PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	"Errors" and omissions in paper-based early warning scores: the association with changes in vital signs. A database analysis.
AUTHORS	Clifton, David; Clifton, Lei; Sandu, Dona-Maria; Smith, Gary; Tarassenko, Lionel; Vollam, Sarah; Watkinson, Peter

VERSION 1 - REVIEW

REVIEWER	John Kellett Thunder Bay Regional Health Sciences Center, Thunder Bay, Ontario, Canada
REVIEW RETURNED	25-Jan-2015

GENERAL COMMENTS	The word "run" has multiple meanings and it should be replaced with something more specific to what the authors are, I think, trying to say (i.e. a series, or sequence of observations that are stable). I would suggest, therefore, that they define a stable sequence as a sequence of three or more consecutive observations with the same early warning score. This definition needs to be made clear in the Abstract Methods should more clearly state that vital signs were recorded as marks on a paper chart, and (?) not as numbers. Hence, I assume the need for "criteria. I found the Tables extremely difficult to interpret and suggest they are replaced with something like the examples below. Percentages should be included in the tables, and they should be ordered so that the reader can make sense of them (i.e. from the most to the least, either in numbers or errors or alerts etc). Table 1 should abandon the Greek letters and use simple terms such as True Alerts, True Non-alerts, Missed Alerts and False Alerts. Table 2 should also be simplified. Am I right in think that by "Other" in Table 2 the authors mean patients with unstable or fluctuating EWS? Paragraph 3.4 of the Results is close to unintelligible. Table 3 should just leave out the patients with no next set, as they are a very small number and make no difference to the points (I think) the authors are trying to make. In case they are not clear from this document I have appended Tables in attached file Incomplete observations

	True Non-Alerts 15,007 (89.4%) 3,038 (20.4%) True Alerts 1,039 (6.2%) 343 (33.0%) Missed Alerts 515 (3.1%) 217 (42.1%) False Alerts 234 (1.4%) 118 (50.4%) Total 16,795 (100.0%) 3,716 (22.1%) Table 1 Complete Observations With Errors Stable sequence EWS = 0 7,332 (56.1%) 219 (3.0%) Stable sequence EWS >=1 1,108 (8.5%) 308 (27.8%) After stable sequence EWS >=1 298 (2.3%) 83 (27.9%) Unstable EWS 3,637 (27.8%) 1,213 (33.4%) After stable sequence EWS = 0 704 (5.4%) 389 (55.3%) Total 13,079 (100.0%) 2,212 (16.9%) Table 2 Next set Total With True Alert True alert – error 157 (1.2%) 92 (58.6% True alert – no error 539 (4.2%) 292 (54.2%) Missed alert – error 296 (2.3%) 78 (26.4%) False alert – error 116 (0.9%) 17 (14.7%) True Non-alert – error 1,628 (12.8%) 118 (7.2%) True Non-alert – no error 10,185 (79.8%) 431 (4.2%) Total 12,764 (100.0%) 1,028 (8.1%) Table 3

REVIEWER	Mark Niegsch Dept. of Anesthesia Juliane Marie Centeret, Rigshospitalet, Denmark
REVIEW RETURNED	03-Feb-2015

GENERAL COMMENTS	The research question or study objective is more clearly defined in the abstract than in the text. A clear definition of the primary outcome are missing. in Abstract, Design are not described, instead are the the outcome presented under methods. Methods: Study design is sparse. What kind of a study have you conducted? how were staff trained in the use of the EWS? Outcomes need to be clearly defined. References regarding following could be useful. - aggregate score versus sigle parameter score. for the discussion of reasons for not reacting on some scores. - Implementaton processes. - References are up to date. regarding are lacking. - electronic EWS Appendi C Figure 4 and corresponding text adresses an error in the transcription of a EWS chart. as far as i can see the the figure illustrates at the bottom blood-presure and heart rate and the plotting at the topper most likely the temperature with has been documented with a wrongful marking.

	Since outcomes are not described it is difficult to answer [if the results are presented clearly]. Results are presented in the 6th paragraph of the discussion. These should be presented under results. Regarding references adressing EWS chart designing, it should be discussed that there are a great risk og reviewing a hawthorne effect. This often glorifies the intervention but with time the effect could be reduced. Parameters such as implementation and staff education that could influence the results, are not discussed. Just because an EWS system has become a standard of care, does not assure that it is well implemented. There are data published which documents poor implementation even after several years of using such a system. Staff might have good reasons for missing observations which could be sought out. Published papers adresses Nurses' role in detecting deterioration and documents that they use observation to validate intuitive feelings. A study published january 2015 proofs a positive effect on in hospital mortality, with the introduction of an electronic physiological surveillance system. Your observations might aid in the reevaluation of the escalation protokol for the future electronic systems. Thereby minimising the risk of missing deteriorating patients and reducing the unnecessary activation of resources. Strengths and weaknesses are more a part of the introduction than a reflection upon the strengths and weaknesses. Reporting Checklist are missing.

REVIEWER	Prof Michael Buist Faculty of Health University of Tasmania Australia
REVIEW RETURNED	16-Feb-2015

GENERAL COMMENTS	This is an excellent analysis of the epidemiology of errors in recording a EWS in a very large data set. Well written with useful information and a thoriugh discussion. My only concern is the title. This paper to my mind is not about "factors underlying perceived 'errors" " but rather adescription of them.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1: John Kellet.

"I would strongly recommend the following: The word "run" has multiple meanings and it should be replaced with something more specific to what the authors are, I think, trying to say (i.e. a series, or

sequence of observations that are stable). I would suggest, therefore, that they define a stable sequence as a sequence of three or more consecutive observations with the same early warning score. This definition needs to be made clear in the abstract".

We agree this is a clearer terminology and have adjusted our text (and abstract) accordingly.

"Methods should more clearly state that vital signs were recorded as marks on a paper chart, and (?) not as numbers. Hence, I assume the need for criteria".

We have adjusted our text to make it clear which vital signs were recorded as marks.

"I found the Tables extremely difficult to interpret and suggest they are replaced with something like the examples below. Percentages should be included in the tables……."

We are grateful for the considerable effort John Kellet put into clarifying our tables, and have in general followed his recommendations, and included percentages. Where we have deviated we have explained below.

"Table 1 should abandon the Greek letters and use simple terms such as True Alerts, True Non-alerts, Missed Alerts and False Alerts"

We have removed the greek letters, but avoided the use of "true" and "false". As we are documenting degrees of concordance, we believe these descriptions could cause confusion.

"Am I right in think that by "Other" in Table 2 the authors mean patients with unstable or fluctuating EWS?"

We have described what constitutes "other" in the text related to the table (were neither part of a stable series, nor occurred after a series). These observations do not necessarily form part of an unstable series, as a series requires at least three observations by the definitions used in our paper. For example, if there were two observations with differing aggregate scores between two stable series of "0" scores, the second observation would not be in any series type and did not directly follow a stable series. We have therefore not altered our original definition. However, we believe that the use of John Kellet's "stable series" description greatly clarifies our definitions.

"Paragraph 3.4 of the Results is close to unintelligible"

We have re-written this paragraph to improve readability.

"Table 3 should just leave out the patients with no next set, as they are a very small number and make no difference to the points (I think) the authors are trying to make"

We have re-ordered this table and included percentages, but have left the patients with no next observation set in, as the percentages would otherwise not add up, which we believe would prevent the reader from checking our findings. We are however, happy to be guided by the editors.

Reviewer 2: Mark Niegsch

Many of Mark Niegsch's comments result from us being insufficiently clear about our trial design, which we have therefore clarified in the methods section.

“The research question or study objective is more clearly defined in the abstract than in the text. A clear definition of the primary outcome are missing”

See above. We normally specify a “primary outcome” when undertaking primary research for which a power calculation is required. We have made it clearer that we have undertaken a secondary analysis of the CALMS-2 dataset. More importantly, we have specified in the methods section what we sought to determine from our secondary analysis.

“In Abstract, Design are not described, instead are the outcome presented under methods”

Corrected.

“Methods: Study design is sparse. What kind of a study have you conducted? How were staff trained in the use of the EWS?”

See above. We have clarified study design and staff training in the text.

“Outcomes need to be clearly defined”

Specified in the methods section

“Appendix C Figure 4 and corresponding text addresses an error in the transcription of a EWS chart. As far as I can see the figure illustrates at the bottom blood-pressure and heart rate and the plotting at the topper most likely the temperature with has been documented with a wrongful marking”.

These are simply examples of transcription error types, which we have redrawn to preserve anonymity. We have adjusted the legend to reflect the correct parameter.

“Since outcomes are not described it is difficult to answer (if the results are presented clearly)”

See above – we have specified in the methods section.

“Results are presented in the 6th paragraph of the discussion. These should be presented under results”.

As we wrote, each of the results discussed in paragraph 6 is shown in the results tables.

“References regarding following could be useful. Aggregate score versus single parameter score. For the discussion of reasons for not reacting on some scores. Implementaton processes. electronic EWS….parameters such as implementation and staff education that could influence the results, are not discussed. Just because an EWS system has become a standard of care, does not assure that it is well implemented. There are data published which documents poor implementation even after several years of using such a system. Staff might have good reasons for missing observations which could be sought out. Published papers addresses Nurses role in detecting deterioration and documents that they use observation to validate intuitive feelings. A study published january 2015 proofs a positive effect on in hospital mortality, with the introduction of an electronic physiological surveillance system. Your observations might aid in the re-evaluation of the escalation protocol for the future electronic systems. Thereby minimising the risk of missing deteriorating patients and reducing the unnecessary activation of resources”

We have referenced our suggestion that clinicians may use additional information to that available

from the vital signs of the patients. We are grateful that the reviewer has seen Professor Smith's work. There are of course many areas of interest in early warning scores that we could have discussed. However we believe our discussion is balanced and have not further added to it, in the interests of clarity and brevity.

"Reporting Checklist are missing"

We initially did not believe there was an appropriate check list for this study. On reflection, we have used the STROBE checklist, which we found to be the most appropriate for this study, although there are some aspects which are not applicable.

Reviewer 3 Prof Michael Buist

"This is an excellent analysis of the epidemiology of errors in recording a EWS in a very large data set. Well written with useful information and a thorough discussion. My only concern is the title. This paper to my mind is not about "factors underlying perceived 'errors" but rather a description of them".

We have adjusted the title to take account of Professor Buist's suggestions.

We are grateful for the constructive criticism, particularly noting the considerable time John Kellet must have taken over his review.

VERSION 2 – REVIEW

REVIEWER	John Kellett Nenagh Hospital Nenagh Ireland
REVIEW RETURNED	21-Mar-2015

GENERAL COMMENTS	This paper continues to be VERY difficult to read and understand. I do appreciate what the authors are trying to express is not easy - but the current version MUST be improved. The author(s) seem to think that everyone else's mental maths are as good as they infer their own to be - well not outside Oxford! I have attached a proposed revision of Table 1. I suggest that the authors consider developing the following definitions that they stick to rigidly throughout the text: True NO Alerts True Alerts - called False Alerts - called True Alerts missed - i.e. not called There are a number of authors to this paper, so I imagine it has been written one of two ways: 1. The main text has been delegated to one author OR 2. The text has been written by committee. If it is the first method (which I suspect) then it is time for the other authors to get round and re-write this paper, circulate it among their lay friends and bang it into some kind of shape that ordinary folk can

	read and understand. If, on the other hand, this paper was written by committee then it is time for ONE author to sit down and re-write it, circulate among their friends etc. This paper simply must be re-written more clearly. It carries and important clinical messages, and it would be too bad if these do not get published because of a failure to master the English language.

REVIEWER	Mark Niegsch Copenhagen University Hospital, Rigshospitlet. Juliane Marie Centeret dept. anesthesia and operations 4014. Copenhagen. Denmark.
REVIEW RETURNED	03-Apr-2015

GENERAL COMMENTS	The manuscript is more clear and concise now. I find that the objective is clear and well adressed in results and discussion. Overall i find i well written and interresting too read.

VERSION 2 – AUTHOR RESPONSE

Reviewer 1: John Kellet.

This paper continues to be VERY difficult to read and understand. I do appreciate what the authors are trying to express is not easy - but the current version MUST be improved. The author(s) seem to think that everyone else's mental maths are as good as they infer their own to be - well not outside Oxford! I have attached a proposed revision of Table 1. I suggest that the authors consider developing the following definitions that they stick to rigidly throughout the text:

True NO Alerts
True Alerts - called
False Alerts - called
True Alerts missed - i.e. not called

There are a number of authors to this paper, so I imagine it has been written one of two ways:

1. The main text has been delegated to one author OR 2. The text has been written by committee.

If it is the first method (which I suspect) then it is time for the other authors to get round and re-write this paper, circulate it among their lay friends and bang it into some kind of shape that ordinary folk can read and understand.

If, on the other hand, this paper was written by committee then it is time for ONE author to sit down and re-write it, circulate among their friends etc.

This paper simply must be re-written more clearly. It carries and important clinical messages, and it would be too bad if these do not get published because of a failure to master the English language.

We have taken John Kellet's advice and Gary Smith has rewritten sections of this paper to improve clarity. We have also had the paper read by colleagues who are not authors, which has also helped

us improve clarity. We have revised our terminology in table 1 in light of John Kellet's comments, though we have used precise terms that differ slightly, in light of feedback from our colleagues.

Reviewer 2: Mark Niegsch

We thank Mark Niegsch for his positive comments about the improved clarity of the paper and hope the further amendments we have made have will have improved it further.

We are grateful for the constructive criticism, particularly noting the considerable time John Kellet must have taken over his review and remain grateful for Michael Buist's positive but useful guidance.