVEL, Netherlands Institute for Health Services Research
Competing Interests: None declared

Throughout the article, it is sometimes confusing which is done regularly by the Norwegian supervisor as a part of its work, or which is going to be carried out 'as a new activity' by the research team. Especially the text under the header 'interventions'. It is the intervention of your study, but normally, this is what the supervisor anyway. Am I right?
You are right. We have re-written the whole method section and tried to explain better what the inspection teams do. We have also included a new table with the 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 is 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 the site visit. The inspection team reviews records of patients with sepsis and collect the same data as they did prior to the site visit.
Report on findings from audit
17 Follow up audit 14 months after the site visit. The inspection team reviews records of patients with sepsis and collect the same data as they did prior to the site visit.
Report on findings from audit

Table 2. Key elements of the intervention

I have the same problem with the process and outcome indicators you are going to collect: are'nt these normally also collected by the supervisor? Why not? Why does the supervisor not monitor these indicators to measure quality of care? In this light: it may help to give some more information about the other methods the supervisor uses to supervise quality of care.

We have made changes throughout the method section to make this more understandable. The process and outcome measures are not available as routine data and are according to Norwegian legislation collected and used as part of the inspection. The same data are then used for research purposes, and the Regional Ethics Committee of Norway Nord (REC) along with the Norwegian Data Protection Authority have ruled that we can use these data collected during the inspection for research purposes.

We have also added a new paragraph with information about other methods used to supervise the quality of care.
The County Governors are responsible for supervising the hospitals in their region. According to Norwegian legislation, hospitals are required to inform the County Governor about serious adverse patient events, and the County Governor investigates such patient events to decide whether the hospital has delivered inappropriate care. Furthermore, the County Governor handle general patient complaints and carry out inspections in different areas on a regular basis. Based on these supervisory activities the County Governors possesses knowledge about risk and vulnerability at the hospitals in their counties, e.g. high turnover of personnel, lack of key competence, or financial constraints.

Occasionally, I am known with a stepped wedge design. However, it is still not totally clear to me why this design is working out well for this study, and I think other people would not understand it at all. At page 10, the sentence before reference no. 51: could you explain more in detail why this design is recommended for evaluating intervention effects?
We have changed the whole methods section and provided more information:

Randomized controlled trials (RCTs) are considered the gold standard for assessing the effects of an intervention.44 However, in the present project, an RCT will not be feasible as it is impossible to establish an appropriate control group. Data regarding detection and treatment of sepsis are not available as routine data in Norway. Such data can only be collected by reviewing individual patient records. According to Norwegian legislation, the inspection teams have access to patient records and can collect relevant data as part of the inspection. If we were to conduct an RCT, the inspection teams would have to collect data from hospitals that were not inspected. Collecting such data is a key ingredient of an inspection and would itself be an intervention. Furthermore, if the data collected from hospitals in the control group were indicative of non-compliant behavior, the inspection teams would have to follow up their findings with those hospitals; thus, it would no longer be a control group. A stepped wedge design has been recommended for evaluating intervention effects when it is not feasible to establish a control group.45 Furthermore, this type of design is recommended for evaluating the effect of service delivery type interventions where it is not possible to expose the whole study population for the intervention simultaneously and where implementation takes time.46 In our case the intervention is aimed at changing service delivery for patient with sepsis, it is not possible to conduct all inspection simultaneously, and implementation of change following the intervention takes time.

A very extensive list of process and outcome indicators is going to be collected. However, patient characteristics such as age, gender, educational level, etc. are not listed. These are all variables that could bias the relationship you study. For instance, higher educated people often have a better health status, people living in urban areas may have less health statuses, older people etc etc. Are you going to collect these data and are you going to correct for these characteristics? And how, in your statistical analysis? I have not enough expertise to assess how this must be analysed, and whether your analysis, the way it is described now, is adequate.

We have re-written the whole section about statistical analysis. In addition we have provided a new paragraph in the discussion section:
Statistical analysis
Descriptive statistics will be used to quantify sample characteristics. All analysis will use patient-level data, collected at four periods for different patients for 24 hospitals – two collections during the control period and two collections during the intervention period for each hospital. In order to compare the various process and outcome measurements (dependent variables) between the intervention and control periods (independent variable), we will use logistic regression models for binary measurements and linear regression models for continuous measurements. The choices of regression methods for the various measurements are outlined in table 3. As recommended in literature, 59 60 all models will include time as a covariate to adjust for potential secular changes in the process and outcome measurements during the study period. The underlying form of time will be included in the models as a linear term, polynomial term, or cubic spline term, as appropriate. As patients are sampled from different hospitals, a between-hospital variation in measurements is likely, introducing correlated data within the hospitals. To account for this intra-cluster correlation, we will use generalized estimating equations methodology,61 specifying an exchangeable working correlation structure, i.e., any two patients are equally correlated within hospitals regardless of time and intervention and control periods. However, as this assumption might not hold for all hospitals, a method for obtaining cluster-robust standard errors of model parameters will be applied.62 Finally, as our repeated sampling of patients with sepsis may not be entirely representative of the total population, difference in certain patient characteristics, including age and sex, between comparison periods might arise. In that case, the abovementioned models will also include such covariates for obtaining correct model means.

Factors like socioeconomic status and co-morbidity can affect the outcomes. We do not have access to such data and can therefore not adjust for these factors. In a stepped wedge design, all sites are exposed for the intervention, and we compare change in the effect measures before and after the intervention. We collect the data in a standardized way, and the intervention itself should not affect which patients that are admitted to the hospitals. Consequently, we have no reason to believe that confounding factors should be unevenly distributed in the study population before and after the intervention.

I understand the relevance of the study, but in the introduction and discussion, you may want to have a closer look at what already has been done in other countries, concerning the effects of supervision and regulation on quality of care. In the Netherlands we are also attempting to measure this. Also see for instance the dissertation of Annemiek Huisman on supervision on suicides reports.
This may also be interesting:
Stoopendaal AMV, De Bree MA, Robben PBM. Reconceptualizing regulation: Formative evaluation of an experiment with system-based regulation in Dutch healthcare. Evaluation, okt. 2016.

http://journals.sagepub.com/doi/abs/10.1177/1356389016667889

We have made changes to the first paragraphs in the introduction and discussion section and included the two references that you suggested. We also refer to other works from the Netherlands:

van Dishoeck AM, Oude Wesselink SF, Lingsma HF, et al. [Transparency: can the effect of governmental surveillance be quantified?]. Ned Tijdschr Geneeskd 2013;157(16):A1676. [published Online First: 2013/04/19]
Oude Wesselink SF, Lingsma HF, Reulings PGJ, et al. Does Government Supervision Improve Stop-Smoking Counseling in Midwifery Practices? Nicotine & Tobacco Research 2014 doi: 10.1093/ntr/ntu190
Oude Wesselink SF, Lingsma HF, Ketelaars CA, et al. Effects of government supervision on quality of integrated diabetes care: a cluster randomized controlled trial. Med Care 2015;53(9):784-91. doi: 10.1097/mlr.0000000000000399 [published Online First: 2015/08/01
Ngo D, Breejen Ed, Putters K, et al. Supervising the Quality of Care in Changing Healthcare Systems - An International Comparison. Rotterdam: Dept. of Healthcare Governance Institute of Health Policy and Management, Erasmus University Medical Center, 2008.
Sparreboom WF. How Effective Are You? A Research on How Health Care Regulators Across Europe Study the Effectiveness of Regulation. [Master]. Amsterdam: VU University, 2009.

Abstract: I miss a reference to 'regulation' or regulatory regimes, to me that would be a word through which I recognize that this is my study field.
We have included regulation as a key word

Reviewer: 2
Reviewer Name: prof. dr. Paul Robben
Institution and Country: Institute of Health Policy & Management, Erasmus University Rotterdam, the Netherlands
Competing Interests: none declared

1. What is incomplete in the stepped wedge design?
We have added a new sentence:
The design is incomplete in that we do not continuously collect data from all the included sites, rather we collect data at four different time points.

2. Why not use the unexpected length of stay as a process indicator instead of length of stay.
Reference: Reducing hospital length of stay by improving quality and safety of care? Thesis Radboud University Nijmegen, H, J. Borghans, 2012

We agree that this is a good idea, and probably a better indicator. However, unexpected length of stay is not available as a routine indicator in Norway. We would have to review all the patient records manually and make a judgement on unexpected length of stay for each patient. That would be too resource demanding.

3. "we do not intend to evaluate the effect of the different components of inspection; rather, we will analyze the effects of inspection as a whole". The qualitative data (focus groups) make it possible to give some insight in the effect of the different components of the inspection process as a complex intervention.

That is a good point. We have added two paragraphs in the discussion section:
The stepped wedge design enables us to track changes in outcome measures for sepsis detection and treatment over time. The changes that we might observe in the outcome measures are not necessarily attributable to the inspections alone. There can be other factors beside the inspections that can affect sepsis detection and treatment during the study period. Our qualitative data can help identifying such factors and provide insight into how they might interact with the inspections. By combining findings from the qualitative and quantitative data, we assert that we can assess how sepsis detection and treatment develops over time and substantiate how inspections along with other factors can affect the development.

Our study is based on an overall framework suggesting that the inspection need to affect organizational ideas and activities to facilitate change in organizational behavior and thereby improve quality of care for patients with sepsis (figure 1). Moreover, we have developed a more detailed theory of change for the inspections, suggesting how they can contribute to affect organizational change. We collect quantitative data that indicate whether organizational behavior and the quality of care improve after the inspections. By combining these findings with our qualitative data that provide insight into how the inspections affect organizational change, we can test our theories about how inspections can contribute to improve the quality of care. We suggest that our study can contribute to build theory about how inspections can improve the quality of care, and thus have relevance for inspections covering other topics.

4. Will it be possible to generalize the outcome of this study (sepsis) to inspection in general?
We have added this paragraph in the discussion section:
Our study is based on an overall framework suggesting that the inspection need to affect organizational ideas and activities to facilitate change in organizational behavior and thereby improve quality of care for patients with sepsis (figure 1). Moreover, we have developed a more detailed theory of change for the inspections, suggesting how they can contribute to affect organizational change. We collect quantitative data that indicate whether organizational behavior and the quality of care improve after the inspections. By combining these findings with our qualitative data that provide insight into how the inspections affect organizational change, we can test our theories about how inspections can contribute to improve the quality of care. We suggest that our study can contribute to build theory about how inspections can improve the quality of care, and thus have relevance for inspections covering other topics.

5. the selection of hospitals needs more explanation. This doesn't seem to be a well controlled process.
We have added more information on how the hospitals are selected:

The County Governors are responsible for supervising the hospitals in their region. According to Norwegian legislation, hospitals are required to inform the County Governor about serious adverse patient events, and the County Governor investigates such patient events to decide whether the hospital has delivered inappropriate care. Furthermore, the County Governor handles general patient complaints and carries out inspections in different areas on a regular basis. Based on these supervisory activities the County Governors possesses knowledge about risk and vulnerability at the hospitals in their counties, e.g. high turnover of personnel, lack of key competence, or financial constraints.

About 40 acute care hospitals in Norway treat patients with sepsis. There is large variation in the size of these hospitals and the number of patients treated. All 40 hospitals are eligible for inspection. The

standard procedure used by the National Board of Health Supervision for conducting nation-wide inspections is followed. This procedure implies that the regional teams decide which hospitals to inspect in their region. The main criterion for selecting which hospitals to inspect is hospital size. The large hospitals treat more patients, and consequently sub-standard care will affect many patients. Moreover, the inspection teams also use their local knowledge about specific risks and vulnerability when selecting hospitals for inspection.

6. there is not enough information about the inspection teams (seize, experience, variation in and between the teams). Is the report a consensus product of the inspection team?
We have provided more information about the teams:
The inspections are conducted by six regional teams from the County Governors in Norway. The teams consist of a minimum of four inspectors. The leader of the team has long experience and particular training in doing inspections. The team has medical and legal expertise. One of the team members is an external medical expert who has special expertise on sepsis. The expert works on a daily basis in a hospital, but has been hired part time by the Norwegian Board of Health Supervision to assist the inspection teams. The clinical experts do not participate in inspections of hospitals where they have their regular work.

Data collection is done by the inspectors and not independently , this seems to be a major limitation of the study but is not mentioned as such.
We have added more information on how the data is collected:
Revised paragraph in method section along with a new figure:
Figure 2 outlines the clusters and the data collection. For each inspection, we collect data at four different time points, referred to as P0, P1, P2, and P3. P0 is the baseline measurement for all hospitals before the Norwegian Board of Health Supervision announced the inspection campaign. The campaign is part of the regular planned inspection activities. These activities are transparent for the hospitals and are announced in advance, in this particular case six months before the first inspection. The baseline measurement is done right before the inspection campaign was announced. All hospitals know that they can be inspected. By collecting data before the campaign was announced we can track changes throughout the inspection cycle and assess to what extent changes are implemented before the inspections are undertaken. The data for P0 will be collected retrospectively at the same time as that for P1. P1 is the pre-inspection measurement, and P2 and P3 are post-inspection measurements. The regional inspection teams collect data during the inspection and audits conducted 8 (P2) and 14 months (P3) following the initial inspection. These data serve two purposes. They are used to guide the judgments on whether the inspected hospitals comply with the requirements, and to evaluate how inspections affect the clinical processes involved in diagnosing and treating sepsis.

Figure 2. Illustration of clusters and data collection

We have also added a new paragraph in the methods section about data monitoring:

Data monitoring
The external medical expert, together with the leader of the inspection team, oversees the data collection process. Inspectors with medical expertise collect data by reviewing electronic patient records. To increase inter rater reliability, the inspectors work in pairs, and they all sit together in the same room and can ask for supervision from the external medical expert when needed. To reduce inter rater bias between the inspection teams, the Norwegian Board of Health Supervision has developed a framework describing in detail the data that should be collected from the patient records and the criteria for judgement. All inspection teams have received special training and participated in

meetings where the audit criteria have been discussed to promote a common understanding. Once a team is assigned to one hospital, they collect data at all four time points. To promote validity and reliability of the collected data the involved hospitals can oversee how data is collected. The entire data collection process is transparent for the inspected hospitals, and the hospitals can verify all the collected data if they wish. The complete data file is checked manually before analysis, and we also apply various procedures of electronic field checks to secure data quality.

New paragraph in discussion section:
The inspections in our study are contemporary and transparent events, thus it is not possible to mask who is exposed to the intervention. Nor is it possible to mask the data collection. Because there is no available routine date about sepsis detection and treatment, such data is collected as part of the intervention in our study. The inspections teams access the patient records manually, and they will therefore know which hospital the patients belong to and whether the patient was admitted before or after the inspection. The fact that data is collected as part of the intervention can be viewed as a limitation. Doing data collection during the inspection is however standard procedure, and thus not atypical for the inspections in our study.48 Given the nature of the intervention – that is collecting data, reporting and giving recommendations – what we are in fact measuring is, at least in part, the effect of data collection.

Reviewer: 3
Reviewer Name: Anne-Marie Hill
Institution and Country: Curtin University Australia
Competing Interests: None declared

Overall
An important subject but a complex study protocol which needs to be more clearly explained to the reader. A major query that needs addressing is that this is not an observational trial, it has an intervention. The methods and procedures do not reflect this clearly and I wonder why it is registered as an observational trial. Authors should clarify their description at various points and also justify why they are talking observation but mention intervention. Authors themselves make the point that inspections can be considered interventions.

General comment
We have rewritten and restructured the entire methods section in line with your recommendations. We agree that this is not an observational trial, because it has an intervention. We wanted to be transparent in what we do and therefore we registered our study in clinicaltrials.gov. The register has two boxes for study types:
1. Clinical trial, in which patients are assigned to an intervention
2. Observational study.
The dilemma is that our study does not fit either of these two categories entirely. On the one hand, it does not fit the definition of a clinical trial where patients are assigned to the intervention. In our study the intervention is on the organizational level, and patients are not assigned to an intervention. On the other hand, it is not an observational trial because it has an intervention. Initially, we registered the study as an observational trial and described that it had an intervention. We have now changed the registration in clinicaltrials.gov to study type clinical trial, and explained that patients are not assigned to an intervention. We have changed the manuscript accordingly.

Title – the manuscript needs to have the words study protocol in the trial
New title:
Effects of external inspection on sepsis detection and treatment: a study protocol for a quasi-

experimental study with a stepped wedge design

Abstract
Clarify the aim – there is no primary aim presented for the study
Quality of care for who and where?
But you have an Intervention clearly described in the procedure which you don't mention in the abstract? Your audit is an intervention and is what you are measuring – this is not an observational study.
Why effect measures and not outcome measures?
Change in process indicators is very vague to have in an abstract this needs to be clarified
A key limitation of the study is the lack of randomisation and blinding this needs to be mentioned as it will clearly affect what sites to prior to and after such external audits.

New abstract:
Introduction
Inspections are widely used in health care as a means to improve the health services delivered to patients. Despite their widespread use, there is little evidence of their effect. The mechanisms for how inspections can promote change are poorly understood. In this study, we use a national inspection campaign of sepsis detection and initial treatment in hospitals as case to:
1. Explore how inspections affect the involved organizations.
2. Evaluate what effect external inspections have on the process of delivering care to patients, measured by change in indicators reflecting how sepsis detection and treatment is carried out.
3. Evaluate whether external inspections affect patient outcomes, measured as change in the 30-day mortality rate and length of hospital stay.

Methods and analysis
The intervention that we study is inspections of sepsis detection and treatment in hospitals. The intervention will be rolled out sequentially during 12 months to 24 hospitals. Our effect measures are change in indicators related to the detection and treatment of sepsis, the 30-day mortality rate, and length of hospital stay. We collect data from patient records at baseline, before the inspections, and at 8 and 14 months after the inspections. We use logistic regression models and linear regression models to compare the various effect measurements between the intervention and control periods. All the models will include time as a covariate to adjust for potential secular changes in the effect measurements during the study period. We collect qualitative data before and after the inspections, and we will conduct a thematic content analysis to explore how inspections affect the involved organizations.

Ethics and dissemination
The study has obtained ethical approval by the Regional Ethics Committee of Norway Nord and the Norwegian Data Protection Authority. It is registered at www.clinicaltrials.gov (Identifier: NCT02747121). Results will be reported in international peer-reviewed journals.

Introduction
Page 4 line 58 – quality of care is vague and gives no background to the scope of the problem please explain what you mean using relevant studies.
We have revised the first paragraph and added a new paragraph in the introduction:
External inspections constitute a core component of regulatory regimes and certification and accreditation processes.1 2 Different terms such as external review, supervision, and audit have been used to describe this activity.3 4 There are differences between these approaches, but they have in common that a health care organization's performance is assessed according to an externally defined

standard. We use the term “external inspection”, which implies that the inspection is initiated and controlled by an organization external to the one being inspected.5 We define external inspection as: a system, process or arrangement in which some dimensions or characteristics of a healthcare provider organization and its activities are assessed or analyzed against a framework of ideas, knowledge, or measures derived or developed outside that organization.6

Inspections are widely used in health care as a means to improve the quality of care delivered to patients.1 7 Quality of care is a complex concept that can be understood in different ways.8 We understand quality of care as: the degree to which health services for individuals and populations increase the likelihood of desired health outcomes, and are consistent with current professional knowledge.9 We found this definition expedient because it highlights that the quality of care encompasses outcomes for patients and populations, and that the outcomes are dependent on the delivery of health services consistent with current professional knowledge. External inspections can be used with the intention to secure that delivery of health services are consistent with current professional knowledge.

Line 23 page 5 should be “have reported..”
Corrected

Line 8 page 6 – again need to be specific about what you mean by process indicators at this point you should discuss the term briefly in the introduction so the reader knows what you are meaning there is no context for why this becomes an aim.
We have rewritten and restructured the whole methods section. We start out by describing our conceptual framework. We have also made a new figure that illustrates the relationship between the framework, the study aims, data and effect measures (figure 1). We have defined what we mean by process indicators and tried to provide the context by explaining our study framework more thoroughly.

Conceptual framework
We take the perspective that quality of care can be considered a system property that is dependent on how the organization providing care performs as a whole.34 Improving the quality of care is thus dependent on changing organizational behavior, which implies changing the way clinicians interact and perform their clinical processes.35 36 Change in organizational behavior is a complex social process that involves a range of different organizational activities.37 If external inspection is to contribute to improvement in the quality of care, it should have an impact on those activities involved in organizational change, here defined as any modification in organizational composition, structure, or behavior.38

We have previously conducted a systematic review of published research to identify the mechanisms of how external inspections can contribute to improving quality of care in health organizations.39 By combining empirical evidence and theoretical contributions, we found evidence to support that external inspections need to affect both organizational ideas and organizational change activities to improve the quality of care. Organizational ideas encompass theoretical constructs like organizational readiness for change, awareness of current practice and performance gaps, and organizational acceptance that change is necessary.40 41 Organizational change activities refer to key activities involved in quality improvement like setting goals, planning and implementing improvement measures, and evaluating effect of such measures.42 43

Figure 1. Conceptual study framework

Figure 1 depicts our overall conceptual framework, how the elements of the framework relate to the different study aims, and the corresponding data and effect measures. We suggest that inspections can affect organizational ideas and initiate change activities, which in turn can lead to organizational change. We collect qualitative data to explore how the inspections affect the involved organizations. Moreover, we suggest that organizational change and change in the process of detecting and treating patients with sepsis can contribute to improve the quality of care. To measure change in the process of detecting and treating sepsis we collect data that reflect this process, e.g. time to triage, time to initial assessment by physician, and time to treatment with antibiotics. We refer to these data as process indicators, because they reflect how the process of detecting and treating sepsis is carried out. To measure change in the quality of care we use two outcome measures, length of hospital stay and 30-day mortality rate.

Methods'
Overall I think the methods section needs to be clarified and presented in a standardised research framework.
This study uses an audit cycle and the measurements need to be presented more clearly around that.

We have rewritten and restructured the whole methods section in line with your suggestions. First, we present our conceptual framework and a figure illustrating the framework, the study aims and the effect measures.
We have tried to present the methods in a more standardized framework (PICO-model). We present the study population, the intervention, the comparison and the outcomes.

Page 6 line 39 – you need to supply a figure of your model you state here that the whole framework of the proposed study is based on it but a short text description doesn't demonstrate how this will be applied to measuring robust outcomes.
We have made a new figure, figure 1. See also comments above.

Page 7 lines 10 to 26 – I don't understand why this is here – you seem to be discussing your outcome measures and concurrently mentioning in part a description how your statistical analysis will work. I think you need to re-sort the methods section including this section

We have taken this part out and moved it to the section with qualitative analysis.

Your procedure should have a table that summarises the whole process as the text is very wordy. This intervention is not a single entity. Therefore it needs to be made much clearer how the outcomes will be robustly measured in relation to when the reports and feedback are provided during the study period. You state that (page 16 lines 44 on) that you measure after the inspection (the intervention) but you do this by conducting another audit – so surely this is not just measurement this is further intervention?

We have made new table summarizing the intervention. We have also tried to clarify that the audits at 8 and 14 months after the inspection is also part of the intervention.
Table 2. 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 is 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 the site visit. The inspection team reviews records of patients with sepsis and collect the same data as they did prior to the site visit.
Report on findings from audit
17 Follow-up audit 14 months after the site visit. The inspection team reviews records of patients with sepsis and collect the same data as they did prior to the site visit.
Report on findings from audit

Data analysis
I think this needs to be much more clearly presented and in much more depth – the aims don't seem to be matched by the analysis and which outcomes exactly are going to be analysed using which methods – your description is so broad it's not possible to know what you are evaluating and how. The boxed texts contain multiple outcomes. Importantly you don't clearly state what you are measuring in terms of the audit – is it baseline with the final data collected? Is it each step in the wedge, you need to present this in a coherent manner for the reader
We have re-written the whole section on data analysis. We have made a new figure, figure 2, that outlines the clusters and the data collection. Furthermore, we have also revised table 1, which displays the overall study design. We have also include a new table 3 that outlines the analysis.

Statistical analysis
Descriptive statistics will be used to quantify sample characteristics. All analysis will use patient-level data, collected at four periods for different patients for 24 hospitals – two collections during the control period and two collections during the intervention period for each hospital. In order to compare the various process and outcome measurements (dependent variables) between the intervention and control periods (independent variable), we will use logistic regression models for binary measurements and linear regression models for continuous measurements. The choices of regression methods for the various measurements are outlined in table 3. As recommended in literature, 59 60 all models will include time as a covariate to adjust for potential secular changes in the process and outcome measurements during the study period. The underlying form of time will be included in the models as a linear term, polynomial term, or cubic spline term, as appropriate. As patients are sampled from different hospitals, a between-hospital variation in measurements is likely, introducing correlated data within the hospitals. To account for this intra-cluster correlation, we will use generalized estimating equations methodology,61 specifying an exchangeable working correlation structure, i.e., any two patients are equally correlated within hospitals regardless of time and intervention and control periods. However, as this assumption might not hold for all hospitals, a method for obtaining cluster-robust standard errors of model parameters will be applied.62 Finally, as our repeated sampling of patients with sepsis may not be entirely representative of the total population, difference in certain patient characteristics, including age and sex, between comparison periods might arise. In that case, the abovementioned models will also include such covariates for obtaining correct model means.
Table 3. Outline of regression models for the various process and outcome measurements

Indicator Dependent variable Type GEE model1
Process Triage within 15 minutes Binary Logistic regression
Process Timely assessment by physician Binary Logistic regression
Process Vital signs evaluated within 30 minutes Binary Logistic regression
Process Blood lactate measured within 30 minutes Binary Logistic regression
Process Supplementing blood samples within 30 minutes Binary Logistic regression
Process Blood culture taken before antibiotics Binary Logistic regression
Process Adequate supplementing investigation within 24 hours Binary Logistic regression
Process Antibiotic treatment within 1 hour Binary Logistic regression
Process Intravenous fluid within 30 minutes Binary Logistic regression
Process Oxygen therapy within 30 minutes Binary Logistic regression
Process Adequate surveillance regime established Binary Logistic regression
Process Adequate discharge from emergency room Binary Logistic regression
Outcome 30-day mortality Binary Logistic regression
Outcome Length of stay Continuous2 Linear regression
1 Regression models with generalized estimating methodology (GEE).
2 Transformed if skewed distribution.

Line 32 page 13 where did you get the ICC of 0.05 from?
Revised paragraph:
The power calculations were performed using the steppedwedge function in Stata/IC version 14.0 (StataCorp LLC, College Station, TX, USA) software for Windows, developed by Hemming and Girling.54 The statistical power in a stepped wedge design depends on the total number of intervention sites, the total number of data collection points for each intervention site, the number of patient records included at each data collection point, the correlation between clustered observations on the same hospital (intra-cluster correlation), and the implementation period.45 We based our calculations on 24 intervention sites, 4 data collection points, and 33 patient records per collection point at each intervention site. As the intra-cluster correlation may vary between samples and between process and outcome measurements, it is not straightforward to specify an intra-cluster correlation in advanced. In addition, we could not find any estimated intra-cluster correlation in previous trials of patients with sepsis. Consequently, we chose an intra-cluster correlation of 0.05, which is in line with that estimated for several patient outcomes in a cluster randomized trials of heart failure patients.55 Type I and II errors were assumed to be 0.05 and 0.20, respectively.

Figure is difficult to grasp- the inspections don’t seem to match the observations (how many inspections for each hospital) and then hard to work out the timeline the reader should not have to add squares manually to work out how the stepped wedge is working
I think the figure needs to be altered to show the inspection (it is not an intervention) and give the totals for the reader in an overall manner showing how the procedure works so a highlighted example in there of one hospital with more text Clarify intervention vs inspection clearly in the figure

See comment above.

It’s also not clear why the baseline is so far before the beginning of the stepped wedge and that it occurs retrospectively and this does not match your description of how it is collected in text? Mention in figure what is occurring at that point? IF baseline done and no blinding are sites altering procedures before you start the inspections?

We have provided more information:

Quantitative data
Figure 2 outlines the clusters and the data collection. For each inspection, we collect data at four different time points, referred to as P0, P1, P2, and P3. P0 is the baseline measurement for all hospitals before the Norwegian Board of Health Supervision announced the inspection campaign. The campaign is part of the regular planned inspection activities. These activities are transparent for the hospitals and are announced in advance, in this particular case six months before the first inspection. The baseline measurement is done right before the inspection campaign was announced. All hospitals know that they can be inspected. By collecting data before the campaign was announced we can track changes throughout the inspection cycle and assess to what extent changes are implemented before the inspections are undertaken. The data for P0 will be collected retrospectively at the same time as that for P1. P1 is the pre-inspection measurement, and P2 and P3 are post-inspection measurements. The regional inspection teams collect data during the inspection and audits conducted 8 (P2) and 14 months (P3) following the initial inspection. These data serve two purposes. They are used to guide the judgments on whether the inspected hospitals comply with the requirements, and to evaluate how inspections affect the clinical processes involved in diagnosing and treating sepsis.

Figure 2. Illustration of clusters and data collection

New paragraphs in discussion section:
The stepped wedge design enables us to track changes in outcome measures for sepsis detection and treatment over time. The changes that we might observe in the outcome measures are not necessarily attributable to the inspections alone. There can be other factors beside the inspections that can affect sepsis detection and treatment during the study period. Our qualitative data can help identifying such factors and provide insight into how they might interact with the inspections. By combining findings from the qualitative and quantitative data, we assert that we can assess how sepsis detection and treatment develops over time and substantiate how inspections along with other factors can affect the development.

The inspections in our study are contemporary and transparent events, thus it is not possible to mask who is exposed to the intervention. Nor is it possible to mask the data collection. Because there is no available routine date about sepsis detection and treatment, such data is collected as part of the intervention in our study. The inspections teams access the patient records manually, and they will therefore know which hospital the patients belong to and whether the patient was admitted before or after the inspection. The fact that data is collected as part of the intervention can be viewed as a limitation. Doing data collection during the inspection is however standard procedure, and thus not atypical for the inspections in our study.48 Given the nature of the intervention – that is collecting data, reporting and giving recommendations – what we are in fact measuring is, at least in part, the effect of data collection.

Additionally its not clear where the report and feedback occurs and hospital action occurs in the figure
We provide this information in the new table 2.

Minor – would be better practice to describe “patients with a diagnosis of sepsis” or patients with sepsis not “sepsis patients)
We have changed it to patients with sepsis

The SPIRIT checklist demonstrates that the study does not fit the criteria for an observational trial – I think authors need to clarify this - perhaps register it as a Quality improvement project

See first comment. We have change our registration to clinical trial and explained that the intervention is on the organizational level.

VERSION 2 – REVIEW

REVIEWER	Anne-Marie Hill Curtin University, Perth, Australia
REVIEW RETURNED	18-Apr-2017

GENERAL COMMENTS	Overall- this revised version of the study protocol has clarified the procedure and presents more depth to the context of the study. It allows the reader to understand the methodology and in particular the statistical analysis. I wish the authors well with the study and have some minor comments only. Abstract: The three aims as presented are a clear means of introducing the research. Methods: The study design is clear now and the authors clearly explain the time covariate of time in the statistical analysis, but in the discussion it should be noted that the PO measurement appears to be retrospective? in that the prospective data collection starts at P1 – can authors clarify? For the outcome measure of 30 day mortality rate it would be helpful for the reader to define this and describe why this is significant, rather than the reader have to go to the references to know why this is chosen and in particular whether it it's a robust timepoint to measure hospital mortality. Discussion: It would be helpful to discuss how different the hospitals are – eg size of ICU, type of patients admitted, health care area –as it would then be clear in explaining how the stepped wedge can cater for these differences between the hospitals baseline characteristics to be understood within the statistical analysis. Minor corrections: Page 4 line 8; Page 16 – line 3; Page 17 - line 15 and line 32 -again suggest throughout don't refer to sepsis patients but patients diagnosed with sepsis or similar phrase Page 18 – line 49 – should read "inspections need" as being plural.

VERSION 2 – AUTHOR RESPONSE

In the following, we will describe our changes to the manuscript.

Comment 1:
Methods:
The study design is clear now and the authors clearly explain the time covariate of time in the statistical analysis, but in the discussion it should be noted that the PO measurement appears to be retrospective? in that the prospective data collection starts at P1 – can authors clarify?
We have added a few sentences to the paragraph to clarify how data collection is carried out.

Figure 2 outlines the clusters and the data collection. For each inspection, we collect data at four

different time points, referred to as P0, P1, P2, and P3. P0 is the baseline measurement for all hospitals before the Norwegian Board of Health Supervision announced the inspection campaign. The campaign is part of the regular planned inspection activities. These activities are transparent for the hospitals and are announced in advance, in this particular case six months before the first inspection. Due to practical reasons, the inspection teams collect data for P0 and P1 at the same time right before the inspection. Data for P0 are thus collected retrospectively, but it is pre-defined that the patients that will be included are the last patients with suspected sepsis that were admitted prior to October 1st 2015.

Comment 2
For the outcome measure of 30 day mortality rate it would be helpful for the reader to define this and describe why this is significant, rather than the reader have to go to the references to know why this is chosen and in particular whether it it's a robust timepoint to measure hospital mortality.

We have added a new paragraph:

We use 30-day mortality rate as our key outcome measure, defined as the ratio of patients with sepsis who are dead within 30 days of hospital admittance. Mortality measures based on in-hospital deaths alone, can be misleading as indicators of hospital performance.55 Our measure also includes out of-hospital deaths. Using the unique personal identification number provided to all citizens of Norway, we are able to link the patient record data with data from the National Registry to calculate the 30-day mortality rate. 30-day mortality rate is an established, national quality indicator for Norwegian hospitals,56 and this indicator has been shown to have better validity as a hospital performance measure than in-hospital mortality for selected medical conditions.57 30-day mortality rate has also previously been used to assess effects of measures to improve care for patients with sepsis.58
Comment 3
Discussion:
It would be helpful to discuss how different the hospitals are – eg size of ICU, type of patients admitted, health care area –as it would then be clear in explaining how the stepped wedge can cater for these differences between the hospitals baseline characteristics to be understood within the statistical analysis.

We have revised the paragraph on statistical analysis:

Norway is divided into 18 counties in addition to its capital city, and there are acute hospitals that treat patients with sepsis in all counties. The population density varies between the counties. The smallest acute hospitals serve a population of about 50 000, while the largest serve a population of about 500 000. The sizes of the ICU units and the number of patients with sepsis treated during a year will therefore differ between the included hospitals. This is a national inspection campaign, and hospital in all counties will be inspected. As patients are sampled from different hospitals, a between-hospital variation in measurements is likely, introducing correlated data within the hospitals. To account for this intra-cluster correlation, we will use generalized estimating equations methodology,64 specifying an exchangeable working correlation structure, i.e., any two patients are equally correlated within hospitals regardless of time and intervention and control periods. However, as this assumption might not hold for all hospitals, a method for obtaining cluster-robust standard errors of model parameters will be applied.65 Finally, as our repeated sampling of patients with sepsis may not be entirely representative of the total population, difference in certain patient characteristics, including age and sex, between comparison periods might arise. In that case, the abovementioned models will also include such covariates for obtaining correct model means.

In addition we have added a new paragraph in the discussion section:
The size of the included hospitals differ. The order of the inspections are randomized in clusters of

four hospitals, and all the clusters include hospitals with different sizes. To account for intra-cluster correlation, we will use generalized estimating equations methodology,64 and we will include hospital size as a covariate in our analytic models. Due to Norwegian topography with long travel distances, patients with suspected sepsis are typically sent to the nearest acute hospital for initial diagnosis and treatment. In some cases patients with septic shock can be transferred to a larger hospital later. We know the number of patients with organ dysfunction admitted to the various hospitals, and can adjust for this in our analysis. We do however not have access to data about comorbidity for the included patients and can therefore not fully adjust for case mix differences between the included hospitals. Our process measures cover the initial steps of the diagnostic and treatment process, which is done in all hospitals irrespectively of size. We compare changes in the effect measures before and after the intervention, and the intervention itself should not affect which patients that are admitted to the different hospitals. Consequently, we have no reason to believe that case mix differences between the inspected hospitals should change before and after the intervention.
Minor corrections:
Page 4 line 8; Page 16 – line 3; Page 17 - line 15 and line 32 -again suggest throughout don't refer to sepsis patients but patients diagnosed with sepsis or similar phrase
Page 18 – line 49 – should read "inspections need" as being plural.

We have made corrections in line with what the reviewer has suggested.

VERSION 3 – REVIEW

REVIEWER	Anne-Marie Hill Curtin University, Australia
REVIEW RETURNED	07-Jun-2017

GENERAL COMMENTS	Authors have clarified the outcome measure well for the reader.