PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Investigating the discriminative value of Early Warning Scores in patients with respiratory disease using a retrospective cohort analysis of admissions to Nottingham University Hospitals Trust over a 2 year period.
AUTHORS	Forster, Sarah; Housley, Gemma; McKeever, Tricia; Shaw, Dominick

VERSION 1 – REVIEW

REVIEWER	Professor Lionel Tarassenko University of Oxford UK
REVIEW RETURNED	07-Nov-2017

GENERAL COMMENTS	The paper does not address the topic described in the title. The NEWS early warning score was designed using a database of vital signs recorded from 100,000 general ward patients in Queen Alexandra Hospital in Portsmouth. This paper should describe how different the vital signs of patients admitted to a tertiary referral centre for respiratory medicine are from those admitted to general wards. The latter should have been the comparator group for the study described in this paper. The authors rightly state in their introduction that a “chronic physiological disturbance caused by COPD may render NEWS less discriminative than in an unselected population; consequently attempts have been made to improve the score in this population.” However, instead of investigating how COPD and other respiratory diseases affect the vital signs of in-hospital patients and hence their EWS score, they use another generic EWS designed locally to compare it with the generic NEWS. The main difference between their local score (the “Notts score”) and NEWS is that they also include urine output as one of the vital signs used to calculate the score. Nowhere in the paper do the authors discuss why this should improve the specificity of their score for patients with respiratory disease. The mortality rate in their study (6%) is much higher than the mortality for the dataset used to design NEWS (1%). The authors do not discuss why the patients in their study had a much higher risk of mortality. As the authors state, both scores perform similarly; however the mandated cut points differ. NEWS is known to have a lower

	specificity and hospitals adjust their cut points and escalation protocols accordingly. The observation that the high sensitivity and low specificity of NEWS means that it acts like a d-dimer is valid, but is not directly related to the title of the paper (the impact of early warning scores on managing patients with respiratory disease). Page 5: “predicating mortality” should obviously be “predicting mortality”.

REVIEWER	Malcolm Brodlie Newcastle University, United Kingdom
REVIEW RETURNED	08-Nov-2017

GENERAL COMMENTS	I enjoyed reading this manuscript. It describes a retrospective analysis of a large dataset from a single tertiary respiratory centre applying 2 different early warning scores (EWS) (a nationally generated NEWS and local EWS) to routinely collected clinical observations data for inpatients. Sensitivity and specificity for mortality within 24 hours along with impact on workload in terms of number of mandated registrar reviews are calculated and compared. The paper is relatively clear and is certainly admirably concise. Overall it provides some interesting observations that would be a useful addition to the literature but is relatively limited in its scope and analysis. Specific points that I would note are: • The title should be more reflective of the actual study design • Clarify in abstract that these were adult patients (and in more detail in the methods section re. characteristics of any particular case mix,e.g. were long term ventilated patients included, etc, etc) • The D dimers analogy does not feel strictly appropriate for an abstract (is ultimately an editorial decision), is justifiable with more description in the discussion in my opinion • I would like to see details of the statistical methods used • Introduction is at the “not really long enough side of concise” in my opinion, there could be more in depth discussion of the relevant research questions and background and challenges in respiratory patients • More information and granularity would strengthen the paper I think, for example o Is data available around ITU or HDU admission rather than just mortality as an outcome? o Is there data on chronic respiratory diagnosis +/- acute problem available, subanalysis along these lines would be interesting o Any data around actual change in management following registrar review? o Workload impact largely focuses on registrars, what about changes in terms of nursing or more junior doctors? • Discussion: Page 5, line 6 … at both extremes of sensitivity comments, is this not stating the obvious? • page 6, line 4. D dimer anecdote needs to be referenced where the analogy is first made • Details could be provided of the relevant trust and R+D approvals at least (I agree that formal ethical approval not required) • Fig 1 is unclear and needs to be redesigned to increase clarity

	• Fig 3 is repeated as B+W and colour, need consistency re. Nottingham EWS versus local EWS terminology • As a general point the discussion could be a little more reflective, this is allowable in relation to this topic and would be of interest in my opinion since the work does touch on workload issues for junior doctors and the cited figure from another study of 47% of critical clinical reviews are generated by nursing staff concern rather than triggering of EWS seems pertinent to me – there is both an art and a science to medicine.

REVIEWER	Dr Ronan O'Driscoll Respiratory Medicine Salford Royal Hospital Salford M6 8HD UK
REVIEW RETURNED	08-Nov-2017

GENERAL COMMENTS	This is an interesting report which describes a retrospective cohort analysis of the vital signs of 8812 respiratory patients in Nottingham. The authors devised a local NEWS score and compared the sensitivity and specificity of this system for predicting mortality with the sensitivity and specificity of the standard NEWS score. They also reported the "callout rate" for both systems. The authors make an important point that the impact on workload for each system can be estimated with precision whereas the effect of track and trigger systems on patient survival is unknown in the absence of large controlled studies. It may be worth mentioning in the paper that a high callout rate might endanger some patients due to "callout fatigue" whilst other apparently stable patients might be harmed by having their care interrupted if their doctors are frequently called away to urgently review patients with elevated NEWS scores. Overall the paper is of interest and clearly written but I have a number of comments and questions. Major comments 1. Was the local NEWS score used on all wards at the author's hospitals or only on the respiratory wards? If it was used on all wards, how did it compare with the standard NEWS score on non-respiratory wards? If it was used only in the respiratory unit, how did hospital staff cope with different scoring systems in different units? 2. How was the local NEWS system devised? I found it difficult to work the key differences from figure 1. The key differences (and the reason for them) should be described in the text. If I read figure 1 correctly, the scores are the same in both systems for respiratory rate, heart rate, blood pressure and temperature but there are differences for the scoring for oxygen use and level of consciousness. The most dramatic difference is that the local NEWS score seems to allocate no points for oxygen saturation but instead adds urine output. The logic behind these choices needs to be discussed and explained. 3. The local scoring system seems to allocate points for oxygen flow rate regardless of the device used. Most patients using nasal cannulae or simple face masks are likely to score zero points (flow ≤9L/min) but the range of 10-14l/min (one local NEWS point) covers a wide range of devices including 31% and 35% Venturi masks which would actually deliver less oxygen than nasal cannulae at 6 l/min or a simple face mask at 5-10 l/min.

	Minor comments 1. Figure 3 appears twice in the pdf version of the file that I reviewed. Strangely, the symbols are of different sizes in the first version but the same size in the second version and therefore the reader cannot tell one from the other. 2. In figure 1, patients with oxygen flow rates of 10-14 l/min are allocated 1 point so I presume that the final box should say ≥15L/min rather than >15 L/min. This is important because the standard flow rate for Reservoir masks is 15L/min. 3. The authors have rightly excluded patients who are receiving end of life care, many hospitals discontinue routine observations in this situation. There is another cohort of patients such as those with advanced lung disease or cancer who are judged to be unsuitable for level 3 or level 3 care but are not requiring end of life care and will often have a "DNA-CPR" order. Some of these patients will have repeatedly high NEWS scores and they will have a high mortality rage but they do not require repeated urgent medical review. Best practice in many hospitals is to modify the scoring system for such patients to avoid "false alarms". Can the local system be modified easily for such patients? 4. Only 54 of 8812 (0.6%) episodes of care were recorded as "End of life care". This seems very low for a respiratory unit.

VERSION 1 – AUTHOR RESPONSE

Response: It has now been clarified in the text that all data was anonymised prior to extraction- Page 4 line 11

Reviewer: 1
Reviewer Name: Professor Lionel Tarassenko
Institution and Country: University of Oxford, UK
Please state any competing interests: None declared

Please leave your comments for the authors below

1. The paper does not address the topic described in the title. The NEWS early warning score was designed using a database of vital signs recorded from 100,000 general ward patients in Queen Alexandra Hospital in Portsmouth. This paper should describe how different the vital signs of patients admitted to a tertiary referral centre for respiratory medicine are from those admitted to general wards. The latter should have been the comparator group for the study described in this paper.

Response: We agree that scores should be looked at in different populations but this piece of work looks at the differing discriminative values of the two scores in a respiratory population and COPD subpopulation rather than looking a two different populations.

2. The authors rightly state in their introduction that a "chronic physiological disturbance caused by COPD may render NEWS less discriminative than in an unselected population; consequently attempts have been made to improve the score in this population." However, instead of investigating how COPD and other respiratory diseases affect the vital signs of in-hospital patients and hence their EWS score, they use another generic EWS designed locally to compare it with the generic NEWS.

Response: The aim was to show that neither system is optimal. The scores have different approaches to managing these patients. One is more sensitive, the other a more pragmatic approach. Our study highlights the difficulties in using such scoring systems in this patient population. Given that respiratory patients constitute a large part of the acute medical take this is an important issue.

3. The main difference between their local score (the “Notts score”) and NEWS is that they also include urine output as one of the vital signs used to calculate the score. Nowhere in the paper do the authors discuss why this should improve the specificity of their score for patients with respiratory disease.

Response: Thank you. The aim of our paper is not to show one scores superiority over another (the authors were not involved in designing the local score) but to highlight the significant difference in outcome with similar types of score. The main difference between NEWS and the local score relates to oxygen delivery scoring and not using saturations. Urine output is included as a general indicator of mortality.

4. The mortality rate in their study (6%) is much higher than the mortality for the dataset used to design NEWS (1%). The authors do not discuss why the patients in their study had a much higher risk of mortality.

Response: Patients with respiratory disease have a higher mortality rate than an unselected admission population, such as the cohort NEWS/ ViEWS was derived from and is typical for a respiratory cohort.

5. As the authors state, both scores perform similarly; however the mandated cut points differ. NEWS is known to have a lower specificity and hospitals adjust their cut points and escalation protocols accordingly. The observation that the high sensitivity and low specificity of NEWS means that it acts like a d-dimer is valid, but is not directly related to the title of the paper (the impact of early warning scores on managing patients with respiratory disease).

Response: In our experience, and following discussion with clinicians in this area, hospitals have very little data and understanding of how changes in cut points affect their manpower and escalation rates, hence the idea behind this paper.

Page 5: “predicating mortality” should obviously be “predicting mortality”.

Response: Many thanks- changed

Reviewer: 2
Reviewer Name: Malcolm Brodlie
Institution and Country: Newcastle University, United Kingdom
Please state any competing interests: None declared

Please leave your comments for the authors below

I enjoyed reading this manuscript.

It describes a retrospective analysis of a large dataset from a single tertiary respiratory centre applying 2 different early warning scores (EWS) (a nationally generated NEWS and local EWS) to routinely collected clinical observations data for inpatients. Sensitivity and specificity for mortality

within 24 hours along with impact on workload in terms of number of mandated registrar reviews are calculated and compared.

The paper is relatively clear and is certainly admirably concise.

Overall it provides some interesting observations that would be a useful addition to the literature but is relatively limited in its scope and analysis.

Specific points that I would note are:

6. • The title should be more reflective of the actual study design

Response: We have reviewed and changed this in line with comments

7. • Clarify in abstract that these were adult patients (and in more detail in the methods section re. characteristics of any particular case mix,e.g. were long term ventilated patients included, etc, etc)

Response: Thank you for pointing this out- it has been clarified on Page 4 lines 5-6.

8. • The D dimers analogy does not feel strictly appropriate for an abstract (is ultimately an editorial decision), is justifiable with more description in the discussion in my opinion-

Response: Thank you- we agree this is a better approach and this has been taken out of abstract

9. • I would like to see details of the statistical methods used.

Response: Increased detail has been added throughout the methods section on page 4

10. • Introduction is at the “not really long enough side of concise” in my opinion, there could be more in depth discussion of the relevant research questions and background and challenges in respiratory patients

Response: We have extended the introduction to add further detail regarding the points raised

11. • More information and granularity would strengthen the paper I think, for example
o Is data available around ITU or HDU admission rather than just mortality as an outcome?
o Is there data on chronic respiratory diagnosis +/- acute problem available, subanalysis along these lines would be interesting- Beyond the scope of this article
o Any data around actual change in management following registrar review?-

Response: These are very valid points but difficult to capture in hospital in large numbers, moreover the data would be confounded by registrar seniority and experience, and would be retrospective.

12. o Workload impact largely focuses on registrars, what about changes in terms of nursing or more junior doctors?-

Response: This has been added, please see table

13. • Discussion: Page 5, line 6 … at both extremes of sensitivity comments, is this not stating the obvious?

Response: One of our internal reviewers thought we should add this line.

14. • page 6, line 4. D dimer anecdote needs to be referenced where the analogy is first made-

Response: We have added a reference to how d-dimer is used.

15. • Details could be provided of the relevant trust and R+D approvals at least (I agree that formal ethical approval not required).

Response: It has now been clarified that information governance approval sought prior to extraction and analysis of data and that all data was anonymised prior to extraction.

16. • Fig 1 is unclear and needs to be redesigned to increase clarity –

Response: This has been done

17. • Fig 3 is repeated as B+W and colour, need consistency re. Nottingham EWS versus local EWS terminology

Response: This has been rectified.

18. • As a general point the discussion could be a little more reflective, this is allowable in relation to this topic and would be of interest in my opinion since the work does touch on workload issues for junior doctors and the cited figure from another study of 47% of critical clinical reviews are generated by nursing staff concern rather than triggering of EWS seems pertinent to me – there is both an art and a science to medicine.

Response: We agree, and the discussion has been changed- page 7 lines 28-31, page 8 lines 7-11 and page 9 lines 2-8 on the marked up version of the manuscript

Reviewer: 3
Reviewer Name: Dr Ronan O'Driscoll
Institution and Country: Respiratory Medicine, Salford Royal Hospital, Salford M6 8HD, UK
Please state any competing interests: None declared

Please leave your comments for the authors below

This is an interesting report which describes a retrospective cohort analysis of the vital signs of 8812 respiratory patients in Nottingham. The authors devised a local NEWS score and compared the sensitivity and specificity of this system for predicting mortality with the sensitivity and specificity of the standard NEWS score. They also reported the “callout rate” for both systems.

The authors make an important point that the impact on workload for each system can be estimated with precision whereas the effect of track and trigger systems on patient survival is unknown in the absence of large controlled studies. It may be worth mentioning in the paper that a high callout rate might endanger some patients due to “callout fatigue” whilst other apparently stable patients might be harmed by having their care interrupted if their doctors are frequently called away to urgently review patients with elevated NEWS scores.

Response: Thank you. This has been changed.

Overall the paper is of interest and clearly written but I have a number of comments and questions.

Major comments
1. Was the local NEWS score used on all wards at the author's hospitals or only on the respiratory wards? If it was used on all wards, how did it compare with the standard NEWS score on non-respiratory wards? If it was used only in the respiratory unit, how did hospital staff cope with different scoring systems in different units?-

Response: Used throughout hospital but data were only available for respiratory patients for this time period

2. How was the local NEWS system devised? I found it difficult to work the key differences from figure 1. The key differences (and the reason for them) should be described in the text. If I read figure 1 correctly, the scores are the same in both systems for respiratory rate, heart rate, blood pressure and temperature but there are differences for the scoring for oxygen use and level of consciousness. The most dramatic difference is that the local NEWS score seems to allocate no points for oxygen saturation but instead adds urine output. The logic behind these choices needs to be discussed and explained.-

Response: Thank you. We have explained the differences in the discussion. The authors did not design the local score. We are aware of individual hospitals tweaking NEWS for their own purposes, but with little data to predict impact on outcome and workload, or on the score's subsequent specificity or sensitivity, hence the reason for this paper, which is primarily to highlight the size and significance of the issue and move towards standardizing scores in a more formal manner.

3. The local scoring system seems to allocate points for oxygen flow rate regardless of the device used. Most patients using nasal cannulae or simple face masks are likely to score zero points (flow ≤9L/min) but the range of 10-14l/min (one local NEWS point) covers a wide range of devices including 31% and 35% Venturi masks which would actually deliver less oxygen than nasal cannulae at 6 l/min or a simple face mask at 5-10 l/min. –

Response: Agreed, this should be examined in the future.

Minor comments
1. Figure 3 appears twice in the pdf version of the file that I reviewed. Strangely, the symbols are of different sizes in the first version but the same size in the second version and therefore the reader cannot tell one from the other.-

Response: Thank you, this will be amended

2. In figure 1, patients with oxygen flow rates of 10-14 l/min are allocated 1 point so I presume that the final box should say ≥15L/min rather than >15 L/min. This is important because the standard flow rate for Reservoir masks is 15L/min.-

Response: Thank you. Will be amended

3. The authors have rightly excluded patients who are receiving end of life care, many hospitals discontinue routine observations in this situation. There is another cohort of patients such as those with advanced lung disease or cancer who are judged to be unsuitable for level 3 or level 3 care but are not requiring end of life care and will often have a "DNA-CPR" order. Some of these patients will have repeatedly high NEWS scores and they will have a high mortality rage but they do not require repeated urgent medical review. Best practice in many hospitals is to modify the scoring system for such patients –

Response: DNACPR is not linked in this data set- but is now being prospectively recorded for future work

VERSION 2 – REVIEW

REVIEWER	Dr Ronan O'Driscoll Dr Ronan O'Driscoll BsC MD FRCP Consultant Physician Respiratory Medicine Level 4 Brooke Building (Orange Area) Salford Royal Foundation NHS Trust Stott Lane Salford M6 8HD Phone 0161 206 5155 Fax 0161 206 4328 ronan.o.driscoll@srft.nhs.uk
REVIEW RETURNED	31-Jan-2018

GENERAL COMMENTS	January 2018 (revised paper) This paper is much edited and much improved. One area that I cannot find addressed was the final two sentences in my minor comment number 3. The authors have discussed the impact of different callout thresholds and they pointed out that a higher threshold will reduce un-necessary callouts of medical staff but will also reduce the sensitivity of the EWS system. The alternative to re-setting the callout threshold (for all patients or for specific disease groups) is to make patient-specific adjustments to the scores, for example adjusting downwards the score for respiratory rate for a patient with chronic lung disease who has a high baseline respiratory rate. The NEWS system advises making such adjustments for patients with chronically abnormal parameters but any such adjustments are not evidence based at present. The paragraph in the discussion section about future studies should mention the option of adjusting specific parameters for individual patients as well as adjustments of the overall system. Controlled trials in this area would be very challenging because of the very large number of options for re-calibration of scores for individual patients. The most important development since this paper was reviewed in November 2017 was the publication of the revised NEWS system (NEWS2) by the Royal College of Physicians in December 2017. The NEWS2 system has introduced a section with modified oximetry scores for patients who are judged to be at risk of hypercapnia. This new section is based on the Salford NEWS system (discussed in Reference 2 by Hodgson et al but applied incorrectly in that study to all patients with COPD, including those with normal CO2 levels). The Nottingham authors have introduced a new table and a new results paragraph related to patients with COPD who constitute the majority of patients at risk of hypercapnia. Unfortunately the main subject of this paper (comparison of local EWS system with “NEWS1”) is now a purely historical comparison because NHS England and Dept of Health have advised the use of NEWS2 in all UK hospitals from now onwards. As a minimum, the discussion section needs to be updated to mention the recent introduction of NEWS2 and to explain that it addresses some of the concerns about patients with chronic

	respiratory disease (although it does not take account of their likelihood to have above average respiratory rate even when clinically stable). It would be unfair to ask the authors to re-run the data compared with NEWS2 but the results of such an analysis would be of great interest. Perhaps the authors could consider it as a future study? Finally, I spotted a minor error on page 5, line 20. The data for COPD patients in in Table 3 (not Figure 3 as stated)

REVIEWER	Malcolm Brodlie Newcastle University, United Kingdom
REVIEW RETURNED	10-Feb-2018

GENERAL COMMENTS	Thank you for taking on board the various reviewer comments. I believe that the title and abstract now more accurately reflect the study and there is greater discussion of the limitations of the work. It is unfortunate that it is not possible to supply any more granularity to the data but this is adequately discussed.

VERSION 2 – AUTHOR RESPONSE

Response:

This has now been changed to:

'Investigating the discriminative value of Early Warning Scores in patients with respiratory disease using a retrospective cohort analysis of admissions to Nottingham University Hospitals Trust over a 2 year period.'

- Please ensure your manuscript fully adheres with the STROBE checklist as we have noticed that some information is lacking, such as a description of the generalisability of the results in the discussion section.

Response:

The article has been reviewed with additions made including an update of the discussion section detailing generalisability of results.

Reviewer Name: Dr Ronan O'Driscoll

Consultant Physician, Respiratory Medicine

Salford Royal Foundation NHS Trust

1- One area that I cannot find addressed was the final two sentences in my minor comment number 3. The authors have discussed the impact of different callout thresholds and they pointed out that a higher threshold will reduce un-necessary callouts of medical staff but will

also reduce the sensitivity of the EWS system. The alternative to re-setting the callout threshold (for all patients or for specific disease groups) is to make patient-specific adjustments to the scores, for example adjusting downwards the score for respiratory rate for a patient with chronic lung disease who has a high baseline respiratory rate. The NEWS system advises making such adjustments for patients with chronically abnormal parameters but any such adjustments are not evidence based at present. The paragraph in the discussion section about future studies should mention the option of adjusting specific parameters for individual patients as well as adjustments of the overall system. Controlled trials in this area would be very challenging because of the very large number of options for re-calibration of scores for individual patients.

Response:

We agree this is an important possible future direction in the development of scoring models and one we have now more fully addressed in the discussion section. Many thanks

2- The most important development since this paper was reviewed in November 2017 was the publication of the revised NEWS system (NEWS2) by the Royal College of Physicians in December 2017. The NEWS2 system has introduced a section with modified oximetry scores for patients who are judged to be at risk of hypercapnia. This new section is based on the Salford NEWS system (discussed in Reference 2 by Hodgson et al but applied incorrectly in that study to all patients with COPD, including those with normal CO2 levels). The Nottingham authors have introduced a new table and a new results paragraph related to patients with COPD who constitute the majority of patients at risk of hypercapnia. Unfortunately the main subject of this paper (comparison of local EWS system with "NEWS1") is now a purely historical comparison because NHS England and Dept of Health have advised the use of NEWS2 in all UK hospitals from now onwards. As a minimum, the discussion section needs to be updated to mention the recent introduction of NEWS2 and to explain that it addresses some of the concerns about patients with chronic respiratory disease (although it does not take account of their likelihood to have above average respiratory rate even when clinically stable). It would be unfair to ask the authors to re-run the data compared with NEWS2 but the results of such an analysis would be of great interest. Perhaps the authors could consider it as a future study?

Response:

Thank you. We have now addressed the release of NEWS2 which is of course an important development in this area. There are two arguments which contribute to the ongoing relevance of this study. First, that as the roll out of NEWS2 will take time due to re-calibration of software platforms and staff training, NEWS will remain the clinical tool deployed by many trusts for monitoring patients. Secondly that the limitations in terms of design and evaluations of EWS highlighted by this study remain valid independent of the scoring system being discussed and should act as a catalyst for designing and funding future studies that are both novel and collaborative.

3- Finally, I spotted a minor error on page 5, line 20. The data for COPD patients in in Table 3 (not Figure 3 as stated)

Response:

Many thanks for spotting this- it has been corrected

Reviewer Name: Malcolm Brodlie

Institution and Country: Newcastle University, United Kingdom

Please leave your comments for the authors below

Thank you for taking on board the various reviewer comments.

I believe that the title and abstract now more accurately reflect the study and there is greater discussion of the limitations of the work.

It is unfortunate that it is not possible to supply any more granularity to the data but this is adequately discussed.