PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Promoting leadership and quality improvement through external inspections of management of sepsis in Norwegian hospitals: a focus group study
AUTHORS	Husabø, Gunnar; Teig, Inger Lise; Frich, Jan; Bondevik, Gunnar; Hovlid, Einar

VERSION 1 – REVIEW

REVIEWER	Ejemai Eboreime University of Alberta, Canada
REVIEW RETURNED	22-Jul-2020

GENERAL COMMENTS	The authors present an interesting qualitative study examining how external inspections may foster clinical improvement, using the case of a nationwide inspection of sepsis treatment in emergency departments at Norwegian hospitals. I find the article well written, the conclusions align with the questions, methods and results. Two key issues to consider: 1. The authors should include a section on context/settings to give non-Norwegian readers a better understanding of the Norwegian health system 2. The section on the sepsis inspections, which is the intervention, may be better at the background section

REVIEWER	Ana Cristina Castro University of York, UK
REVIEW RETURNED	24-Jul-2020

GENERAL COMMENTS	Thanks for the opportunity to review this fascinating article reporting on the findings of focus groups exploring the views clinicians, managers, and inspection teams involved with the inspections of sepsis treatment in emergency departments at four different hospitals in Norway. I am adding below some comments that I hope the authors find useful. General comment. Maybe I am too cynical, but the results and the discussion are too optimistic, and perhaps, they do not present a balanced picture of what it means to receive an inspection team. I do not doubt that hospitals respond to inspections, but they impose a burden and put everything under scrutiny, which hospitals not always enjoy. Also, this is focused on particular inspections. I wonder how many more inspections these hospitals receive. How much more data they submit to other bodies performing quality improvement or performance monitoring? That might make a big difference regarding how open they are to the comments from the inspectors.

	Specific comments. (After I finished, I realised that there are page numbers at the top and the bottom of the page. I followed the ones at the top) Abstract Page 3 Lines 10 to 17: I think this is part of the background to your research, not the setting where it was conducted. Line 24: I don't think you need to specify an intervention in this case. That can be explained as part of the setting. Introduction Page 4, lines 33 to 42. I would start this paragraph with the argument that quantitative evidence tends to show that inspections are not effective. (line 36: Moreover, recent research…) and then, highlight that a deeper understanding of the mechanism of change is needed. You should explain why inspections of sepsis care in emergency departments. It sounds very niche, so it is not very clear why you chose that diagnosis and setting. It is briefly explained in the methods section. Still, I wonder whether the argument about prevention of sepsis-related deaths is more about being preventable and the presence of unwarranted variation in early diagnosis and treatment, instead of the number of patients who have sepsis. Line 45: After looking at the supplementary file 1, I am not sure about your sampling method. The four hospitals chosen had a very poor performance in the indicator; therefore, they did not need to do much to improve. If their performance was so poor, why didn't they realise and did something before the inspection happened? I think that having bad performance partly explains why they saw inspections as a positive thing. Methods You should include an example of the questions you used in your focus groups or the topics that were covered. Page 5 Lines 3 to 7: According to the reporting checklist at the end of the document, in this paragraph, you are reporting your qualitative approach and research paradigm, but this is not present here. Lines 25-26. Maybe a reference for "the set of quantitative criteria for recommended diagnosis and treatment of sepsis." Line 27: I'm sorry to be annoying, but why 33 patients? Line 31: How many days does the inspection last? How big are the teams? What are the consequences if a hospital performs poorly in the inspection? We need more context around these inspections and, maybe, the regulatory environment for hospitals in Norway. Last paragraph under "the sepsis inspections". Is there a reference or website that explains the inspection process? It is worth adding it. Page 6 Lines 36 and 37: Should go at the end of the heading "study design". You should also include in that section, any protocol that might have been published in advance and you should mention which (if any) reporting guideline you followed. Line 40: add the name of the company that owns Nvivo Line 48: I am not a fan of using "emerging" to refer to themes. Sorry. It makes it sound like themes come up to the surface like gas bubbles in a champagne glass when, in reality, it takes hard work to think about them. Page 7: I feel I need more context about what kind of clinical meeting these hospitals were holding. I understand the point about their limited ability to see up-to-date data. Still, I am wondering whether they were holding regular morbidity and mortality reviews that could highlight suboptimal care.

	Page 8 lines 43-47: What I gather from this paragraph is that the medical experts in the inspection team have some specialist knowledge, but not so much to be biased, which could be considered as a strength. If that is the point you are trying to make, I think it is not entirely clear. It is open to interpretation. Lines 49-60: It is not clear how. I can see how the quote highlights what you are saying in the paragraph, but if you mention, specifically, what about the feedback from the inspection had a direct impact on patient care, then it would be easy to understand the connection. Discussion I do not have specific comments on the discussion, except that it is not well balanced. Most of the literature presented support your findings and explanations. Most of the quantitative research shows that effects on patient's outcomes are minimal. As I said before, I know that hospitals respond to inspections (and you show it in your results), but after reading this article, I keep wondering, why those hospitals did not realise they were performing so poorly before being inspected?

REVIEWER	Clair Sullivan University of Queensland Brisbane, Queensland Australia
REVIEW RETURNED	17-Aug-2020

GENERAL COMMENTS	Thank you for this paper: ongoing quality improvement is such an important topic. This is a qualitative observation of the perceptions of hospital teams of external quality inspections (focussing on sepsis). The inspections were quite small, with only 33 patient records assessed. This is very small sample size given the high prevalence of sepsis (as quoted in the article) . Was there any regular sepsis outcomes that were assessed in addition to these "deep dive" chart reviews? e.g overall inpatient mortality rate from sepsis in each hospital? Unexpected admission rate to the intensive care unit as a result of sepsis? These rates could be measured in every patient, every time, to avoid sample size issues that the 33 cases in this paper may present. These rates can then be continuously monitored, to allow the impact of any quality improvement efforts to be demonstrated. If the method of monitoring is an infrequent point-prevalence sample size of 33 charts, how will the impact of continuous quality improvement efforts be assessed? I worry that a 33 chart sample size may present a less than representative sample and the focus may be on whether the pathway was followed, rather than patient outcomes? The term "nonconformities" in the paper, made me think of non- adherence to a pathway , rather than adverse patient outcomes. It is hard to imagine the power generated from 33 cases could be statistically meaningfully analysed and benchmarked? What was measured and presented from these 33 patients? How were these measures benchmarked in order to define "best practice"? The method of interviewing staff was robust and much care was taken with sampling. It sounds as though everyone found the

	process positive: were there any neutral or negative comments at all? I would be careful about claiming that the inspections improved care: the paper does not present any data to show that patient outcomes have improved, rather that the hospital teams found this the inspection a positive experience. Did patient outcomes from sepsis actually in association with the inspections (e. g inpatient mortality rates from sepsis)? Thank you for all your work on this,

VERSION 1 – AUTHOR RESPONSE

Response to reviewers

We thank the reviewers for valuable comments.

As the reviewers have raised some issues in common, we will first discuss these. The changes in the manuscript are listed according to the corresponding reviewer comments in the table below, which also includes comments related to issues confined single reviewer comments. Formatting amendments requested by the editor and other changes to the manuscript is listed at the bottom of this document.

General comments

1) The number of electronic health records reviewed

In the original draft, we wrote that the electronic health records (EHR) of 33 recent patients with sepsis had been reviewed. This was unintentionally misleading. The inspection teams sampled data from four different time intervals. For each interval they included 33 patients. The first interval included the most recent patients admitted before 1 October 2015, which was immediately before NBHS announced the inspections. The second interval included the most recent patients before the inspection of each hospital. The third and fourth interval were follow-ups, with new samples at 8 and 14 months after the initial inspections. Our interviews were conducted after the initial inspection, and after the written reports based on the first two patient samples had been presented to the hospitals. The correct number of EHRs is 66 per hospital. (We choose not to include the later follow-up samples here, as these were not a part of the inspections proper and were not reported to the hospitals at the time of interviewing.)

There are two main reasons why EHRs were drawn from different time periods. First, NBHS wanted to account for possible changes in clinical performance over time. Second, the inspections were planned in conjunction with the research project. The data from the assessment of EHRs from all hospitals were subsequently combined in a single data set and used as a basis for quantitative analyses. The time periods for collecting data followed a stepped wedge design where clusters of

hospitals received the inspections in a randomized order within a time period of one year, with the first period serving as a common baseline measurement.

As explained in the manuscript, assessing the EHRs was a laborious effort. The inspection teams started out with a list of patients with ICD-10 codes related to sepsis and infection. To avoid the inclusion of false negative and false positive cases, they first determined whether the patients fulfilled the criteria for having sepsis at arrival. They then had to manually extract data from the EHR related to a number of care processes, diagnostic procedures and blood samples, and enter these into a database. The close scrutiny of the EHRs gave the inspectors a picture of the care provided to the individual patient, and served as an important source of information for assessing the care processes of the emergency departments and identify systemic weaknesses that should be addressed and discussed with the staff and management. Even though not used for statistical inference, the samples of 33+33 records served as sufficient grounds for identifying weaknesses and documenting nonconformities in the care provided. Small sample sizes are in general considered to be acceptable in quality improvement, audit and inspection work when the goal is to demonstrate a gap in system performance, rather than securing precise effect estimates.[1] The follow-up record

reviews at 8 and 14 months after the inspection provided the inspection teams with data to assess the impact of the corrective measures implemented by the hospitals.

The exact number of EHRs per sample (33) was arrived on through power calculations determining the sample size needed for the quantitative analyses performed as a part of the research project.[2]

Replicating and expanding the record review for this particular subgroup of patients using automated data extraction tools would be difficult, given how documentation practices currently are. The data in the EHRs are entered in such a way that it was necessary to make a qualitative judgement in each case regarding the presence and severity of sepsis (SIRS criteria and presence of organ dysfunction). Relying on ICD-10 codes alone would result in the inclusion of too many false positive and false negative cases.[3] Conceivably, NBHS could try to develop an automated process for extracting the data once a list of eligible patients had been produced. However, because the inspections were conducted in 24 different hospitals, each with their own documentation procedures and technical implementation of the nationally mandated EHR system, the manual extraction of data was judged to be a less time consuming and more reliable approach than automating this procedure.

The details and merits of the internal improvement efforts of the hospitals are beyond the scope of the present study. However, we agree with the reviewers that the hospitals should consider developing solutions enabling automatic tracking of such data in a systematic fashion. Sepsis screening has been recommended as a way to aid earlier recognition of sepsis in emergency departments.[4]

2) Performance improvement in the inspected hospitals

In order to focus the present study on the change mechanisms related to the inspections, we have tried not to go into too much details regarding quantitative instruments, data collection procedures and measurements of inspection effects. These questions will be addressed in a later report from the research project. We acknowledge the need for an account of the improvements in the hospitals included in this study, as our sampling approach predicates that improvements have taken place. This is the reason why we included the supplement file with the study.

We have critically reviewed the manuscript and found that an error occurred when copying the numbers into the word table in the supplement file. The correct numbers for Hospital B are 35% pre-inspection and 59% post-inspection, and we have corrected this. Further, we have also concluded that this supplemental file was over-simplified. There is a lack of context here – especially regarding what the average delay was for all hospitals in the study. The time to antibiotics was in fact not particularly delayed in the hospitals we selected into this study. While the time to antibiotics prior to the inspections was 25% on average for all 24 hospitals, average time to antibiotics for “our” four hospitals was 23%. Furthermore, time to antibiotics in our study was comparable to that of previous international studies, as we have explained in a previously published paper from this research project.[5]

Why did not the hospitals realize how poorly they were performing? This is an interesting question. The answer involves more than one factor. Importantly, sepsis diagnosis can be challenging in emergency settings and sometimes sepsis is not recognized early enough.[6] Adding to this, work in emergency departments involve different groups of medical professionals that cooperate under

sometimes stressful conditions, resulting in increased risk of loss of information. An important factor making it especially problematic for emergency departments to conduct a systematic follow-up is

that patients entering the emergency department often are transferred to other departments in the hospital. If the patient has an adverse outcome further down the line, it will not necessarily be evident to the personnel caring for that patient in the emergency department.

These challenges are of course an important reason for why NBHS decided to prioritize this area for inspection. This is an aspect that was not communicated clearly enough in the original draft of this study.

3) The cost of inspections

There are transaction costs of inspections in the form of time and resources spent to provide information to the regulatory authorities.

While there are no other regulatory bodies other than NBHS and the County Governors performing inspections of the services provided by healthcare organizations, there are several different regulators tasked with following up other aspects, such as HSE (Health, Safety and Environment), data protection and food safety. In a wider sense, one could also include the instructions and governance from the central government (in the dual role of owners and issuers of regulation).

In the interviews, we explored both negative and positive experiences with the inspections. Although the degree of enthusiasm varied between the interviewees and a few commented how inspections meant extra work, the positive experiences clearly dominated in the interviews. This can perhaps partly be explained by our selection of inspections with positive developments in care processes. More importantly, however, the aim of this study is to explore change mechanisms. This implies that costs are not a focus of this study. As such, we do not intend this study to be an evaluation of the pros and cons of the inspections.

#	*Comment*	Response
Reviewer 1: Ejemai Eboreime		
1	*The authors should include a section on context/settings to give non-Norwegian readers a better understanding of the Norwegian health system*	Changes to the manuscript: - Included information about the healthcare system and the role of NBHS under the heading “context” (previously named “the sepsis inspections”)

2	*The section on the sepsis inspections, which is the intervention, may be better at the background section*	Changes to the manuscript: - We have moved these paragraphs to the introduction.

Reviewer 2: Ana Cristina Castro		
0	*Maybe I am too cynical, but the results and the discussion are too optimistic, and perhaps, they do not present a balanced picture of what it means to receive an inspection team.* *I do not doubt that hospitals respond to inspections, but they impose a burden and put everything under scrutiny, which hospitals not always enjoy. Also, this is focused on particular inspections. I wonder how many more inspections these hospitals receive. How much more data they submit to other bodies performing quality improvement or performance monitoring? That might make a big difference regarding how open they are to the comments from the inspectors.*	Comment: See general comment 3. Changes to the manuscript: - Included a paragraph in the "strengths and limitations" section mentioning potential costs and negative side-effects of the inspections.
1	*Abstract Page 3 Lines 10 to 17: I think this is part of the background to your research, not the setting where it was conducted.*	Changes to the manuscript: - Included some of the background information in "Objective"
2	*Line 24: I don't think you need to specify an intervention in this case. That can be explained as part of the setting.*	Changes to the manuscript: - Removed accordingly
3	*Page 4, lines 33 to 42.* *I would start this paragraph with the argument that quantitative evidence tends to show that inspections are not effective.* *(line 36: Moreover, recent research…) and then, highlight that a deeper understanding*	Changes to the manuscript: - Revised the paragraph, starting with an overview of the findings from effect studies. - Revised the argument regarding the NBHS' prioritization of sepsis in ED as

	of the mechanism of change is needed. You should explain why inspections of sepsis care in emergency departments. It sounds very niche, so it is not very clear why you chose that diagnosis and setting. It is briefly explained in the methods section. Still, I wonder whether the argument about prevention of sepsis-related deaths is more about being preventable and the presence of unwarranted variation in early diagnosis and treatment, instead of the number of patients who have sepsis.	suggested

4	*Line 45: After looking at the supplementary file 1, I am not sure about your sampling method. The four hospitals chosen had a very poor performance in the indicator; therefore, they did not need to do much to improve. If their performance was so poor, why didn't they realise and did something before the inspection happened? I think that having bad performance partly explains why they saw inspections as a positive thing.*	Comment: See general comment 2. Changes to the manuscript: - Included statistics from the pre- and post-inspection period of all hospitals in supplementary file 2.
5	*Methods* *You should include an example of the questions you used in your focus groups or the topics that were covered.*	Changes to the manuscript: - Included a table under the “participants and data collection” subheading describing interview topics.
6	*Page 5* *Lines 3 to 7: According to the reporting checklist at the end of the document, in this paragraph, you are reporting your qualitative approach and research paradigm, but this is not present here.*	Changes to the manuscript: - Reported our research paradigm (realism) under “study design”
7	*Lines 25-26. Maybe a reference for "the set of quantitative criteria for recommended diagnosis and treatment of sepsis."*	Changes to the manuscript: - Added references to guidelines.
8	*Line 27: I'm sorry to be annoying, but why 33 patients?*	Comment: See general comment 1. Changes to the manuscript: - Corrected the number of patients (66) - Explained the review of EHR more in-depth in supplementary file 2.

9	*Line 31: How many days does the inspection last? How big are the teams? What are the consequences if a hospital performs poorly in the inspection? We need more context around these inspections and, maybe, the regulatory environment for hospitals in Norway.*	Comment: The size of the teams (3 to 4 regular inspectors and 1 expert) has been described in the paragraph above. Changes to the manuscript: - Included information about the healthcare system and the role of NBHS (in the introduction). - Included information about the NBHS' approach to penalization (in the introduction). - Provided a translated version of one of the inspection reports to provide more context (supplementary file 1).

10	*Last paragraph under "the sepsis inspections". Is there a reference or website that explains the inspection process? It is worth adding it.*	Changes to the manuscript: - We have added a reference to a NBHS web pages with information in English (in the introduction) - Provided a translated version of one of the inspection reports to provide more context.
11	*Page 6 Lines 36 and 37: Should go at the end of the heading "study design". You should also include in that section, any protocol that might have been published in advance and you should mention which (if any) reporting guideline you followed.*	Comment: The protocol was referenced in this paragraph, but it was not explicitly mentioned in the text that the reference was to the protocol. Changes to the manuscript: - Reference to ethics committee approval moved accordingly - Explicitly mentioned the study protocol. - Mentioned the SRQR guideline.
12	*Line 40: add the name of the company that owns Nvivo*	Changes to the manuscript: - Added version and name to Nvivo
13	*Line 48: I am not a fan of using "emerging" to refer to themes. Sorry. It makes it sound like themes come up to the surface like gas bubbles in a champagne glass when, in reality, it takes hard work to think about them.*	Changes to the manuscript: - We have replaced “emerging” with less bubbly alternatives under the subheading “transcription and analysis” and in the first paragraph of the results section.

14	*Page 7: I feel I need more context about what kind of clinical meeting these hospitals were holding. I understand the point about their limited ability to see up-to-date data. Still, I am wondering whether they were holding regular morbidity and mortality reviews that could highlight suboptimal care.*	Comment: By law, the Norwegian hospitals are mandated to have a quality management system and to ensure the participation of staff and management in the quality improvement work. Additionally, all hospitals have teams dedicated to assessing patient safety through regular reviews of samples of health records using the Global Trigger Tool methodology. However, these activities are usually not specifically targeted at patients presenting with sepsis to the emergency department. As this is a group of patients who have heterogenous symptoms and often are transferred to other departments, chances are that potential weaknesses in the care processes for these patients may have gone undetected (ref general comment 2) Changes to the manuscript: - Included information in the introduction about the legal requirements regarding the hospitals' quality management systems. - Revised the explanation (now in the introduction) of why NBHS prioritized this area for inspection so that the challenges with sub-optimal care are more clearly explained
15	*Page 8 lines 43-47: What I gather from this paragraph is that the medical experts in the inspection team have some specialist knowledge, but not so much to be biased,*	Comment: Our intention was to underline the importance of the experts' medical expertise and "real-world" experience, and that the

	which could be considered as a strength. If that is the point you are trying to make, I think it is not entirely clear. It is open to interpretation.	inclusion of experts was seen as enhancing the inspections' legitimacy (a point of view shared by hospitals and inspection teams). Changes to the manuscript: - Revised the paragraph in order that this point should come across more succinctly.

16	*Lines 49-60: It is not clear how. I can see how the quote highlights what you are saying in the paragraph, but if you mention, specifically, what about the feedback from the inspection had a direct impact on patient care, then it would be easy to understand the connection.*	Comment: We understand that this paragraph was phrased ambiguously. The implied meaning of "system" here (and in the quote) was that of system analysis/system thinking. The inspection sought to analyze why delays and non-completion of processes happened, and to do so, the inspection teams focused on how the care processes were interconnected. By focusing on the system, the inspection could help the hospitals to identify solutions that would leverage improvement throughout the interconnected care processes. Changes to the manuscript: - The paragraph is rephrased so that it explains the system-level focus of the inspections more clearly.
17	*I do not have specific comments on the discussion, except that it is not well balanced. Most of the literature presented support your findings and explanations. Most of the quantitative research shows that effects on patient's outcomes are minimal. As I said before, I know that hospitals respond to inspections (and you show it in your results), but after reading this article, I keep wondering, why those hospitals did not realise they were performing so poorly before being inspected?*	Comment: See general comments 2 and 3 and our response to comment #14. Changes to the manuscript: - Revised the review of past research of external inspections in the introduction. - Included (in the introduction) more context related to the challenges with recognizing sepsis in the ED setting. - Included a paragraph in the "strengths and limitations" section mentioning potential costs and negative side-effects of the inspections to improve the balance of the discussion.

		- Included statistics related to the pre- and post-inspection of all hospitals in supplementary file 2.

Reviewer 3: Clair Sullivan		
1	*The inspections were quite small, with only 33 patient records assessed. This is very small sample size given the high prevalence of sepsis (as quoted in the article) . Was there any regular sepsis outcomes that were assessed in addition to these "deep dive" chart reviews? e.g overall inpatient mortality rate from sepsis in each hospital? Unexpected admission rate to the intensive care unit as a result of sepsis?* *These rates could be measured in every patient, every time, to avoid sample size issues that the 33 cases in this paper may present. These rates can then be continuously monitored, to allow the impact of any quality improvement efforts to be demonstrated. If the method of monitoring is an infrequent point-prevalence sample size of 33 charts, how will the impact of continuous quality improvement efforts be assessed?*	Comment: Patient outcomes like inpatient mortality rate from sepsis were not covered in the inspections As argued in general comment 1, the recording practices presently in use at the hospitals present an obstacle to developing accurate monitoring routines for this group of patients. Changes to the manuscript: - Corrected the number of patients (66) - Explained the health record review process more in detail in supplementary file 2.
2	*I worry that a 33 chart sample size may present a less than representative sample and the focus may be on whether the pathway was followed, rather than patient outcomes? The term "nonconformities" in the paper, made me think of non- adherence to a pathway , rather than adverse patient outcomes.* *It is hard to imagine the power generated*	Comment: Please see general comment 1. The focus of the inspection was on (timely) completion of processes, rather than patient outcomes. Processes were assessed using either binary measures (e.g. completed or not), or ordinal scales (e.g. within 1 hour, within 2 hours, more than 3 hours etc.).

	from 33 cases could be statistically meaningfully analysed and benchmarked? *What was measured and presented from these 33 patients? How were these measures benchmarked in order to define "best practice"?*	Changes: - corrected the number of patients (66) - provided more background information about the inspections (introduction). - provided a translated version of one of the inspection reports as supplementary file 1 which shows how the inspection used the different measures/performance indicators.

5	*The method of interviewing staff was robust and much care was taken with sampling. It sounds as though everyone found the process positive: were there any neutral or negative comments at all?*	Comment: See general comment 3. Changes to the manuscript: - Included a paragraph in the "strengths and limitations" section mentioning potential costs and negative side-effects of the inspections.
6	*I would be careful about claiming that the inspections improved care: the paper does not present any data to show that patient outcomes have improved, rather that the hospital teams found this the inspection a positive experience. Did patient outcomes from sepsis actually in association with the inspections (e. g inpatient mortality rates from sepsis)?*	Comment: We agree that the study does not show improvements in patient outcomes. We have chosen to use the phrase "quality of care". referencing care that at the time it was given conformed to the practice that could have been expected to achieve the best results.[7] We use the term "care processes" in reference to the actions undertaken with a specific expected medical outcome in the form of diagnosing, maintaining or treating patients. When it comes to the effects of mortality, this will be analyzed in a paper we are currently working on. Changes to the manuscript: - revised the manuscript where needed to ensure consistent usage of these terms

Formatting amendments

1. Please revise the title to include the study's settings (bearing in mind BMJ Open has an international readership)

2. Required Supplementary format:
- Please re-upload your Supplementary files in PDF format.

3. Patient and Public Involvement:

- The title has been changed to: “Promoting leadership and quality improvement through external inspections of management of sepsis in Norwegian hospitals: a focus group study”
- The supplementary files have been uploaded in PDF format
- We have added a new subheading to the methods section describing patient and public involvement

Additional changes

In addition to the changes directly related to the issues raised by the reviewers, we made some minor changes to fix typographical errors and to improve certain phrases. We have also added an acknowledgement section at the end of the manuscript.

Because of the revisions, what was referred to as supplementary file 1 in the original draft is now supplementary file 2 (file: *Supplemental Material S2 inspection findings.pdf*), while supplementary file 1 is the translated version of an inspection report (file: *Supplemental Material S1 inspection report.pdf*).

As per instructions from the BMJ Open editorial office we have also revised the contributorship statement and added a data availability statement in the manuscript document.

References

1. Etchells E, Ho M, Shojania KG. Value of small sample sizes in rapid-cycle quality improvement projects. BMJ Qual Saf 2016;25(3):202-6 doi: 10.1136/bmjqs-2015-005094.

2. 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(9):e016213 doi: 10.1136/bmjopen-2017-016213.

3. Cluzet VC, Lautenbach E. Editorial Commentary: We Are Seeing More Sepsis …. But Are We Seeing the Whole Picture? Clin Infect Dis 2016;62(6):704-06 doi: 10.1093/cid/civ1023.

4. Gatewood MOK, Wemple M, Greco S, et al. A quality improvement project to improve early sepsis care in the emergency department. BMJ Qual Saf 2015;24(12):787-95 doi: 10.1136/bmjqs-2014-003552.

5. 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.

6. Morr M, Lukasz A, Rübig E, et al. Sepsis recognition in the emergency department – impact on quality of care and outcome? BMC Emerg Med 2017;17(1):11 doi: 10.1186/s12873-017-0122-9.

7. Donabedian A. The quality of care. How can it be assessed? JAMA 1988;260 doi: 10.1001/jama.1988.03410120089033.

VERSION 2 – REVIEW

REVIEWER	Ana Cristina Castro University of York United Kingdom
REVIEW RETURNED	29-Sep-2020

GENERAL COMMENTS	I want to congratulate the authors for all the changes that they have implemented in this new revised version of their article. I can see they have put great effort into making it more robust. I have very few comments this time: From page 3 line 43 up to page 4 line 34: I've seen that one reviewer asked you to put more details about the inspections in the introduction, but I think this information would fit better in the methods section. Also, I know we need to know the details, but it feels a bit too long. I do not have strong views about the length of that section, but it might be better if it were more concise. Page 4, line 54: I am not convinced that your study followed a realist paradigm. To the best of my knowledge, realist evaluation seeks to identify CMO (context + mechanism=outcome) configurations when there is a complex intervention under assessment. The goal is to identify middle-level theories that explain why, how, for whom, and under what circumstances an intervention leads to an outcome. However, you are presenting your results as themes that are common to all the hospitals where you conducted your interviews. I am not saying you need to redo your analysis, but, perhaps, the realist paradigm is not the paradigm you are following. Discussion and policy implications: I must insist that these need to be more nuanced. I appreciate that you have added some paragraphs mentioning potential opportunity costs hospitals face when preparing for an inspection. But, the style of an inspection might vary from a tick-box exercise up to a very flexible, adaptive inspection regime. Both extremes and everything in the middle have their pros and cons. If it is a tick-box exercise, the reliability of the inspection might be high, but you miss the complexities of the processes under assessment. If it is a flexible, adaptive regime, reliability might be poor, and it might be difficult to make sense of the findings. As it is presented, it seems like you have found the perfect style of inspections. If that it is the case, please tell us, specifically, what other countries need to do to replicate your results. If it is not the case, we need to know what the specific issues are where this system fails. Related to the previous topic, your policy implications are very broad; therefore, it is not very clear what your specific (i.e. actionable) recommendations are.

REVIEWER	Clair Sullivan University of QLD
REVIEW RETURNED	15-Sep-2020

GENERAL COMMENTS	Thank you for largely addressing the concerns

VERSION 2 – AUTHOR RESPONSE

Comment 1:

From page 3 line 43 up to page 4 line 34: I've seen that one reviewer asked you to put more details about the inspections in the introduction, but I think this information would fit better in the methods section. Also, I know we need to know the details, but it feels a bit too long. I do not have strong views about the length of that section, but it might be better if it were more concise.

Response:

We have moved these details back to the methods section and revised the text in order to make it more concise.

Comment 2:

Page 4, line 54: I am not convinced that your study followed a realist paradigm. To the best of my knowledge, realist evaluation seeks to identify CMO (context + mechanism=outcome) configurations when there is a complex intervention under assessment. The goal is to identify middle-level theories that explain why, how, for whom, and under what circumstances an intervention leads to an outcome. However, you are presenting your results as themes that are common to all the hospitals where you conducted your interviews. I am not saying you need to redo your analysis, but, perhaps, the realist paradigm is not the paradigm you are following.

Response:

The reviewer makes an important point here. We agree that we are not doing a realist evaluation.

Our intention was simply to state that we are subscribing to a realist epistemology and ontology.

The term paradigm can have many meanings. In this paper we have adopted the usage from the Standards for Reporting Qualitative Research guidelines: “identifying the research paradigm (e.g., postpositivist, constructivist/ interpretivist) is also recommended”.

Realism, the way we use term, refers to a middle position in between the purely positivist and purely constructivist camp. That is: realists presume the existence of a ‘real world’. They believe that causal mechanisms exist, but the realists’ focus is on necessities and contingencies rather than regularity.

Furthermore, they view their access to understanding social phenomena as highly complicated: social phenomena are concept-dependent and understanding them involves interpretation and hermeneutics. We believe that this epistemological and ontological outlook is compatible with a multitude of different methodological approaches. The application of realist philosophy does not require a strict configuration of CMOs. We thus maintain that we have been informed by the realist paradigm when conducting this study.

However, we concede that the inclusion of the reference to Pawson’s Science of evaluation might confuse the reader, as Pawson is associated with the methodology for realist evaluation, which follows along the lines sketched out by the reviewer. Therefore, we have substituted Science of evaluation with Andrew Sayer’s Realism and Social Science.

Comment 3:

Discussion and policy implications: I must insist that these need to be more nuanced. I appreciate that you have added some paragraphs mentioning potential opportunity costs hospitals face when preparing for an inspection. But, the style of an inspection might vary from a tick-box exercise up to a very flexible, adaptive inspection regime. Both extremes and everything in the middle have their pros and cons. If it is a tick-box exercise, the reliability of the inspection might be high, but you miss the complexities of the processes under assessment. If it is a flexible, adaptive regime, reliability might be poor, and it might be difficult to make sense of the findings.

As it is presented, it seems like you have found the perfect style of inspections. If that it is the case, please tell us, specifically, what other countries need to do to replicate your results. If it is not the case, we need to know what the specific issues are where this system fails.

Related to the previous topic, your policy implications are very broad; therefore, it is not very clear what your specific (i.e. actionable) recommendations are.

Response:

We have tried to accommodate the reviewer's discerning comments regarding the need for greater nuance in the discussion:

1. In "strengths and limitations", we have specified that we have used "disease-specific" rather than hospital-level indicators to assess which hospitals to include in our study, including a new reference that describes the difference between the two types of indicators in an inspection setting. By noting that we have used disease-specific indicators, we acknowledge that we can not tell whether the quality of care in the hospital as a whole was influenced by the inspections. We have deleted the last sentence, which seemed superfluous as it essentially just conveyed our belief in the study approach.

2. In "interpretations", we have added a new paragraph that addresses the reviewer's request for highlighting the specific limitations of this approach to inspections. We highlight the importance of the organization's responsiveness and learning culture, and we also point to other contextual factors related to the organizations' improvement capabilities. We then reiterate that there are important limitations when it comes to the generalizability of our findings.

3. The last two paragraphs of "policy implications" have been revised in order to provide more specific recommendations:

   We start by explaining that in the realm of inspections of patient care, finding clinically relevant indicators means operationalizing clinical standards, and that such indicators must be reliable in the sense that they can identify substandard performance, and that they must be sensitive to improvement.

   Bringing this recommendation to its conclusion, we suggest that external clinical experts can play an important role throughout the whole inspection process.

   We have removed a portion of the last paragraph describing the importance of "responsive"/balanced approaches. (Though we still believe this to be a useful strategy, we took the reviewer's comments regarding avoiding broad recommendations to heart.)

   Replacing the deleted text from the last paragraph are a couple of sentences that highlight the major possibilities and limitations of the approach used in the sepsis inspections, as per the reviewer's suggestions.

Other changes

We have included a reference to an upcoming study (accepted for publication) that investigates the effects of the sepsis inspections. This reference directs the reader to more information related to the inspection effects.

We have fixed some minor grammatical and orthographic errors.

VERSION 3 – REVIEW

REVIEWER	Ana Castro University of York, UK
REVIEW RETURNED	21-Oct-2020

GENERAL COMMENTS	Thanks for the opportunity to review this article. The authors have done a marvellous job addressing the comments.