Peer Review File

Article information: https://dx.doi.org/10.21037/atm-22-3303

First External Peer Review

Reviewer A

The author should be commended on their manuscript. They describe the benefit of HRnV in additional HRV measures in predicting serious bacterial infections in young febile infants.

I have a few minor questions/comments:

Introduction: describes the difficulties with identifying serious bacterial infections in infants and the limitations of current methods.

Methods: The authors note that the used and infant friendly prototype to record ECG and HRV parameters were derived in 5 min recordings. Can the authors provide information in regards to:

Comment 1: where measurements made in the same positions (e.g. supine, upright/held) and do the authors believe there is any risk that positioning during measurement may have impacted on HRV

Reply: We performed the ECG recording for recruited infants either while the child was placed supine in the cot bed, or held supine in the caregivers' arms. We were careful to keep the infant comfortable and to minimise movement artefacts, hence we requested for caregivers to hold the infants in the supine position. We believe that by ensuring consistency in infants' position, the HRV and HRnV results would be more easily reproducible.

Changes in the text (Page 8, Lines 194 – 196): Infants were placed either supine in the cot bed or held supine in the caregivers' arms, with our priority being the child's

comfort and to minimise movement artefacts.

Comment 2: How long did the authors record HRV parameters. Did the recording last longer than 5 min and if so were they able to derive any long term HRV parameters. Did infants receive continuous ECG monitoring or just discrete episodes.

Reply: We were not able to derive any long term HRV parameters and did not monitor the HRV continuously. This was a single discrete 5- minute episode, since our goal was to understand if HRV, as an early measure of autonomic dysfunction, was useful to predict the presence of serious bacterial infections (SBIs) in this high-risk population. We recognise that valuable information may be obtained from continuous ECG monitoring and have included this in the limitations.

Changes in the text (Page 13 Lines 326 – 328): "Although we designed the study to use a discrete 5-minute ECG recording for HRV, we recognise that continuous ECG monitoring may yield richer information for diagnostic and prognostic purposes."

Comment 3: Can the authors provide details on the resolution / frequency they recorded ECG at (e.g. 250 Hz)

Reply: We used a sampling frequency of 250Hz, which has been reported to be acceptable for HRV analysis for both time- and frequency-domain parameters. (Kwon O et al. Electrocardiogram Sampling Frequency Range Acceptable for Heart Rate Variability Analysis. Healthc Inform Res. 2018 Jul;24(3):198-206. doi: 10.4258/hir.2018.24.3.198. Epub 2018 Jul 31. PMID: 30109153; PMCID: PMC6085204.) We have clarified this under the Methods section

Changes in the text (Pages 8: Lines 192– 194): "5-minute single lead electrocardiogram (ECG) tracings were obtained using a paediatric-friendly device, with a sampling frequency of 250Hz."

Comment 4: What method did the authors use to identify QRS peaks?

Reply: Our software used the Physionet HRV toolkit (https://archive.physionet.org/tutorials/hrv-toolkit/) where QRS peaks were

automatically identified. (Behar J, Johnson A, Clifford GD et al. A comparison of single channel fetal ECG extraction methods. Ann Biomed Eng. 2014 Jun;42(6):1340-53.) In the infant population, the QRS peaks were correctly and automatically identified in most cases. However, there were a few cases with motion artifacts and abrupt baseline drift. In these instances, our software allowed manual editing to add or delete the QRS peaks so that the HRV parameters could be derived more accurately.

Changes in the text (Page 8, Lines 197 – 202): "Our software used the Physionet HRV toolkit where QRS peaks were automatically identified. (31,32) In the infant population, the QRS peaks were correctly and automatically identified in most cases. However, there were a few cases with motion artifacts and abrupt baseline drift. In these instances, our software allowed manual editing to add or delete the QRS peaks so that the HRV parameters could be derived more accurately.

Comment 5: Was there any influence in the timing of ECG measurement? (e.g. did morning, evening or overnight ECG recordings have any impact on the ECG parameters obtained?

Reply: We did not record the timing of the ECG recording, and acknowledge that there may be diurnal variations that we were not able to account for. We have included this in the Limitations.

Changes in the text (Page 13, Lines 328 – 329): "We did not account for possible diurnal variation in our HRV analyses."

Results

Comment 6: The authors mention that the heart rate in infants with SBIs was significantly greater than those without SBIs, to what extent do the authors believe that they differences in HRV measures can be account for simply due to heart rate and did the authors make any attempt to mitigate for this. (e.g.Sacha J. Interaction between heart rate and heart rate variability. Ann Noninvasive Electrocardiol. 2014 May;19(3):207-16. doi: 10.1111/anec.12148. Epub 2014 Mar 6. PMID: 24602150; PMCID: PMC6932101.)

Reply: As explained in the paper cited by the reviewer, the interactions between HR and HRV are both physiological and mathematical. From a physiological perspective, the elevated HR of patients with SBI may be attributed to the infection and the host response. Thus, for the sake of predicting SBI, the elevated HR or HRV caused by infection would be a good predictor.

For the mathematical part, we conducted collinearity elimination before building logistic models, and HR was not selected into the final logistic models. Therefore, the mathematical interaction from HR has been somewhat mitigated.

We believe that mitigation of HR interaction with HRV would be a great adjustment to measure HRV differences in an intra-patient manner (i.e., to investigate the true differences of HRV measured at different times from the same patient), and will take this into account for future research.

Comment 7: figure 2: the figure reports AUC for models using including more variables. The authors have labelled the best performing AUC curve ALL but the picture does not describe well that ALL refers to vital signs, HRV parameters and blood results) perhaps including within the legend that ALL included bloods would help the reader.

Reply: We thank the Reviewer and have made the change to (revised) Figure 3's Legend

Overall this is a good piece of work that adds well to the increasing literature on HRV and HRnV. The authors should be commended in their thorough reporting of HRV parameters used. With minor changes this will be a good addition to the literature on HRV.

Reviewer B

Overview comments: This is an interesting paper on a potentially important new method of characterizing HRV.

Comment 1: The main deficiency of this paper is that the authors have not shown how the improved performance of the models could potentially result in improvements to care and outcomes among patients.

What would make this paper much stronger is to draw an arc from a quantified problem in existing standard of care, through a hypothesis related to actionability, to a potential improvement in care and/or outcomes.

Reply: We thank the Reviewer for this important advice. In the Introduction, we detailed the implication of over-investigation and liberal use of antibiotics in this young infant population due to fear of missing serious bacterial infections (SBIs). These include causing pain to infants through extensive invasive investigations, anxiety for caregivers, as well as the financial burden of unnecessary procedures and hospitalisations.

Recognising the need to reduce unnecessary testing, researchers have derived algorithms to define a low-risk population who may benefit from less investigations. (citations 11,12) The gap remains, however, that these algorithms are not generalisable to all populations. More importantly, there is no guidance on which febrile infant should receive priority for urgent intervention. We have now added in the introduction that this lack of prioritisation results in infants who are at high risk of SBIs, who truly require urgent investigations and antibiotics, facing delays in time-to antibiotics (Yang J, Ong WJ, Piragasam R, et al. Delays in Time-To-Antibiotics for Young Febrile Infants With Serious Bacterial Infections: A Prospective Single-Center Study. Front Pediatr. 2022;10:873043.)

Changes in the text (Page 4, Lines 111-112) : "These invasive procedures are also painful for the infants and cause caregivers additional anxiety. (10) ..... **This lack of prioritisation causes delays in time-to-antibiotics for febrile infants who are at high risk for SBIs. (14)** Researchers continue to pursue predictive models for SBIs

using clinical and biochemical predictors. (14–16)"

Please refer to the next reply for the changes made to the Aims statement and to the Discussion.

Comment 2:

1) Lines 132-135: the authors report that patients <28 days old "routinely receive broad-spectrum antibiotics". And that among patients 28-90 days old, the "majority of these receive antibiotics until culture results are known".

2) Lines 111-113: the authors hypothesize "that the addition of HRV and HRnV measures to existing triage tools will enhance the ability of ED providers to identify SBIs early in the course of disease". Arguably, ED triage is not early in the course of the disease. In any case, the hypothesis of identifying SBIs points toward a test of improved sensitivity, yet the authors have chosen ROC as the primary means of characterizing their model. ROC is an important and useful means of characterizing model performance. Yet clinical implementation often revolves around one or more action thresholds.

Reply: With regards to ED triage not necessarily representing an early phase of disease, we have made the change accordingly.

Changes in the text (Page 5, Lines 125-127): "We hypothesized that the addition of HRV and HRnV measures to existing triage tools will enhance the ability of ED providers to identify SBIs early in the course of the patient's journey."

We agree that clinical implementation revolves around an actionable threshold, for which we should report improved sensitivity. The goal of this paper is to assess the potential value-add of HRnV to existing tools. We are currently not ready to recommend actionable thresholds. We agree with the Reviewer and are actively working on a larger cohort of infants where HRV and HRnV parameters used to validate our model can subsequently provide a risk score. With a risk score, we can then recommend thresholds of action and report on sensitivity, specificity, positive and negative likelihood ratios. These steps are necessary before we can move to clinical implementation.

We have revised the Aims statement so that it is not an over-reach of what we can report

in this manuscript.

Changes in the text (Page 5, Lines 124 – 125): "We therefore sought to explore the potential of HRV and HRnV measures to predict for SBIs in young febrile infants."

We have also made extensive changes to the Discussion (Page 12, Lines 296 – 304): "HRnV holds promise in the young infant age group due to the non-invasive nature of this technology.(18) Importantly, we demonstrated that HRnV provides additional information at triage, early in the infants' ED journey. Future studies should focus on deriving actionable thresholds using HRV and HRnV, together with other clinical predictors. Recommendations for change in clinical practice can only be made after studying the sensitivity and specificity of these thresholds. HRnV has the potential to reduce recognition delays and contribute to prioritisation of at-risk infants for early antibiotics.(14) In implementing an actionable algorithm, we can then study if the time-to-antibiotics is indeed improved for infants at high risk of SBIs."

Comment 3: The qualitative labels (routine and majority) lead to skepticism that a successful hypothesis (enhanced ability to identify SBIs at triage) would lead to improvements in care. If all or most patients are being treated with ABX, it is hard to see how identifying more SBIs changes anything.

If the authors believe that a significant fraction of patients are not receiving ABX despite SBIs, then they should quantify this deficiency in standard of care, and structure their hypothesis around improving this problem (one of many potential examples: that HRV/HRnV improve sensitivity at a constant alarm rate).

If, on the other hand, the authors believe that inclusion of HRV/HRnV into models predicting SBI may improve the timing of ABX administration among patients with SBI, then this has not been made clear throughout the paper.

Finally, if the authors believe that better discrimination could lead to reduction in ABX administered to patients with no SBI, then that should be made clear and a hypothesis generated around that concept.

There is the basis of a very strong paper here. But the authors need to think about clinical actionability and either restructure, or better explain, how an improved model

for assessing risk of SBI could lead to better outcomes for patients.

Reply: We thank the Reviewer for the feedback and have appraised the gap, our goal, hypothesis, approach and next steps. The clinical problem is that febrile infants are over-investigated, resulting in unnecessary testing and over-use of empirical antibiotics. While current published algorithms (11,12) seek to delineate a low-risk population, a large number of infants still undergo extensive investigations, resulting in delays to time-to-antibiotics for those febrile infants who are truly at high risk of SBIs. This gives rise to the need to derive predictive models for febrile infants, early in the course of the patient journey.

Changes in the text:

(Page 4, Lines 111 – 113): "This lack of prioritisation causes delays in time-to-antibiotics for febrile infants who are at high risk for SBIs. (14) Researchers continue to pursue predictive models for SBIs using clinical and biochemical predictors. (15–17)"

(Page 5, Lines 124 – 127): "We therefore sought to explore the potential of HRV and HRnV measures to predict for SBIs in young febrile infants. We hypothesized that the addition of HRV and HRnV measures to existing triage tools will enhance the ability of ED providers to identify SBIs early in the course of the patient's journey."

(Page 12, Lines 296 – 304): "HRnV holds promise in the young infant age group due to the non-invasive nature of this technology.(18) Importantly, we demonstrated that HRnV provides additional information at triage, early in the infants' ED journey. Future studies should focus on deriving actionable thresholds using HRV and HRnV, together with other clinical predictors. Recommendations for change in clinical practice can only be made after studying the sensitivity and specificity of these thresholds. HRnV has the potential to reduce recognition delays and contribute to prioritisation of at-risk infants for early antibiotics.(29) In implementing an actionable algorithm, we can then study if the time-to-antibiotics is indeed improved for infants at high risk of SBIs."

(Page 13, Lines 321 – 323): "We also recognise that the derivation and validation of clinical prediction models may potentially result in cases of missed SBIs, and the safety profiles of these should be evaluated carefully before recommendations for clinical

implementation.(36,37)”

Specific critiques:

Comment 4: Line 43: Tables 1, 2, and 3 were not included in the pdf and were not reviewed.

Reply: The tables were attached in the online manuscript submission system. We now attach them in the manuscript as well, for easy review.

Comment 5: Line 63-64: Suggest changing to: 74/312 infants (23.7%) had SBIs with the most common being UTIs (66/72, 91.7%).

Changes in text (Page 3, Lines 73-74): “74/312 infants (23.7%) had SBIs with the most common being UTIs (66/72, 91.7%). 2 infants had co-infections.”

Corresponding changes in text (Page 10, Lines 244 – 246): “74/312 infants (23.7%) had SBIs with the most common being UTIs (66/72, 91.7%) and bacteraemia (6/72, 8.3%). Among the remaining 2 infants, 1 infant had meningitis and UTI, and another infant had bacteraemia and UTI.”

Comment 6: Line 85: Suggest adding “including meningitis, bloodstream infection, and urinary infection” where SBIs are defined.

Changes in text (Page 4, Lines 96 – 98): “Serious bacterial infections (SBIs) (including meningitis, bloodstream infections and urinary tract infections) in young febrile infants (< 90 days old) pose a diagnostic dilemma to emergency department (ED) physicians and paediatricians.(1–3)”

Comment 7: Line 92: Suggest deleting “cumulatively to” and insert “over seven years” after dollars (if I am correctly interpreting the intended meaning).

Changes in text (Page 4, Lines 102 – 104): “In the United States of America (USA), the cost of hospitalisation and discharge after no more than 3 days of antibiotics for these febrile infants added to more than USD 76 million dollars over seven years.(9)”

Comment 8: Line 95: Suggest replacing "deserves" with "might benefit from"

Changes in text (Page 4, Lines 106 – 108): "Recognising the need to reduce unnecessary testing, researchers have derived algorithms to define a low-risk population that might benefit from a less aggressive approach.(11,12)"

Comment 9: Line 102: Suggest inserting "among febrile infants" after SBIs.

Changes in text (Page 4, Lines 114 – 115): "We previously demonstrated that heart rate variability (HRV) adds to triage performance in predicting for SBIs among febrile infants.(18)"

Comment 10: Line 107-109: Another sentence is would be helpful here to describe what HRnV is, and how it augments… and enhances….

Changes in text (Page 5, Lines 119 - 123): "Heart rate n-variability (HRnV), constructs new signals based on the R-to-R peak intervals (RRI) used in the conventional HRV analysis. It was more recently invented as a novel tool to augment the number of calculated parameters from the same segment of signals, with the potential to enhance the prognostic information provided by traditional HRV parameters.(26,27)"

Comment 11: Line 155: better wording/description of "contention with such diagnoses at this young age" would be helpful.

Changes in text (Page 7, Lines 172 – 175): "We chose to align our definition with the updated literature (12) and did not include pneumonia or lower respiratory tract infections stated in our original protocol (NCT04103151) because there were concerns over inconsistencies in final diagnoses when their records were reviewed."

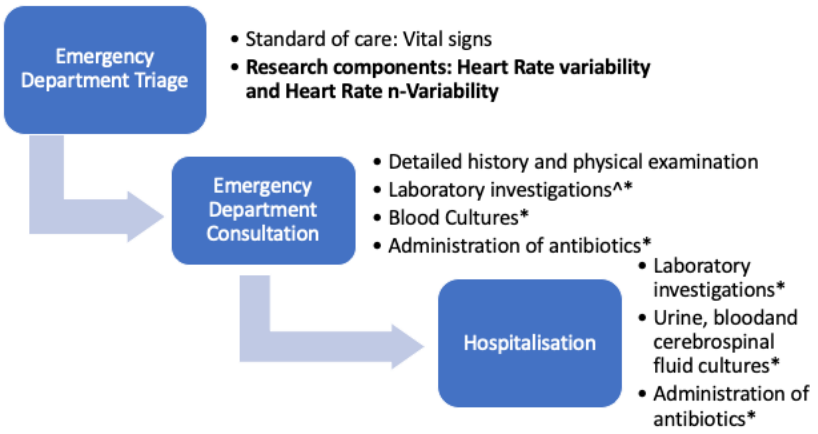
Comment 12: Line 170+: a better description of the temporal relationship between triage, abx administration, and ECG collection would be helpful. And perhaps availability of CBC/CRP (see the comment below regarding lines 242-243). Perhaps a diagram or flowchart walking the reader through the diagnostic cascade, including the

point at which the improved model might influence decision making in a positive way.

Reply: In the revised manuscript, we have detailed the diagnostic cascade for febrile infants in our ED (Figure 1).

Changes in text (Page 6-7, Lines 163 – 169): "Laboratory investigations such as the total white blood cell count, the absolute neutrophil count (ANC), haemoglobin, platelets, C-reactive protein (CRP) and procalcitonin were recorded. If the infants are stable, these investigations are carried out at the ward after hospitalisation. If they are considered to be high risk for SBIs (by vital signs or clinician assessment), the investigations and administration of antibiotics are administered expeditiously in the ED. The diagnostic cascade is detailed in Figure 1."

Figure 1. Diagnostic cascade for febrile infants in the Emergency Department



^Laboratory investigations include total white blood cell count, the absolute neutrophil count (ANC), haemoglobin, platelets, C-reactive protein (CRP) and procalcitonin
*in the event that an infant is considered high risk for serious bacterial infection (SBI) at triage, the above are performed expediently in the Emergency Department. Otherwise, the investigations, cultures and administration of antibiotics are carried out after hospitalisation.

Comment 13: Line 190-191: the impact of excluding >5% of patients due to movement artefact should be addressed in the Results (characteristics and outcomes) and Discussion.

Reply: We have included the information in the Results and the Discussion

Changes in the text (Pages 10, Lines 239 – 242): "We recruited a total of 330 febrile

infants, among whom 18 (5.5%) patients did not complete the study because movement artefacts rendered the HRV and HRnV parameters unsuitable for analysis (Figure 2). Among the 18 excluded infants, the median age was 11.0 days (IQR 3.0 - 50.5 days) and 4 (22.2%) had SBIs, all of which were UTIs. Among 312 infants analysed…"

Changes in the text (Page 13, Lines 323 – 326): "We excluded 18 infants (5.5%) because their HRV could not be analysed. However, the SBI rate was comparable between the excluded infants and the analysed study population (22.2% versus 23.7%)."

Comment 14: Line 200: please expand on why/how variables were entered based on clinical discretion

Changes in the text (Page 9, Lines 227 – 228): "The variables were determined after reviewing known predictors in the literature, as well as availability of these data at ED triage and then at consultation."

Comment 15: Line 203: how were the 95% CIs calculated? Also, the authors can greatly strengthen the validity of the result by using techniques such as k-fold cross-validation with repeats to mitigate overfitting and provide more robust estimates of model performance; otherwise, all results should be reported as "apparent" performance.

Reply: The 95% CIs were calculated using the confidence intervals from Wald tests of logistic regression coefficients. For each of the CIs of unadjusted odds ratios, a single-variable logistic regression model was fitted for each of the variables. The Wald confidence interval was then calculated before being converted to the CI of the corresponding odds ratio. For adjusted odds ratios, the Wald confidence intervals were calculated for all coefficients of the final multivariate logistic regression.

The reason we did not implement k-fold cross-validation is that our goal for this paper is to investigate whether HRnV metrics would be beneficial to SBI prediction. The current sample size would also be a limitation for the development and validation of a robust triage model. However, in future work and with a greater sample size, we would like to pursuit a HRnV-based triage model.

Comment 16: Lines 209-210: TRIPOD is more appropriate

Reply: We have replaced the STARD checklist with the TRIPOD Checklist.

Changes in the text (Page 5, Lines 128): "We reported our findings according to the TRIPOD Checklist for Prediction Model Development.(28)"

Citation: Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. Ann Intern Med. 2015;162(1):55-63

Comment 17: Line 239: The text indicates components of the model were age, sex, day of fever, and respiratory rate. Yet the referenced figure (Figure 2) uses the label "Vitals" for this model.

Reply: We have changed the (revised) Figure 3 Legend accordingly.

Comment 18: Lines 242-243: if the authors are going to make this statement, then they should compare it with a model with demographics/vitals + labs versus demographics/vitals + labs + HRV/HRnV. But even so, is the statement relevant to the hypothesis of improving discrimination at the time of triage, as there is some time lag between triage and availability of CBC/CRP. Perhaps that is a secondary hypothesis that could be more fully described.

Reply: We thank the Reviewer for this feedback. We have added the new Figure 1 (Diagnostic cascade of febrile infants in the Emergency Department) to help the readers to understand that HRV and HRnV are meant to be available at the ED triage and applied before the return of blood investigations. This is why the ROCs are built incrementally, based on best available information.

Under Methods (Page 9, Lines 232-236): "We first assessed the AUC of vital signs and clinical assessment in the prediction of SBIs, then added on HRV, HRnV and laboratory investigations, incrementally. We chose to do so because HRV and HRnV are non-invasive biomarkers and have potential at triage to provide early discrimination, before

subsequent laboratory investigations (which require turnaround time) yield results."

Comment 19: Lines 249-250: same as above.

Reply: Thank you, please refer above to the major changes made**.**

Comment 20: Lines 264-265: It seems important here to quantify the difference in performance between HRV and HRnV.

Reply: We have added in the information following this feedback.

Changes in the text: (Page 12, Lines 292 – 295): "We demonstrated that HRnV, in addition to HRV, adds value for risk stratification among young infants at risk of SBIs. In our study, the ROC improved from 0.776 (95%CI 0.718 - 0.835) to 0.805 (95%CI 0.750 – 0.860) after HRnV was added to both HRV and existing triage information."

Comment 21: Lines 268-269: This needs to be expanded upon greatly per the overview comments above.

Changes in the text (Page 12, Lines 298 – 304): "Future studies should focus on deriving actionable thresholds using HRV and HRnV, together with other clinical predictors. Recommendations for change in clinical practice can only be made after studying the sensitivity and specificity of these thresholds. HRnV has the potential to reduce recognition delays and contribute to prioritisation of at-risk infants for early antibiotics.(29) In implementing an actionable algorithm, we can then study if the time-to-antibiotics is indeed improved for infants at high risk of SBIs."

Comment 22: Lines 272-274: Unclear.

Changes in the text (Page 12, Lines 304 – 308): "Future prospective research could focus on the development of an automated triage algorithm similar to that derived for chest pain in adults, providing real time risk stratification for febrile infants at risk of SBIs.(26,33) An effective risk stratification approach, when validated, may address the current burdens of diagnostic inefficiencies.(34,35)"

Comment 23: Lines 284-285: Clarify. Is this alluding to the fact that 28-90 day patients only had CSF obtained based on the discretion of the medical team, or something else?

Changes in the text (Page 13, Lines 318 – 319): "Some infants did not have a complete septic workup (including a lumbar puncture) performed, therefore allowing a theoretical risk of missed SBI."

Comment 24: Line 338: Figure 1, the rightmost box has some text cropped

Changes in (revised) Figure 2 – We have reformatted the Figure accordingly.

Comment 25: Line 343: Figure 2, the legend and the caption do not fully/accurately describe the models (see comment regarding line 239), both as it relates to the use of "Vital Sig

Changes in (revised) Figure 3 – We have relabeled the ROC curves to describe the variables more accurately.

The legend reads

Triage information

Triage information and HRV

Triage information, HRV and HRnV

Triage information, HRV, HRnV and blood test results

*Triage information refers to age, sex, day of fever, and respiratory rate

Reviewer C

Nicely done, well organized, well written paper. This study evaluates the addition of heart rate variability and n-variability in discriminating between febrile infants < 90 days old who have serious bacterial infection vs not.

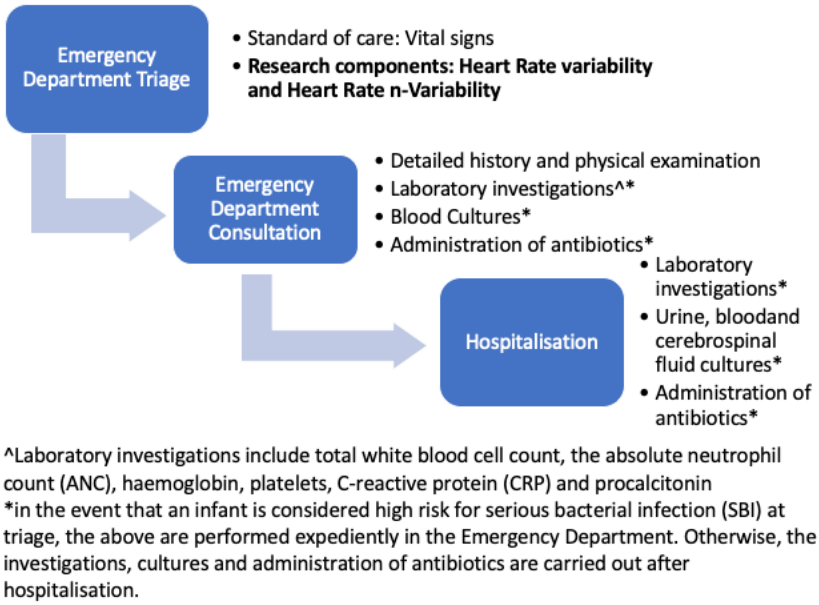
Comment 1: It appears that the addition of these variables increases the AUC. However, the most significant increase comes from adding labs. The utility of adding this simple and non-invasive test should be weighed against the cost and the impact. Will these infants get labs anyway? If so, why add this test? Important end points in follow up studies will be whether this reduces overall cost, reduces unnecessary labs obtained, or reduces unnecessary antibiotics administered.

Reply: We have made major changes in response to both Reviewer B and C.

There is now a new Figure 1, that helps readers to understand the Diagnostic Cascade. Febrile infants need an early risk stratification approach at triage that can prioritise those who require early investigations and antibiotic administration. This is because laboratory tests require turnaround time. An effective risk stratification at ED triage will help to ensure that infants at high risk of SBIs will be prioritized, therefore reducing delays in time-to-antibiotics.

Under Methods (Page 9, Lines 232-236): "We first assessed the AUC of vital signs and clinical assessment in the prediction of SBIs, then added on HRV, HRnV and laboratory investigations, incrementally. We chose to do so because HRV and HRnV are non-invasive biomarkers and have potential at triage to provide early discrimination, before subsequent laboratory investigations (which require turnaround time) yield results."

Figure 1. Diagnostic cascade for febrile infants in the Emergency Department



^Laboratory investigations include total white blood cell count, the absolute neutrophil count (ANC), haemoglobin, platelets, C-reactive protein (CRP) and procalcitonin
*in the event that an infant is considered high risk for serious bacterial infection (SBI) at triage, the above are performed expediently in the Emergency Department. Otherwise, the investigations, cultures and administration of antibiotics are carried out after hospitalisation.

Comment 2: How applicable is this technology? Will any center be able to modify the software prototype? How was the modification achieved – was it with the help of a software engineer? How can this be replicated elsewhere if not described in the methods? (Perhaps you may state that this was previously described in a paper).

Reply: The HRV prototype was developed and modified by an engineer. We have provided more details in Methods to detail that our software was built upon the Physionet HRV toolkit. The device for measuring 5-min ECG in this study is currently under regulatory application in Singapore. In fact, our HRV/HRnV calculation software is agnostic to the ECG hardware and can accept ECG signals in text format, excel format and Kubios format as input. The research team is planning to release and share this software on Github via a GNU General Public License so that it can be used worldwide.

Changes in the text (Page 8, Lines 197 – 202): "Our software used the Physionet HRV toolkit where QRS peaks were automatically identified. (31,32) In the infant population, the QRS peaks were correctly and automatically identified in most cases. However, there were a few cases with motion artifacts and abrupt baseline drift. In these instances, our software allowed manual editing to add or delete the QRS peaks so that the HRV parameters could be derived more accurately.

Comment 3: Did any of the patients have co-infections? For example, it is common for some

young infants to have RSV and a concurrent UTI. Was this evaluated in the study population?

Reply: In this paper, we focused the outcome variable collection on SBIs only. We agree with the Reviewer that there are common co-infections with viruses (specifically RSV and UTIs have indeed been reported in the literature). We will take this into account for future studies.

Comment 4: Do we understand enough about HRV to say that children with abnormal HRV have autonomic nervous system dysregulation? Perhaps in adults or critically ill infants this is true, but in this study the vast majority were not critically ill. Infants in this study with abnormal HRV and HRnV may have an appropriate response to infection with increased variability, and this may not mean autonomic nervous system dysregulation. If it can be stated simply, could you qualitatively state how variability I different in SBI vs no SBI in the results? For example, it appears that the group with SBI has lower variability than the no SBI group (from Supplementary Table 1).

Reply: We agree with the reviewer that in our study population, infants with SBIs were hemodynamically stable. As part of ethical recruitment of patients, those who required urgent resuscitation were not included in this study. We did however, find that both HRV and HRnV indices (see both Table 3 and Supplementary Table 1) had lower variability among infants with SBIs compared to those without. We have provided clarifications in the text.

Changes in the text (Page 10, Lines 257 – 258): "Infants with SBIs had lower variability that those in the non SBI group."

Comment 5: In Figure 1, the text is cut off in one of the boxes. In supplementary Table 2, the numbers don't align with the rows in the first column.

Changes in (revised) Figure 2 – We have reformatted the Figure accordingly.

Changes in (revised) Supplementary Table 2 – We have reformatted the Supplementary Table accordingly.

Second External Peer Review

Overview comments:

The authors' revisions have significantly improved the manuscript.

Reply: Thank you. We have responded to the feedback below. Line numbers refer to the tracked manuscript with Simple Markup.

Comment 1: Line 84: Suggest inserting "at ED triage" after "…among febrile infants"

Reply 1: We have made the relevant change.

Changes in the text (Abstract, Page 3 Line 83-84) "Addition of HRV and HRnV to current assessment tools improved the prediction of SBIs among febrile infants at ED triage."

Comment 2: Lines 140-142: Suggest modifying the hypothesis to better align with the Methods/Results as follows: "We hypothesized that the addition of HRV and HRnV measures to existing triage tools will enhance the discriminatory ability of models to identify SBIs at the time of ED triage."

Reply 2: We have made the relevant change.

Changes in the text (Page 5, Lines 127 – 129) "We hypothesized that the addition of HRV and HRnV measures to existing triage tools will enhance the discriminative ability of models to identify SBIs at the time of ED triage.

Comment 3: Line 218: Suggest inserting "analysis" after "HRV".

Reply 3: We have made the relevant change.

Changes in the text (Page 8, Lines 205-206): "This enabled us to perform and complete the HRV analysis on most of these young infants."

Comment 4: Lines 318-320: Suggest clarifying on the timing of the availability of lab markers as follows: "By adding laboratory results that became available after ED consultation (absolute neutrophil count, haemoglobin, and C-reactive protein), the performance of the model improved…"

Reply 4: We have made the corresponding change.

Changes in the text (Page 11, Lines 270 – 273): "By adding laboratory results that became available after ED consultation (absolute neutrophil count, haemoglobin, and C-reactive

protein), the performance of the model improved to 0.875 (95%CI 0.828 – 0.921). (Figure 3)"

Comment 5: Lines 347-359: I suggest re-ordering the sentences for clarity, as follows (note some wording changes as well):

"HRnV holds promise in the young infant age group due to the non-invasive nature of this technology.(18) Importantly, we demonstrated that HRnV provides additional information at triage, early in the infants' ED journey. HRnV has the potential to reduce recognition delays and contribute to prioritisation of at-risk infants for early antibiotics.(14) Future studies should focus on deriving actionable thresholds using HRV and HRnV, together with other clinical predictors, similar to an automated triage algorithm derived for chest pain in adults, providing real time risk stratification for febrile infants at risk of SBIs.(26,33) Recommendations for change in clinical practice can only be made after studying the performance metrics of these thresholds. In implementing an actionable algorithm, we can then study if the time-to-antibiotics is indeed improved for infants at high risk of SBIs. An effective risk stratification approach, once validated, may address the current burdens of diagnostic inefficiencies.(34,35)"

Reply 5: The changes have been made accordingly:

Changes in the text (Page 12, Lines 298 – 309): "HRnV holds promise in the young infant age group due to the non-invasive nature of this technology.(18) Importantly, we demonstrated that HRnV provides additional information at triage, early in the infants' ED journey. HRnV has the potential to reduce recognition delays and contribute to prioritisation of at-risk infants for early antibiotics.(14) Future studies should focus on deriving actionable thresholds using HRV and HRnV, together with other clinical predictors, similar to an automated triage algorithm derived for chest pain in adults, providing real time risk stratification for febrile infants at risk of SBIs.(26,33). Recommendations for change in clinical practice can only be made after studying the performance metrics of these thresholds. In implementing an actionable algorithm, we can then study if the time-to-antibiotics is indeed improved for infants at high risk of SBIs. An effective risk stratification approach, once validated, may address the current burdens of diagnostic inefficiencies.(34,35)"

Comment 6: Lines 406-407: I suggest aligning the first sentence of the conclusion better with the Results as follows, "We found that the addition of HRnV and HRV to demographics and vital signs improved model performance at ED triage. Further research…"

Reply 6: The changes have been made.

Changes in the text (Page 13, Line 339-340): "We found that the addition of HRnV and HRV

to demographics and vital signs improved model performance at ED triage."

Comment 7: Line 502: Suggest inserting "demographics" in front of "vital signs". I also suggest expanding the table to include all four models by adding additional columns, and with NAs in the cells where the variable was not used.

Reply 7: We have made the change to the title for Table 3. We thank the Reviewer for the comment on expanding Table 3 (multivariable logistic regression) to include all 4 models. After careful consideration, we are of the opinion that the unadjusted adds ratios (for all variables) and adjusted odds ratios (for the final model) provide the clearest representation, while the performance of the other models are represented in Figure 3. This is because there are already a rather large number of variables, with their corresponding point estimates (and 95% CIs) in Table 3.

Changes made to the text (Page 18): "Table 3. Multivariable logistic regression for demographics, vital signs, blood investigations, heart rate variability (HRV) and heart rate n-variability (HRnV)."