PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The nature of adverse events with opioids in hospitalized patients: a post-hoc analysis of three patient record review studies.
AUTHORS	Schutijser, Bernadette; Jongerden, Irene; Klopotowska, Joanna; Moesker, Marco; Langelaan, Maaike; Wagner, Cordula; de Bruijne, Martine

VERSION 1 – REVIEW

REVIEWER	David Williamson Université de Montréal
REVIEW RETURNED	13-Apr-2020

GENERAL COMMENTS	General comments: The subject of this paper is both interesting and pertinent. However, as the number of events reported in the study sample is small (28 patients), the conclusions are limited. Why were weren't cases with a lower likelyhood reported? Specific comments: Methods Page 4 Line 53: How were the patient records randomly selected? Please specify. Page 5 Line 19: Did the physicians reviewignt he ADE receiving any sort of training? Page 5 Line 29: Was the scoring system validated? How sensitive and specific is it? Page 5 Line 44: How can you affirm that a score of 4-6 signifies a 50% chance or greater of being potentially preventable? How was this determined? Page 6 Line 30-21: This sentence is unclear. Please re-write. Page 6 Line 39-40: We was the inter-rater reliability for nurses reported in the results Results Page 6 Line 51-52: This is the overall agreement for the entire study. Do you have the results for the ORADES? Discussion P8 Line 11-13: Should we have selected the cases with causality likelihood scores of 1-3 as well, then we could determine at least 2500 additional cases on whether medication. Given the small number of reported cases, the paper would benefit from a description of the events. and opioids were related.

	P9 Line 17: The positive agreement between physicians for detecting and assessing positive agreement was fair at best. This should be added to the limits.

REVIEWER	Nicole Heneka University of Technology Sydney
REVIEW RETURNED	21-May-2020

GENERAL COMMENTS	Thank you for the opportunity to review this paper which explores an important patient safety topic. Please find some considerations for the authors below. Methods The methods are described in detail, however, there are some aspects that would benefit from further explanation. My main concern with the methods is the absence of established taxonomies and ratings to classify causality, preventability, attributable factors and patient harm. It is often challenging to determine how the authors came to attribute these factors to the ADEs. How was the scoring system to identify causality developed? Is it based on an existing system currently in use? Similarly, is the preventability score based on established criteria or developed by the authors? There are some inconsistencies in Box 1 regarding what has been classified as preventable versus non-preventable (see further comment below). What was the process for determining if the ADE contributed to the patient's death or harm? How were attributable factor domains determined? Was an established taxonomy used? Results The authors report only ADEs resulting in severe patient harm were included. There is wide variability in the harm experienced by patients in the ADEs reported in Box 1. How was patient harm classified? The use of a tool to clearly identify the level of harm is recommended, e.g., the NCCMERP Index for Categorizing Medication Errors. A number of presumptions have been made regarding whether harm was directly attributable to an opioid. Box 1 reports the ADE's analysed in this study. Why don't all the drug events list the dosage of opioid prescribed versus administered? What is considered a 'too high dose'? The ordered and administered doses should be listed for all ADEs. It would be beneficial to include the nature of the error made for each ADE, e.g., prescribing error, administration error, etc., as this is discussed at length in the Discussion. Are there standardised opioid prescribing guidelines against which the ADEs have been checked? Multiple examples state too high a dose was given, or the dose was inappropriate, but dosage ordered or administered is not reported. How was the judgement made that the dose was too high? Case 4: What was the PRN opioid order? How did this differ from what was administered? Did the patient experience any symptoms of opioid toxicity? Case 5: States the type of analgesic is unknown, how have the authors determined this is an opioid? Case 6: Please expand on the pump mode units – 8 and 13 of what? The authors state 'this possibly resulted in an epileptic insult requiring ventilation'. How was this determined?

	Case 9: No dosage of morphine given, so '2-6 injections a day' is ambiguous. It is unclear why the majority of the ADEs classified as non-preventable were identified as such. Non-preventable ADEs are generally defined as harmful and unintended reactions to a drug after its appropriate use. Many of non-preventable ADEs resulted in opioid toxicity, but there is not enough information to determine whether this was due to errors in prescribing or administration, or a reaction to the drug following appropriate use. Some patients clearly had previously unknown sensitivities to opioids (non-preventable), but in other cases, e.g., case 17, the wrong dose of opioid was administered; why was this classified as non-preventable? I recommend the ADEs are more consistently reported so it is clear to the reader how decisions on causality and preventability were made, and a standardised taxonomy for classifying patient harm is used. Discussion The review studies were conducted starting in 2008. Have there been any changes to the health system specific to opioid management policy/protocols etc. in that time, and how might this influence the patterns of ORADEs the authors have identified?

VERSION 1 – AUTHOR RESPONSE

Reviewer #1 comments
We thank David Williamson for carefully reading our manuscript. Please find below our response to each of the points raised (the reviewer's comments are in italics).

Comments raising issues with our manuscript:
1. The subject of this paper is both interesting and pertinent. However, as the number of events reported in the study sample is small (28 patients), the conclusions are limited. Why were weren't cases with a lower likelihood reported?

Reply:
We fully agree that the number of events are small. In our study, we focused on cases with a higher likelihood, i.e. a causation rating of at least 4 on a 6-point Likert scale. This method is in line with other international studies (Baker et al., 2004; Brennan et al., 1991). Furthermore, we maintained this definition in all of the adverse events (AEs) studies of our research group, including the studies of 2008, 2011/2012 and 2015/2016 described in this manuscript. Should we have reported cases with a lower likelihood scale (3 or less on the 6-point Likert scale), then these cases could not be defined as real AEs, and the results could be confusing. We changed a sentence about this in the Method section, Design and setting (page 4): "A detailed description of the methodology used in these studies was previously published and comparable to other international AEs studies".

2. Page 4 Line 53: How were the patient records randomly selected? Please specify.

Reply:
The selection of patient records was conducted by the participating hospitals. We provided clear instructions and the inclusion and exclusion criteria. To verify whether the sample of records

was representative, we also asked the hospitals to compare the sample data with the summary data of the relevant period of their hospital. Besides, the sample consisted also of some extra records, in the case a record could not be retrieved. We included a sentence about the selection in the Method section, Design and setting (page 5): “The random selection of patient records was conducted by the participating hospitals with clear instructions of the researchers.”

3. Page 5 Line 19: Did the physicians reviewing the ADE receiving any sort of training?

Reply:
Indeed both nurses and physicians received multiple training moments prior to the reviewing of patient records. The training included a detailed explanation of the study method, the screening process, the database used and a discussion of previous (anonymous) cases to enhance the quality of the review process. During the reviewing process, physicians also had reflection moments, to discuss current cases. We included a sentence about the training in the Method section, Review procedure AE studies (page 5): “Prior to the study, both nurses and physicians had training sessions in which cases were discussed to enhance the quality and standardization of the review process.”

4. Page 5 Line 29: Was the scoring system validated? How sensitive and specific is it?

Reply:
The scoring system has been used in different studies on adverse events for over a decade, such as the Canadian Adverse Events Study.(Baker et al., 2004) The scoring system has not been validated. Reliability has been tested by double reviewing 10% of the patient records. Positive and negative agreement frequencies are moderate. More detailed information about the inter-rater reliability is presented in Supplemental Table 1. We changed a sentence about this in the Method section, Review procedure AE studies (page 5): “Validity of this scoring system has not been tested, but it has been used widely in AE studies for over 20 years and the ratings of the system did not change in that time.”

5. Page 5 Line 44: How can you affirm that a score of 4-6 signifies a 50% chance or greater of being potentially preventable? How was this determined?

Reply:
Before the physicians answered the question whether in their opinion an AE was potentially preventable or not, they were required to respond to 13 questions to add more structure to the review process. For example, if there was a complex medical history, if the patient had co-morbidity and whether another physician would repeat this treatment. These questions gave the physicians an impression how to judge the preventability of an AE. The scores on this question were also a 6-point Likert scale and was previously used in the Canadian Adverse Events Study.(Baker et al., 2004) We trusted the preventability scores of the physicians and, congruent with other studies, concluded that a judgment of 4-6 equals a more than 50% chance or greater of being potentially preventable. We added some sentences about this in the Method section, Review procedure AE studies (page 5): “Before the physicians answered the question about preventability, they were required to respond to 13 questions to add more structure to the review process. For example, if there was a comlex medical history, if the patient had co-morbidity and whether another physician would repeat this treatment.”

6. Page 6 Line 30-21: This sentence is unclear. Please re-write.

Reply:
We assume the reviewer meant the sentence about the comparison between preventable and non-preventable ORADEs. We have rewritten this sentence in the Outcomes section (page 6): "Furthermore, in order to identify risk factors, we compared the outcome variables between preventable and non-preventable ORADEs."

7. Page 6 Line 39-40: We was the inter-rater reliability for nurses reported in the results

Reply:
We think the reviewer asks about the not presented inter-rater reliability of the nurses. In this study, we presented positive and negative agreement frequencies. The interpretation of the Kappa is not straightforward, and it is influenced by the number of categories of each variable and the prevalence of the given scores. It is therefore possible that despite a high agreement, the Kappa is low. This occurs in studies with few adverse events. For this reason we chose to present positive and negative agreement frequencies. It helps to answer questions such as: 'if one expert finds a preventable adverse event, what is the probability that another expert will also find a preventable adverse event?' All agreement frequencies are presented in Supplemental Table 1, but we agree with the reviewer that the agreement frequencies for nurses can be presented for completeness. We added them in the Results section (page 7): "When detecting the predefined triggers, positive agreement between nurses varied between 76.0-91.5%."

8. Page 6 Line 51-52: This is the overall agreement for the entire study. Do you have the results for the ORADES?

Reply:
The agreement frequencies are indeed for the entire adverse event studies. To determine these frequencies, 10% of all patient records were assessed twice by nurses and physicians. It is unknown whether records with ORADEs were among these 10%. Besides, the aim of these studies was to identify AEs and ADEs in Dutch hospitals. Since this study is a post-hoc analysis of these three patient record review studies, it is not possible to present these frequencies on such detailed level. If the primary aim of the studies was to determine ORADEs, then in would certainly be interesting to present the agreement frequencies for ORADEs only.

9. Page 8 Line 11-13: Should we have selected the cases with causality likelihood scores of 1-3 as well, then we could determine at least 2500 additional cases on whether medication. Given the small number of reported cases, the paper would benefit from a description of the events.
and opioids were related.

Reply:
We agree that it would be interesting to determine whether among those 2500 additional cases would be cases with opioids. However, when we designed the study and our research question, we did not know the number of ORADEs would be as low as 28. With today's knowledge we probably would have changed the research question and method up front, but we do not think it is ethical to do this afterwards. Another consideration is that we wanted to stay true to the definition of AEs, which we used in all three AE studies and which was also used in previous international studies. An AE was defined as having a causation rating of at least 4 on the 6-point Likert scale. Cases with causation

ratings of 1-3 can therefore not be defined as AE. It would be too uncertain whether opioids really caused the patient harm. We added a sentence about our considerations in the Discussion section (page 8): “However, we did not determine these 2500 cases, since we wanted to stay true to the definition of an AE (at least 4 on the 6-point Likert scale) and we did not consider it ethical to change the method of the study afterwards.”

10. Page 9 Line 17: The positive agreement between physicians for detecting and assessing positive agreement was fair at best. This should be added to the limits.

Reply:
We agree with the reviewer that the agreement frequencies may have limit the results and added a sentence about this in the Discussion section (pages 9-10): “Secondly, overall agreement frequencies between physicians were moderate. This could have led to different assessments or different scores if other experts were involved. This should be taken into account when interpreting our results. However, a previous review of studies focusing on assessing AEs showed also moderate to substantial inter-rater reliability. For this reason, patient records in all Dutch AE studies have been assessed by the same experts as much as possible and over the years these experts have not become stricter or lenient in their judgment of AEs and their preventability.”

Reviewer #2 comments
We thank Nicole Heneka for carefully reading our manuscript. Please find below our response to each of the points raised (the reviewer’s comments are in italics).

Comments aggregated since no issues were raised:
Thank you for the opportunity to review this paper which explores an important patient safety topic.

Reply:
We thank reviewer #2 for this feedback and appreciate her opinion that our manuscript explores an important safety topic.

Comments raising issues with our manuscript:
1. The methods are described in detail, however, there are some aspects that would benefit from further explanation. My main concern with the methods is the absence of established taxonomies and ratings to classify causality, preventability, attributable factors and patient harm. It is often challenging to determine how the authors came to attribute these factors to the ADEs.
How was the scoring system to identify causality developed? Is it based on an existing system currently in use? Similarly, is the preventability score based on established criteria or developed by the authors? There are some inconsistencies in Box 1 regarding what has been classified as preventable versus non-preventable (see further comment below).

Reply:
The scoring system to identify causality and preventability was, as described in the response for reviewer #1, originally developed by Brennan et al.(Brennan et al., 1991) and slightly adapted and used in previous international studies on adverse events (AEs), such as Baker et al.(Baker et al., 204) and Vincent et al.(Vincent et al., 2001) Both Likert-scales with the 6-point distribution of causality and preventability were also used in our national patient record review studies. Hence, these scales are widely used in AE studies for over 20 years and the ratings of the systems did not change in that

time. We included sentences about this point in the Method section, Design and setting (page 4): “A detailed description of the methodology used in these studies was previously published and comparable to other international AEs studies.”

2. What was the process for determining if the ADE contributed to the patient’s death or harm?

Reply:
Determining whether the ADE contributed to the patient’s death have been retrospectively assessed based on the judgment of the experienced and trained physicians who had insight in all the data in the patient records. The assessment was made by using various extra questions to add more structure to the review process. No additional information was requested from
the involved healthcare professionals to ensure their privacy. A retrospective judgment can always lead to “hindsight bias”: being familiar with the final outcome can influence the judgment. Therefore, the results should be interpreted with caution. To explicit this, we used the term ‘probably/possibly’ in the whole manuscript.

3. How were attributable factor domains determined? Was an established taxonomy used?

Reply:
The attributable factors were assessed based on the judgment of the experienced and trained physicians. The factors were previously established in the taxonomy of the Eindhoven Classification Model.(van Vuren et al., 1997) The main categories were: human, organizational, technical and patient-related. We added this taxonomy in the Method section, Review procedure AE studies (page 5): “The attributable factors were based on the taxonomy of the Eindhoven Classification Model and consisted of the main categories: technical, care, organizational, patient related, violation and other.”

4. The authors report only ADEs resulting in severe patient harm were included. There is wide variability in the harm experienced by patients in the ADEs reported in Box 1. How was patient harm classified? The use of a tool to clearly identify the level of harm is recommended, e.g., the NCCMERP Index for Categorizing Medication Errors. A number of presumptions have been made regarding whether harm was directly attributable to an opioid.

Reply:
Indeed, there is wide variability in the harm experienced by patients in the ADEs reported in our study. For all AEs with causality score 4 or higher, severity of harm was classified using several follow-up questions, with one question focusing on the consequences of harm for the patient. Possible categories were whether the harm possibly resulted in an intervention or treatment, a prolonged hospital admission or death. These categories are similar to the NCCMERP Index for Categorizing Medication Errors. However, it would be interesting to further analyze the type of harm by using the NCCMERP. We did not let the physicians categorize the harm according to the NCCMERP, since the focus was not only on medication errors, but on all potential adverse events. In a future study, focusing on ADEs We added a sentence about this in the Results section, Nature of opioid related ADEs: medication errors & phase (page 7): “Finally, the experts assessed the consequences of the ORADEs (multiple options possible). In eight ORADEs, an intervention or extra treatment was needed, in two ORADEs the patients had a prolonged hospital stay and four preventable ORADEs possibly contributed to the death of the patient (cases #5, #6, #8, #9).” Also, we added a sentence in the Strengths and limitation section (page 10): “Furthermore, this was

also the reason that the harm could not be further categorized according to the NCCMERP Index for Categorizing Medication Errors."

5. Box 1 reports the ADE's analysed in this study. Why don't all the drug events list the dosage of opioid prescribed versus administered? What is considered a 'too high dose'? The ordered and administered doses should be listed for all ADEs.

Reply:
We agree with the reviewer that it would help to present the prescribed and administered medication name and dose. However, as stated in the limitations, our post-hoc analysis was based on the information previously recorded by the experts in an AE database, and on the assessment conducted by these physicians. This means that not all information about which drugs were prescribed was collected and recorded in the database. Moreover, it could also mean that information could not be found by the physicians in the patient records. In some cases, the physicians noted this latter reason. Whether a dose was too high, was also assessed by the physicians. We took another look at the database and added any extra information that we could find about the dosages of opioids in Box 1 on pages 18-20.

6. It would be beneficial to include the nature of the error made for each ADE, e.g., prescribing error, administration error, etc., as this is discussed at length in the Discussion.

Reply:
We agree with the reviewer that this could help to overview the ORADEs. We added this in Box 1 on pages 18-20.

7. Are there standardised opioid prescribing guidelines against which the ADEs have been checked? Multiple examples state too high a dose was given, or the dose was inappropriate, but dosage ordered or administered is not reported. How was the judgment made that the dose was too high?

Reply:
The judgment whether a dose was too high or not was made by the reviewing physicians, based on their clinical experience and knowledge of professional standards. Due to the broad nature of this study, no explicit professional standards for clinical details were used in the review process. The reviewing physicians were highly experienced and specialized in surgery, internal medicine or neurology. During the record review studies, they had access to all information in the electronic patient record and we trusted their judgment based on all the information. In some cases where information was insufficient this was recorded by the reviewing physicians. In most cases, they had access to a lot of information. It could have happened that the reviewing physicians made a decision about the dosage, but forgot to record this in the AE database. As said, the primary goal of the record review studies was determining AEs and ADEs and recording specific detailed information such as a prescribed or administered dose may not be recorded. We added a sentence about this in the Method section, Review procedure AE studies (page 5): "The scoring system was used in all three record review studies and the physicians made the judgments about causality and preventability based on all the available information of the patient's condition and taking into account the guidelines."

8. Case 4: What was the PRN opioid order? How did this differ from what was administered? Did the patient experience any symptoms of opioid toxicity?

Reply:
The exact dose could unfortunately not be retrieved from our AE database. As a consequence of the high dose, the patient became drowsy. We added this in Box 1 (page 18): “This resulted in drowsiness.”

9. Case 5: States the type of analgesic is unknown, how have the authors determined this is an opioid?

Reply:
Based on the information in our AE database, both the two junior and two senior researchers assessed this AE as an opioid related AE. Mainly due to the fact that an anesthetic in combination with pain medication was given through an infusion, we concluded that this was a common opioid anesthetics combination, mostly because a combination of
anesthetic with NSAIDs, acetaminophen or other analgesic medications is very uncommon. We added “intravenous” in the text of this case in Box 1 (page 18).

10. Case 6: Please expand on the pump mode units – 8 and 13 of what? The authors state ‘this possibly resulted in an epileptic insult requiring ventilation’. How was this determined?

Reply:
The exact mode unit could not be retrieved from our AE database. However, in our organization the usual concentration of morphine is 1mg/ml with a pump mode of 1-10ml/hour. A pump mode of 13ml/hour is thus too high and too fast and based on this we assume the mode unit is ml/hour. We added this in the case in Box 1 (page 18): “The pump mode was set at 13 ml/hour instead of
8 ml/hour as ordered.

11. Case 9: No dosage of morphine given, so ‘2-6 injections a day’ is ambiguous.

Reply:
We found the exact dosage in the database and added this in case 9 in Box 1 (page 18): “[…] varying from 2-6 subcutaneous injections of 2,5 mg per day along with transdermal fentanyl 12 mcg hourly. Severe hypercapnia eventually caused her death.”

12. It is unclear why the majority of the ADEs classified as non-preventable were identified as such. Non-preventable ADEs are generally defined as harmful and unintended reactions to a drug after its appropriate use. Many of non-preventable ADEs resulted in opioid toxicity, but there is not enough information to determine whether this was due to errors in prescribing or administration, or a reaction to the drug following appropriate use.
Some patients clearly had previously unknown sensitivities to opioids (non-preventable), but in other cases, e.g., case 17, the wrong dose of opioid was administered; why was this classified as non-preventable? I recommend the ADEs are more consistently reported so it is clear to the reader how decisions on causality and preventability were made, and a standardised taxonomy for classifying patient harm is used.

Reply:
As mentioned in a comment to reviewer #1 in this letter, the physicians assessed whether in their opinion an AE/ADE/ORADE was non-preventable. They were required to respond to 13 questions to add more structure to the review process. For example, if there was a complex medical history, if the patient had co-morbidity and whether another physician would repeat this treatment. These questions gave the physicians an impression how to judge the preventability of an AE. The scores on this question were also a 6-point Likert scale and was previously used in the Canadian Adverse Events Study.(Baker et al., 2004) We trusted the preventability scores of the physicians and, congruent with other studies, concluded that a judgment of 3 or less equals less than a 50% chance of being potentially preventable.
We agree with the reviewer that the preventability in case 17 can raise questions. This case has also caught our attention and we have discussed this case intensively. However, we wanted to stay 100% true to the assessment of the physicians. At the moment of the record review, the specific physician had all information in which he/she made the decision to label this case as non-preventable. We think it is not ethical to change that based on the current information only. We already included this point in the limitation section on page 10. Furthermore, we included all dosages that we could additionally find in the database in Box 1, including the type of error (e.g. prescribing, administration, other) on pages 18-20.

13. The review studies were conducted starting in 2008. Have there been any changes to the health system specific to opioid management policy/protocols etc. in that time, and how might this influence the patterns of ORADEs the authors have identified?

Reply:
The patient record review studies were conducted in 2008, 2011-2012 and 2015-2016. In the eight years between them there have been changes in the Netherlands in the opioid management. For example the Dutch guideline 'Recognition and treatment of pain in frail elderly' was developed and updated in these years.(Verenso, 2016) Furthermore, a special website about opioids was developed for health care professionals, patients and researchers. Overall, the temporal view on opioids may have changed in these years. While the current opinion is that prescribing opioids should be minimized due to the harm of opioids, this changed throughout the years and may not have been recognized 15 years ago, when the focus was mainly on alleviating suffering of pain. This change in opinion may have increased alertness when prescribing or administering opioids, which could have led to less ORADEs. However, our study showed that ORADEs still occur and publishing about them could serve as a method of increasing awareness. We added this sentence in the Strengths and Limitation section (page 10): "The current opinion is that prescribing opioids should be minimized due to the harm of opioids, which is supported by updated guidelines. This view changed throughout the years and may not have been recognized 15 years ago, when the focus was mainly on alleviating suffering of pain. This change in opinion may have increased alertness when prescribing or administering opioids, which could have led to less ORADEs. However, our study showed that ORADEs still occur and publishing about them could serve as a method of increasing awareness."

References
Baker, G.R., Norton, P.G., Flintoft, V., Blais, R., Brown, A., Cox, J., Etchells, E., Ghali, W.A., Hebert, P., Majumdar, S.R., O'Beirne, M., Palacios-Derflingher, L., Reid, R.J., Sheps, S., Tamblyn, R., 2004. The Canadian adverse events study: the incidence of adverse events among hospital patients in Canada. Cmaj 170, 1678-1686.

Brennan, T.A., Leape, L.L., Laird, N.M., Hebert, L., Localio, A.R., Lawthers, A.G., Newhouse, J.P., Weiler, P.C., Hiatt, H.H., 1991. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. N Engl J Med 324, 370-376.
van Vuren, W., Shea, C.E., van der Schaaf, T.W., 1997. The development of an incident analysis tool for the medical field. . Eindhoven University of Technology, Eindhoven.
Verenso. Richtlijn Pijn. Herkenning en behandeling van pijn bij kwetsbare ouderen.; Available from: https://www.verenso.nl/_asset/_public/Richtlijnen_kwaliteit/richtlijnen/database/VER-003-32-Richtlijn-Pijn-deel2-v5LR.pdf.
Vincent, C., Neale, G., Woloshynowych, M., 2001. Adverse events in British hospitals: preliminary retrospective record review. Bmj 322, 517-519.

VERSION 2 – REVIEW

REVIEWER	David Williamson Université de Montréal, Canada
REVIEW RETURNED	07-Jul-2020

GENERAL COMMENTS	The authors have answered my comments. I have no further comments.

REVIEWER	Nicole Heneka University of Technology Sydney
REVIEW RETURNED	12-Jul-2020

GENERAL COMMENTS	I welcome to the opportunity to review this manuscript again and thank the authors for their considered and detailed responses to the initial feedback. The changes/additions you have made to the manuscript have clarified all the questions raised in my feedback and have strengthened the paper. Box 1 is now also clearer and it's very helpful for the reader to see the ORADEs include error types. I'm wondering if it is possible to include a supplementary file, or reference, with the 13 questions physicians responded to prior to determining preventability? This would be an interesting, albeit optional, addition to the paper. I wish the authors well in their future research!

VERSION 2 – AUTHOR RESPONSE

Furthermore, we thank the reviewers for carefully reading our manuscript again and we feel that their review contributed in making this a much better manuscript. We are delighted to read that both reviewers suggested publication in BMJ Open.

We agree with reviewer 2 that the 13 preventability questions would be a good addition to the paper and included a Supplemental Table with the questions. These questions were also previously published in the paper of Baines et al. (2013).[1]

Finally, our study was funded by the Ministry of Health, Welfare and Sports which does not provide grant numbers, but only project titles. We added our project title (Monitor Zorggerelateerde Schade 2015-2018) in the manuscript.

Please find attached the revised document of the manuscript and the two Supplemental Tables. We highlighted the minor changes with track changes.

We look forward to hear from you.

Yours sincerely, on behalf of all authors.

Reference:

1. Baines RJ, Langelaan M, de Bruijne MC, et al. Changes in adverse event rates in hospitals over time: a longitudinal retrospective patient record review study. BMJ Qual Saf. 2013;22(4):290-8.