

Dear dr. van Rompaey,

We are hereby submitting the *revised* manuscript entitled “**Validation of the VitalPAC Early Warning Score at the Intermediate Care Unit**”

We would like to thank you for considering our article for publication. We also express our gratitude towards the reviewers, whose comments have been of great value in the improvement of this manuscript. A detailed point-by-point response to their comments is provided in the next pages.

Sincerely,

Joost D.J. Plate, MD and Falco Hietbrink, MD, PhD.

Reviewer 1

C1. I commend the author for addressing this important topic of clinical relevance. There are certain areas which need modification/clarifications. ViEWS IQR needs to be defined as many readers may not be familiar with that.

R1. Thank you for your extensive review and thoughtful comments. Admittedly, we are not quite sure what you mean by defining the ViEWS IQR. The IQR is the interquartile range, which is explained in page 5, paragraph 4 “median and interquartile range (IQR) for non-normally distributed continuous variables”.

The ViEWS is one of the early warning scores and is explained at page 4, paragraph 2: “Because the VitalPAC – Early Warning Score (ViEWS) has the highest discriminative ability compared to 34 other EWS [1], we chose to validate this EWS. The ViEWS is based on the heart frequency, respiratory rate, temperature, systolic blood pressure, oxygen saturation, inspired oxygen (yes/no) and consciousness (see Additional file 1). As a threshold, it often uses an aggregated score of 5 [4].” Additional file 1 contains the items and respective weights which make up the total score.

In response to your comment and to clarify the abbreviation, the ViEWS was also added in the list of abbreviations (additional to the EWS).

C2. Limitations need to be expanded as many institutions do not allow norepinephrine or any vasopressors in their IMU.

R2. We agree that this is important to realize. Therefore, this was added to page 10, 2nd paragraph, 4th sentence: “in particular as our IMCU may be unique in its possibility to administer vasopressors.”

C3. The NNT, PPN and NPV need to be done with all the factors as nursing worries, continuous O₂ and ViEWS and see what NNT we get.

R3. We agree that this indeed is important and therefore this NNT (NNT = 11) is provided in Table 1 and in text at page 10, 3rd paragraph.

C4. A flow diagram may help to go through the process of consideration for transfer may help.

R4. We agree that a flow diagram might be insightful, but we believe that this could be misinterpreted as a flow diagram would indicate that the ViEWS does have its use at the IMCU whereas this article concludes that it has not. However, if the editor wishes, we could add this diagram.

C5. Also as far as the nursing worry score it may be good to see if the worry varies with the nursing years of experience and if the logistic regression shows any different results

R5. This important factor indeed could change the predictive performance of the nurse worry score, as it seems reasonable to assume that more experienced nurses are more correct in assessing whether a patient deteriorates or not.

Reviewer 2

C1. This is an interesting and clinically relevant article exploring the use of EWS in IMCU. the study found that the system is less than satisfactory in IMCU and should be reconsidered for its use in IMCU. generally, this is a well written article. Several concerns are raised after reading the paper. 1.

R1. Thank you for your extensive review and thoughtful comments.

C2. The problem of multiple measurement is not explicitly addressed. for example, a patient can have 10 EWS scores during his stay, but the deterioration occurred at the last warning score, while the preceding 9 scores all indicate a deterioration. this will make the score a very poor predictive performance. hOW did you deal with such kind of data?

R2. We agree that it is indeed important to consider the problem of multiple measurements. We have not corrected for this in our primary analysis, but instead we have performed two sensitivity analyses to assess whether the existence of multiple measurements affected our results.

First, we assessed the performance with only the first measurement, which resulted in an AUROC of 0.72. Second, we used only the measurements which were part of the regular measurement interval (6-10 hours after the previous measurement), which resulted in an AUROC 0.70. As these both had similar discriminative performances, we think it is reasonable to assume that the effect of multiple measurements on this model's performance are negligible.

C3. It appears that a number of patients were hemodynamically unstable, how can they stay in IMCU, rather than ICU? explain this.

R3. We assume that you base this on figure 1, with the separate scores per parameter. Please tell us otherwise, if this is not the case. The presence of hemodynamically unstable patients can be explained by the possibilities of this IMCU (e.g. vasopressor administration), as well as by the nature of the repeated measurements: patients may become hemodynamically unstable during their admission, for which interventions may take place (e.g. start of noradrenalin, high-flow nasal cannula oxygen therapy or ICU transfer)

C4. also suggest to show the calibration of the EWS, since the calibration can reflect the consistency of prediciton and observation over entire range of EWS scores. suggest to use rms package in R.

R4. Thank you for this suggestion. Calibration plots of the model with only EWS and the full model (with all additional parameters) has been added (Figure 3).

C5. patients in IMCU usually have some baseline critical illness, and did you consider try to find subgroups that may present better discrimination for EWS?

R5. As we believe that the EWS should be used to predict deterioration in *all* admitted patients (as it is currently used), we did not try to find subgroups which

may have better discrimination as this does not align with the aim of the EWS. Also, it was not developed for subgroups.

C6. the study showed that building new predictive models by including other variables improves predictive performance. however, this can be due to the fact that the same cohort from the same center was used for both model training and validation. in such a circumstance, the overfitting problem is poorly addressed. furthermore, how did you choose the additional variables? model specification is very important in model building.

R6. We agree, overfitting indeed is a problem in model training and, as such, will also be a problem in the model updating, as is done here.

We have added this sentence (page 13, paragraph 2): “Furthermore, as these additional IMCU-specific clinical parameters are used to fit a model in our own cohort, there is a risk of overfitting and thus the real performance of this updated model is likely even less.”

We chose the variables based on what we thought may clinically be predictive of deterioration in IMCU patients. As IMCU admissions’ vital parameters are susceptible to supportive therapy, we chose noradrenalin and oxygen use as possible predictors. Further, we chose nurse worry as we believe that the clinical interpretation by nurses may be of value. And, we chose the change in ViEWS since changes in vital parameters were hypothesized to be suggestive for (further) deterioration instead of static (cross-sectional) values. We did not select on our data (e.g. via stepwise backwards selection) as this also may lead to overfitting.

C7. some minor comments: 1. suggest to express normal data as mean and standard error; and non-normal data as median and IQR. this is not compulsory.

R7. We changed the reporting of the baseline mean and confidence interval (i.e. age) to mean and standard deviation. These sentences (page 11) have been added:

“The calibration of the ViEWS and the ViEWS with additional parameters (noradrenalin, nurse worry score, oxygen (continuous) and change in ViEWS) is shown in Figure 3. From this, it follows that the ViEWS underpredicts clinical deterioration for predicted probabilities <0.05 , while it overpredicts clinical deterioration for probabilities >0.05 .”

C8. insert a reference for the statistical description (Ann Transl Med. 2016 Mar;4(5):91. doi: 10.21037/atm.2016.02.11.).

R8. Thank you for this suggestion. This has been added [reference 12].