PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (see an example) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	The Orthopaedic Error Index: development and application of a novel national indicator for assessing the relative safety of hospital care using a cross-sectional approach
AUTHORS	Panesar, Sukhmeet; Netuveli, Gopalakrishnan; Carson-Stevens, Andrew; Javad, Sundas; Patel, Bhavesh; Parry, Gareth; Donaldson, Liam; Sheikh, Aziz

VERSION 1 - REVIEW

REVIEWER	Adrian Barnett Associate Professor Institute of Health and Biomedical Innovation Queensland University of Technology Australia
REVIEW RETURNED	08-Aug-2013

THE STUDY	The paper on hospital performance by Lilford and Pronovost is cited to provide evidence against not using mortality to judge performance, but the Lilford and Pronovost paper also has some salient points for non-fatal outcomes: 1. What is the ratio of avoidable versus non-avoidable errors (the signal to noise ratio)? If this ratio is low then using errors to judge performance will face the same issues as using deaths. 2. The Lilford and Pronovost paper recommends using error rates, but with a focus on fixing specific problems (with an example of anticoagulants not being given before hip replacement surgery). The approach given here is not concerned with using data to identify fixable problems, but instead is focused on identifying outlying hospitals which then face potential unwarranted stigma, a consequence that Lilford and Pronovost specifically warn against (their paragraph beginning "Public inquiries may be needed from time to time..."). Lilford and Pronovost favour methods that bring about gradual improvements in the whole sector, rather than punitive fixes in individual hospitals or doctors. The data come from reports from staff, therefore we would expect a variation in reporting according to hospital culture. This should be discussed. A hospital that does poorly may just be one that encourages it staff members to highlight safety issues. The validation of the method is implied from it's ability to find one outlying hospital (mid Staffs). Whilst it is comforting that the method found this hospital, the method would need far more scrutiny and validation before it could be widely recommended. When designing systems to find outliers the key thought should always be the consequence of being named as an outlier. In this

	case the action is 'closer scrutiny', but there are costs to this, both financial (in terms of the time needed for the closer scrutiny) and reputational costs for the hospital named. These issues should be discussed.
RESULTS & CONCLUSIONS	The funnel plot in Figure 2 shows a clear association with larger hospitals having fewer errors (in an almost linear association). This pattern is not discussed. As the best performers are the largest, the plot creates an argument for shifting orthopaedic surgery to large hospitals only, but there may be logistical or geographical difficulties for doing this, or differences between large and small hospitals in terms of procedure types. These types of differences create an argument for adjusting for patient case-mix. I would guess that if hospital size were adjusted for the three outliers below 2,000 procedures would come closer to the mean. The information given in table 1 may be sufficient for someone (e.g., a journalist) to name the five hospitals. This could be achieved because the number of orthopaedic procedures are likely to be unique. I would recommend rounding these numbers to make them non-identifiable. I also note that Table 1 is not discussed, so perhaps it could even be removed.
GENERAL COMMENTS	Minor comments - Abstract. I wouldn't think about exporting such a system internationally until it has been robustly validated. Also on page 14 the paper is referred to as 'exploratory work'. - Article summary. "Surveillance of risk through patient safety reporting systems has not been undertaken to date" I don't agree with this. For example, the paper: "Evaluating the performance of Australian and New Zealand intensive care units in 2009 and 2010" Statistics in Medicine, http://dx.doi.org/10.1002/sim.5779. I also know that hospitals are routinely compared in terms of their infection rates. - Page 8, line 28, "a baseline for assessing and evaluating interventions", a good point. The next sentence ("This is necessary...") needs to be better linked in with the previous paragraph. - "Monitoring time trends" is mentioned on page 8, but there's no further discussion of this issue. - Figure 1 shows a clear outlier that deserves more scrutiny. It may uncover a data collection issue. - The denominator is only approximated (page 9). There should be some discussion of the impact of this approximation.

REVIEWER	Daniel Ray Director of Informatics University Hospital Birmingham NHS FT England
REVIEW RETURNED	08-Aug-2013

THE STUDY	I think the authors have done a good job in moving patient safety analysis forward away from just looking at death and HSMRs. This is a good idea and a good advancement. I think that the links between the two datasets HES and NRLS as they stand are tenuous in order to draw valid assumptions from as they stand. For me I do not feel that enough detail has been given as to what constitutes a Trauma and Orthopaedic 'procedure'. In HES data subcutaneous injections

	are recorded as procedures as well as hip replacements for example. Are only theatre based procedures included? Its also not clear if the incidents or 'errors' as the authors refer to them as, were actully linked from the HES records to the NRLS records (which essentially come from hospitals Datix systems) are they directly linked? if not they should be so that the authors can be sure that the incidents occurred are directly related to the procedures in terms of dates for example. Furthermore greater detail could be paid to the subcategory of incidents that have occurred and some mapping to the procedures that the patients have had to identify whether the incidents that have occurred actually relate to the procedures the patients have had.

VERSION 1 – AUTHOR RESPONSE

Comment 1:
Interestingly, the two reviewers take different lines on identification of the hospitals. One is identified as Stafford in the discussion, and data that could identify them is included, but they are not named. We are happy to leave the decision over anonymising the data up to you. We disagree with Adrian Barnett's suggestion below to change/round up the data however, as this reduces the scientific accuracy of the results. We think you should either include it as is or omit it. If you do decide to name the hospitals, please give us your view on whether this might prompt legal action by either the hospitals or by affected patients. If this does seem to bear risk we could ask BMJ's libel lawyer to look at the revised paper. We have assumed that you have already weighed this up in the light of past and current studies about the safety and quality of identifiable NHS units.

Response:
We have removed Table 1 to minimise the risk of disclosing the identity of the hospitals. Please see below the response to comment 6 by Dr Adrian Barnett.

Comment 2:
Separately, one of the reviewers points out that the paper is about *voluntarily reported* incidents. So it's possible that reporting rate is related to hospital culture and degree of openness and a high rate isn't necessarily bad. Could you discuss this point more fully?

Response
Please see below the response to comment 2 by Dr Adrian Barnett in which we mention the following: ‘We are also cognisant of the fact that there is likely to be a variation in reporting according to the patient safety culture within hospitals; so a hospital with a high OEI may be one that has an open culture and encourages staff to report patient safety incidents. Of equal importance is the fact that the NRLS was a voluntary reporting system until April 2010, when mandatory reporting was introduced for serious untoward incidents.’

Comments from reviewer 1: Dr Adrian Barnett

Comment 1:
The paper on hospital performance by Lilford and Pronovost is cited to provide evidence against not using mortality to judge performance, but the Lilford and Pronovost paper also has some salient points for non-fatal outcomes:

1. What is the ratio of avoidable versus non-avoidable errors (the signal to noise ratio)? If this ratio is low then using errors to judge performance will face the same issues as using deaths.

Response:
The purpose of this exercise did not include classifying incidents as avoidable versus non-avoidable. We have added the following sentence to paragraph 2, page 12
'Although we did not distinguish between avoidable and non-avoidable incidents in our analysis, previous work by the group has shown that most orthopaedic incidents that result in harm could have been prevented if safety measures had been implemented[21].'

2. The Lilford and Pronovost paper recommends using error rates, but with a focus on fixing specific problems (with an example of anticoagulants not being given before hip replacement surgery). The approach given here is not concerned with using data to identify fixable problems, but instead is focused on identifying outlying hospitals which then face potential unwarranted stigma, a consequence that Lilford and Pronovost specifically warn against (their paragraph beginning "Public inquiries may be needed from time to time..."). Lilford and Pronovost favour methods that bring about gradual improvements in the whole sector, rather than punitive fixes in individual hospitals or doctors.

Response:
We would like to stress that system-wide outlier analysis is one lens with which we can view the safety of a hospital. This method must never be used in isolation. The following addition has been made to the start of paragraph 3, page 12:
'The process of identifying outliers could be associated with stigma and extensive resource allocation, both financial and reputational. [11] Nevertheless, in organisations that foster a culture of reporting and learning, this method should be viewed as one of the many tools that need to be used in parallel to understand how unsafe the care is in that particular organisation. Use of outlier analysis in singularity does not ensure safe care; it merely acts as a trigger for further checks.'

Comment 2:
The data come from reports from staff, therefore we would expect a variation in reporting according to hospital culture. This should be discussed. A hospital that does poorly may just be one that encourages it staff members to highlight safety issues.

Response:
Agree with this point and changes have been made to paragraph 1, page 14 as follows:
'We are also cognisant of the fact that there is likely to be a variation in reporting according to the patient safety culture within hospitals; so a hospital with a high OEI may be one that has an open culture and encourages staff to report patient safety incidents.[31] Of equal importance is the fact that the NRLS was a voluntary reporting system until April 2010, when mandatory reporting was introduced for serious untoward incidents.[32]'

Comment 3:
The validation of the method is implied from its ability to find one outlying hospital (Mid Staffs). Whilst it is comforting that the method found this hospital, the method would need far more scrutiny and validation before it could be widely recommended.

Response:
We now make clear when considering the interpretation of the findings that further validation work is necessary before this approach can be rolled out. This point has been made to the last paragraph on page 12 – 'However, the work requires greater scrutiny and validation; the purpose of this undertaking was to see if national patient safety reporting systems can be used for surveillance of unsafe care.'

Comment 4:
When designing systems to find outliers the key thought should always be the consequence of being named as an outlier. In this case the action is 'closer scrutiny', but there are costs to this, both financial (in terms of the time needed for the closer scrutiny) and reputational costs for the hospital named. These issues should be discussed.

Response:
Please see response to comment 1.

Comment 5:
The funnel plot in Figure 2 shows a clear association with larger hospitals having fewer errors (in an almost linear association). This pattern is not discussed. As the best performers are the largest, the plot creates an argument for shifting orthopaedic surgery to large hospitals only, but there may be logistical or geographical difficulties for doing this, or differences between large and small hospitals in terms of procedure types. These types of differences create an argument for adjusting for patient case-mix. I would guess that if hospital size were adjusted for the three outliers below 2,000 procedures would come closer to the mean.

Response:
We have added the following sentence before the discussion:
'Of note, there is an almost linear association with larger hospitals having fewer errors' and discussed this further in the strengths and limitations section (pages 14 - 15): 'In Figure 2 we showed that large hospitals (number of orthopaedic procedures) are associated with fewer errors. This must be interpreted with caution as we have not been able to adjust for patient or procedure case-mix due to paucity and anonymity of the data. Based on work elsewhere, it has been stipulated that specialised surgical services should be provided in tertiary hospitals, although geographical or logistical impediments may occur.[33] We cannot make this claim based on our findings.'

Comment 6:
The information given in table 1 may be sufficient for someone (e.g., a journalist) to name the five hospitals. This could be achieved because the number of orthopaedic procedures are likely to be unique. I would recommend rounding these numbers to make them non-identifiable. I also note that Table 1 is not discussed, so perhaps it could even be removed.

Response:
We have removed Table 1 to minimise the risk of disclosing the identity of the hospitals.

Comment 7:
- Abstract. I wouldn't think about exporting such a system internationally until it has been robustly validated. Also on page 14 the paper is referred to as 'exploratory work'.

Response:
The abstract has been revised to omit any mention of exporting this approach until further validational work has been undertaken: 'Further validation and scrutiny of the method will be required to assess its potential to be extended to other hospital specialties in the UK and also internationally to other health systems that have comparable national databases of patient safety incidents.'

Comment 8:
- Article summary. "Surveillance of risk through patient safety reporting systems has not been undertaken to date" I don't agree with this. For example, the paper: "Evaluating the performance of Australian and New Zealand intensive care units in 2009 and 2010" Statistics in Medicine, http://dx.doi.org/10.1002/sim.5779. I also know that hospitals are routinely compared in terms of their infection rates.

Response:
Agreed and amended accordingly: '• Surveillance of risk through routinely collected patient safety incidents reported to national patient safety reporting systems has been undertaken in limited situations.'

Comment 9:
- Page 8, line 28, "a baseline for assessing and evaluating interventions", a good point. The next sentence ("This is necessary...") needs to be better linked in with the previous paragraph.

Response:
We have omitted the last sentence as the point about large-scale incident reporting systems has been made in the previous paragraph.

Comment 10:
- "Monitoring time trends" is mentioned on page 8, but there's no further discussion of this issue.

Response:
This point has now been discussed further in the second paragraph on page 13: 'Monitoring trends in unsafe care over time would be invaluable. They would help, in addition to identifying outliers, in evaluating the effect of safety initiatives and case-mix of patients during different periods of the year.'

Comment 11:
- Figure 1 shows a clear outlier that deserves more scrutiny. It may uncover a data collection issue.

Response:
The limitations of the data have been discussed in detail in the section on strengths and limitations on page 14.

Comment 12:
- The denominator is only approximated (page 9). There should be some discussion of the impact of this approximation.

Response:
The last paragraph on pages 15-16 discusses this:

'The main limitations are those inherent to any secondary analysis of data, including absence of specific information needed and necessities of using proxies. Ideally, we would have preferred to link HES data to corresponding NRLS incidents. At present, this is not possible, as the latter does not allow for patient identification via NHS identification numbers. HES data will also give an approximation of orthopaedic and trauma procedures due to coding inaccuracies. However, these

are, we believe, largely mitigated in the present analysis by the fact that the data were collected to study error and we refer to our analyses as secondary only because the analysis approach we employed was unanticipated when the study was designed. However, the OEI has several potential limitations. Reporting of patient safety incidents is a subjective exercise and variation in the dataset is bound to exist. Biases also exist at several levels: reporting of harmful versus non-harmful events and correct classification of categories of harm. Underlying factors for these biases, such as level of patient safety culture within institutions, were not assessed. Further work on measuring the extent and likely impact of such biases is therefore now needed.'

Comments from reviewer 2: Dr Daniel Ray

Comment 1:
I think the authors have done a good job in moving patient safety analysis forward away from just looking at death and HSMRs. This is a good idea and a good advancement. I think that the links between the two datasets HES and NRLS as they stand are tenuous in order to draw valid assumptions from as they stand. For me I do not feel that enough detail has been given as to what constitutes a Trauma and Orthopaedic 'procedure'. In HES data subcutaneous injections are recorded as procedures as well as hip replacements for example. Are only theatre-based procedures included?

Response:
Regarding the point about linking HES and the NRLS, we have answered this below. We were interested in any procedure classified under the category of 'Trauma and Orthopaedics' as provided by HES and the NRLS. We have used specialty rather than procedural data to be more encompassing. HES data is not limited to theatre-based procedures, nor is NRLS data.

Comment 2:
It's also not clear if the incidents or 'errors' as the authors refer to them as, were actually linked from the HES records to the NRLS records (which essentially come from hospital Datix systems) are they directly linked? If not they should be so that the authors can be sure that the incidents occurred are directly related to the procedures in terms of dates for example.

Response:
The NRLS is an anonymous reporting system and even though Datix systems are the primary providers of both HES and NRLS records, the latter do not have an NHS identification number. No linkage is therefore possible. This is now explained when considering the limitations of this work (paragraph 2, page 15). We also discuss why the lack of linkage should not be a barrier to exploring the potential of a patient safety reporting system to allow for surveillance of unsafe care.

Comment 3:
Furthermore greater detail could be paid to the subcategory of incidents that have occurred and some mapping to the procedures that the patients have had to identify whether the incidents that have occurred actually relate to the procedures the patients have had.

Response:
Again, this is not possible. Please see point above in response to comment 2.

Comment 4:
I think the authors will be asked to detail the actual hospitals in my view these should be included in

the paper. Analysis presented could be expanded to show types and category of incidents that have occurred and also what procedures are included.

Response:
As noted above, we have decided at this stage not to disclose the identities of the specific hospital

VERSION 2 – REVIEW

REVIEWER	Daniel Ray Director of Informatics University Hospital Birmingham NHS FT England
REVIEW RETURNED	10-Oct-2013

GENERAL COMMENTS	I have 1 concern still not addressed, orthopaedic procedures are referred to a lot but it is still not stated what one is defined as in their analysis. Is it any patient with a present OPCS code in OP1 field of the HES data with a treatment specialty code of 110? or is the AHRQ listed theatre procedures used as an operative procedure? addressing this simply would give the reader much more understanding as to the context. I mentioned this in my initial review and it has not been addressed. I think the authors have done a good job in addressing all other concerns especially the limitations, linkage and context, it is not perfect but a good step in the right direction for assessing quality of safe care delivery. Although not perfect will drive forward discussion and debate on improvements to data flows that will address the current limitations. If the first point is addressed I would accept.

VERSION 2 – AUTHOR RESPONSE

Comment:
I have 1 concern still not addressed, orthopaedic procedures are referred to a lot but it is still not stated what one is defined as in their analysis. Is it any patient with a present OPCS code in OP1 field of the HES data with a treatment specialty code of 110? or is the AHRQ listed theatre procedures used as an operative procedure? addressing this simply would give the reader much more understanding as to the context. I mentioned this in my initial review and it has not been addressed. I think the authors have done a good job in addressing all other concerns especially the limitations, linkage and context, it is not perfect but a good step in the right direction for assessing quality of safe care delivery. Although not perfect will drive forward discussion and debate on improvements to data flows that will address the current limitations. If the first point is addressed I would accept.

Response:
We have added the following to the second paragraph on page 9:

‘‘We defined an orthopaedic procedure as any patient entry that involves an OPCS (Office of Population Censuses and Surveys Classification of Interventions and Procedures) code in the OP1 field of HES with a treatment speciality code of 110.’’