Peer Review File

Article information: https://dx.doi.org/10.21037/atm-23-1896

Reviewer A

The manuscript "A two-phased study on the use of remote photoplethysmography (rPPG) in paediatric care" by Ahmad Hatib et al. is devoted to investigating the feasibility, acceptability, and accuracy of using rPPG on pediatric patients.
Findings on feasibility, acceptability, and unacceptable performance in children younger than 10 years old are quite interesting.
While the article is interesting in general, it has a significant design flaw, which can potentially nullify their findings.

Comment 1. The use of unknown equipment provided by Nervotec Pte. Ltd represents the major design flaw. For example, there is no information on the clinical performance of this equipment (e.g., in the adult population). While HR extraction from rPPG seems quite feasible, RR and SO2 are much trickier and not out of the box technology, especially in darker skin tones. Without information on validation/verification of this equipment (e.g., peer-reviewed publications) and/or proper information about the device/algorithm, the "black box" approach use is not justified.

Reply 1. Thank you for highlighting this concern. Nervotec Pte Ltd is currently in the process of conducting clinical validation studies in the Singapore healthcare ecosystem to assess the performance of this technology. Their rPPG technology was developed based on well-validated methods described in existing literature (references have been added to the amended text – see below) and is Nervotec's proprietary rPPG software for contactless vital signs monitoring. The device used in our research study was registered as a Clinical Research Material (Notification number CRM2200314) with Singapore's Health Sciences Authority.

This software uses computer vision techniques to locate faces within incoming video frames, identify Regions of Interest (ROIs) then take a spatial and temporal average of ROI pixels in each frame to corresponding RGB (Red, Green, Blue) values. RGB signals are refined and undergo processing to produce a Blood Volume Pulse (BVP) signal. A further module then interprets the BVP signal to estimate vital signs. In the text (see below) we added explanations on how each vital sign (HR, RR and SpO2) is derived.

We acknowledge the challenges associated with precise estimation of RR and SpO2, especially with diverse skin tones and demographic variations. Within the phase of RGB Extraction and Filtering, a component critical to the technology's precision is the implementation of the Bandpass filter. This filtering technique plays a central role in isolating specific frequency

components essential for RR and SpO2 estimation. This information has also been added into the text.

Changes in the text 1. We have modified the text to add in further details on the rPPG technology used in our study as follows (See Page 5, Lines 132-161):

This rPPG technology was developed based on well-validated methods described in existing literature (5,16,17), and is Nervotec's proprietary rPPG software for contactless vital signs monitoring. This software uses computer vision techniques to first identify and locate faces within incoming video frames. Regions of Interest (ROIs) are then identified within the facial area. Facial data is further refined, eliminating non-informative regions such as the hair and eyes. Once ROIs are defined, the software takes a spatial and temporal average of ROI pixels in each frame to corresponding Red, Green, Blue (RGB) values. Signal pre-processing techniques are then applied to remove noise, and further refine the raw signal. Following the tracking of color changes within the video stream over a specific duration, the filtered RGB signals undergo processing to produce Blood Volume Pulse (BVP) signals. HR, RR, and SpO2 values are calculated by analysing BVP signals in either a time or frequency domain by using established mathematical formulae and peak detection techniques (17).

In our study, HR was primarily extracted by analysing the maximum frequency peaks in the BVP signal corresponding to heartbeats (the Power Spectral Density (PSD) of the selected BVP signal). The following equation was used to calculate HR: $HR = 60 \times f$ (highest peak in the frequency spectrum). RR was derived from the PSD of the BVP signal. Band pass filtering was applied to each component with the cut-off frequencies in the normal human breathing range. The component with the strongest peak would be the best candidate for the respiration signal. RR was calculated by identifying the peaks in the resultant signal and converting the frequency to breaths per minute. The equation used was: $RR = 60 \times f$ (peak frequency within the appropriate range). SpO2 was determined by analyzing the ratio of the alternating current (AC) and direct current (DC) components of signals from the Red and Blue colour channels in the RGB images. The formula used for SpO2 extraction was as follows: $SpO2 = (\alpha - \beta) \times (AC_{red}/DC_{red} \div AC_{blue}/DC_{blue})$, where α, β are mathematical constants with values 1 and 0.02 respectively. This ratio was then converted into percentage SpO2. The implementation of Band pass filtering allowed for more precise estimation of RR and SpO2, especially within diverse skin tones and demographic variations.

The experimental setup comprised of a Logitech C920 High Definition Pro Webcam connected to a standard laptop which ran the rPPG application. This device was registered as a Clinical Research Material (Notification number

CRM2200314) with Singapore's Health Sciences Authority.

Minor shortcomings:
Comment 2. Lines 98-100: In the sentence "These data can potentially be used to extrapolate parameters such as heart rate (HR), heart rate variability (HRV), respiratory rate (RR), oxygen saturation (SpO2) and blood pressure (BP)" references are required.

Reply 2. The following references have been added for this statement:

8. Capraro GA, Balmaekers B, den Brinker AC, Rocque M, DePina Y, Schiavo MW, Brennan K, Kobayashi L. Contactless vital signs acquisition using video photoplethysmography, motion analysis and passive infrared thermography devices during emergency department walk-in triage in pandemic conditions. The Journal of emergency medicine. 2022 Jul 1;63(1):115-29.

9. Lee H, Ko H, Chung H, Nam Y, Hong S, Lee J. Real-time realizable mobile imaging photoplethysmography. Scientific Reports. 2022 May 3;12(1):7141.

Changes in the text 2. The new references have been added to the text (See Page 4, Line 102) and the References section.

Reviewer B
Overall an interesting and important topic looking into rPPG which if accurate in children would allow for wireless monitoring.

Methods: why were children between neonates to 6 years of age not included? What age cutoff are you considering a neonate?

Reply 1. We apologise that there was a slight error in the original manuscript. The youngest age for the paediatric age group was 5.2 years, thus the age range for the paediatric age group in Phase 1 should be 5 to 16 years. We have corrected this error in the manuscript.

For the pilot study (Phase 1), our patients were divided into 2 groups – neonatal (patients from the Special Care Nursery with age cutoff of ≤ 28 days old) and paediatric (all other patients in the general wards). The pilot study was first performed in the Special Care Nursery, before moving on to the paediatric general wards. We completed the recruitment of neonatal patients first. The reason children between 29 days and 5 years were not included in the rest of the pilot study (Phase 1) was due to very discrepant results obtained from the neonatal patients compared to the ground truth. Test runs on 2 children aged above 28 days and below age 5 years were also significantly

discrepant, just by observation. Thus, as we had already collected feasibility data from the neonates, and there seemed to be minimal benefit in continuing the device on younger children below 5 years, we decided to stop recruiting children between 29 days to 5 years, and to focus on children 5 years and above for the rest of the pilot study (Phase 1).

Changes in the text 1. We have modified the text to clarify the patient age groups as follows:

See Page 2, Line 66. Ten neonates and 28 children aged 5-16 years were recruited for Phase 1 (765 datapoints).

See Page 2, Lines 75-77. Our study showed that rPPG is acceptable and feasible for neonates and children aged 5-16 years, and HR values in children aged 12 to 16 years correlated well with the current standard.

See Highlight Box. Our study is the first to assess the feasibility and accuracy of rPPG in the paediatric population across varying ages, including neonates up to 28 days old, and children aged 5 to 16 years.

See Page 4, Lines 122-124. Neonates $\leq$ 28 days old and children $\leq$16 years old undergoing vital signs monitoring as per usual clinical protocol or as ordered by the physician, were identified and recruited.

See Page 7, Lines 230-236. A total of 38 patients were recruited for Phase 1, of whom 10 (26.3%) were neonates, and 28 (73.7%) were aged 5 to 16 years (Table 2). From this population, 765 data points were obtained (mean number of recordings per subject of 21.6$\pm$16, range of 0 to 51 measurements per patient). As we performed the pilot study in the neonatal ward first, we found that results were significantly discrepant compared to the ground truth. Test runs in 2 children aged 29 days to below 5 years were also discrepant. As there seemed to be minimal benefit in continuing to run the device on children below 5 years, we continued the rest of Phase 1 on children above 5 years in the paediatric wards. A breakdown of the number of recordings per age group is provided in Supplementary Table 1. The 2 test recordings were not included in the analysis in Table 2.

See Page 7, Lines 249-260. Further results from the pilot study showed that rPPG-derived vital signs values were clinically discrepant from the actual vital signs even for children aged between 5 to 10 years, despite accounting for movement and lighting. For example, HR values differed by as much as 30 to 50 beats per minute compared to pulse oximetry readings, with values being more discrepant the younger the child. A decision was made not to perform formal data analysis for all children below 10 years in view of the above

finding, and to focus on the older children. For patients aged 10 to 16 years (21 patients, 524 data points), the Rs value obtained for HR was 0.50 (95% CI 0.42, 0.57). Rs improved to 0.56 (95% CI 0.47, 0.64) at HR values below 100. Values for RR and SpO2 were 0.07 (95% CI -0.02, 0.16) and -0.03 (95% CI -0.12, 0.05) respectively.

For Phase 2, we further narrowed our focus to older children aged 12 to 16 years, assuming their baseline heart rates would mainly fall below 100 at rest.

See Page 8, Lines 282-283. Our study found that the use of rPPG technology is feasible and acceptable across varying age groups of paediatric patients, from neonates (up to 28 days old) to children aged 5 to 16 years.

See Page 8, Lines 292-293. To our best knowledge, this is the first study to evaluate the use of rPPG in obtaining paediatric vital signs across varying age groups.

See Page 9, Lines 312-315. In contrast, our study was conducted on clinically stable children of varying ages using available ambient lighting, and as such, we propose that our findings may be more generalizable, but require further validation by other investigators.

I commend the authors for collecting skin tone data given the concerns for pulse oximetry accuracy relative to skin pigmentation which could in theory be issues with rPPG too.

Results: Feasibility and parent acceptability data is strong. I wonder though if storing the images would be helpful to further evaluate the technology and assess why it wasn't as accurate in younger ages.

Reply 2. As an extension to our study, we obtained ethical approval and informed consent to store videos of recruited children, and these were shared with the industry partner to help further evaluate the technology and refine it for younger ages.

Changes in the text 2. We have modified the text to include the above information as follows (See Page 9, Lines 334-337):
To further evaluate this technology in younger children, subsequent ethical approval and informed consent was obtained for video images to be recorded, and these were sent to our collaborator for analysis. These video images were taken from 7 children ranging in age from 3 days to 6 years.

How many measurements were done for the different age groups? I see your break down of the enrollment but don't see how many measurements were done within

each age group which makes me wonder if that played a role in the discrepancies seen in younger ages.

Reply 3. The breakdown of measurements by age group for Phase 1 is as follows:

Age group	Number of patients	Number of recordings
≤ 28 days old	10	172
29 days to < 5 years	2	2
≥ 5 years to < 12 years	11	248
≥ 12 years to 16 years	17	345

Changes in the text 3. We have added a supplementary table in Page 21 (under Supplementary Material). We have also modified the text as follows (See Page 7, Lines 234-235): A breakdown of the number of recordings per age group is provided in Supplementary Table 1.

Discussion: Thank you for noting the limitation of not recruiting patients with Fitzpatrick zone I or II (or lighter skin). I think you should note the small number recruited from the darkest skin zone V too given the concerns with pulse oximetry accuracy in darker pigmented patients. In your reference to Heiden et al showing no impact of Fitzpatrick skin type to accuracy, was the comparison to pulse oximetry and rPPG? Or was it compared to blood gas SpO2? If the comparison was to pulse oximetry then the rPPG could be just as inaccurate as pulse oximetry for SpO2 measurement in darker patients and thus not a correct measurement of arterial oxygen which the gold standard measurement would be blood gas measurement.

I think you should also note a limitation of no patients between neonate to 6 years old. Or at least note why this gap in age was done in the pilot phase.

Reply 4. We have modified our text as advised to include a statement on the small number with the darkest skin tone.

For the Heiden et al study, they compared rPPG RR to manual counting over 60 secs, and HR & BP with a "standard clinical automatic sphygmomanometer on one arm, allowing both to be measured simultaneously". Their study did not use blood gas measurements.

We have clarified the age ranges of the patients as in the above reply (Reply 1) and modified the text accordingly (Changes in the text 1).

Changes in the text 4. We have modified the text as follows:

See Page 10, Lines 357-358. In addition, we only had 1 patient with Fitzpatrick

skin phototype VI, which is the most pigmented skin type.

See Page 10, Lines 363-364. In this study, the authors compared rPPG RR to manual counting over 60 secs, and HR & BP with a standard clinical automatic sphygmomanometer on one arm.

Reviewer C

The authors present an evaluation of the feasibility, acceptability, and accuracy of obtaining heart rate (HR), respiratory rate (RR) and oxygen saturation (SpO2) using remote PPG in children.

Minor issues:
Several places in the manuscript claim that this study assessed the feasibility of rPPG "across all age groups". However, Phase 1 did not include subjects aged between 1 and 5 years, and Phase 2 did not include subjects under 12 years of age. The results shown in Figures 1 and 2 appear to have been drawn from Phase 2 of the trial which only included ages 12 to 18. Claims about the age ranges of participants included in the study should be clarified throughout the manuscript. This includes point 2 in the "What is known and what is new" section of the Highlight Box, and the sentences on lines 75, 158, 234, 244 and 262.

Reply 1. Feasibility analysis was performed only for Phase 1, and patients included were neonates (≤ 28 days old) and children aged 5 to 16 years. As such, we have clarified on the age groups for Phase 1 and amended the text to reflect this, as per our response to Reviewer B (Reply 1).

Changes to the text 1. We have modified the text to clarify the patient age groups as follows:

See Page 2, Line 66. Ten neonates and 28 children aged 5-16 years were recruited for Phase 1 (765 datapoints).

See Page 2, Lines 75-77. Our study showed that rPPG is acceptable and feasible for neonates and children aged 5-16 years, and HR values in children aged 12 to 16 years correlated well with the current standard.

See Highlight Box. Our study is the first to assess the feasibility and accuracy of rPPG in the paediatric population across varying ages, including neonates up to 28 days old, and children aged 5 to 16 years.

See Page 4, Lines 122-124. Neonates ≤ 28 days old and children ≤16 years old undergoing vital signs monitoring as per usual clinical protocol or as ordered by the physician, were identified and recruited.

See Page 7, Lines 230-236. A total of 38 patients were recruited for Phase 1, of whom 10 (26.3%) were neonates, and 28 (73.7%) were aged 5 to 16 years (Table 2). From this population, 765 data points were obtained (mean number of recordings per subject of 21.6±16, range of 0 to 51 measurements per patient). As we performed the pilot study in the neonatal ward first, we found that results were significantly discrepant compared to the ground truth. Test runs in 2 children aged 29 days to below 5 years were also discrepant. As there seemed to be minimal benefit in continuing to run the device on children below 5 years, we continued the rest of Phase 1 on children above 5 years in the paediatric wards. A breakdown of the number of recordings per age group is provided in Supplementary Table 1. The 2 test recordings were not included in the analysis in Table 2.

See Page 7, Lines 249-260. Further results from the pilot study showed that rPPG-derived vital signs values were clinically discrepant from the actual vital signs even for children aged between 5 to 10 years, despite accounting for movement and lighting. For example, HR values differed by as much as 30 to 50 beats per minute compared to pulse oximetry readings, with values being more discrepant the younger the child. A decision was made not to perform formal data analysis for all children below 10 years in view of the above finding, and to focus on the older children. For patients aged 10 to 16 years (21 patients, 524 data points), the Rs value obtained for HR was 0.50 (95% CI 0.42, 0.57). Rs improved to 0.56 (95% CI 0.47, 0.64) at HR values below 100. Values for RR and SpO2 were 0.07 (95% CI -0.02, 0.16) and -0.03 (95% CI -0.12, 0.05) respectively.

For Phase 2, we further narrowed our focus to older children aged 12 to 16 years, assuming their baseline heart rates would mainly fall below 100 at rest.

See Page 8, Lines 282-283. Our study found that the use of rPPG technology is feasible and acceptable across varying age groups of paediatric patients, from neonates (up to 28 days old) to children aged 5 to 16 years.

See Page 8, Lines 292-293. To our best knowledge, this is the first study to evaluate the use of rPPG in obtaining paediatric vital signs across varying age groups.

See Page 9, Lines 312-315. In contrast, our study was conducted on clinically stable children of varying ages using available ambient lighting, and as such, we propose that our findings may be more generalizable, but require further validation by other investigators.

Are the algorithms used in the Nervotec device proprietary? If so then this should be mentioned in the paragraph starting on line 129, particularly as the authors flag

the lack of transparency of other rPPG algorithms as a limitation in the Discussion section (line 260).

Reply 2. Yes, the algorithms used in the Nervotec device are proprietary.

Changes in the text 2. We have modified the text as follows:
See Page 5, Lines 132-133. This rPPG technology was developed based on well-validated methods described in existing literature (5,16,17), and is Nervotec's proprietary rPPG software for contactless vital signs monitoring.

The paragraph starting at line 157 and the paragraph starting at line 168 describe generation of a Bland-Altman analysis for Phase 1 and Phase 2 of the study respectively. However, it is unclear whether the results shown in Figures 1 and 2 are from Phase 1 or Phase 2. The captions for Figures 1 and 2 should state whether it shows the results from Phase 1 or Phase 2 of the study. The source of the data used in Figures 1 and 2 should also be made clearer in the text (line 244).

Reply 3. We have modified the captions for Figures 1 and 2 and clarified in the text on the source of the data.

Changes in the text 3.
We have modified the caption for Figure 1 as follows: Bland-Altman plots showing the measurements of HR, SpO2 and RR by pulse oximetry (X-axis) and HR, SpO2 and RR by rPPG (Y-axis), in children aged 12 to 16 years (Phase 2).

We have modified the caption for Figure 2 as follows: Correlation scatterplots for rPPG-derived HR, SpO2 and RR (X-axis) and pulse oximetry-derived HR, SpO2 and RR (Y-axis) with corresponding R values, in children aged 12 to 16 years (Phase 2).

We have modified the text as follows:
See Page 8, Lines 272-273. Figure 1 presents Bland-Altman plots of rPPG HR, SpO2 and RR values compared to the corresponding reference standards, for children aged 12 to 16 years (Phase 2).

The paragraph starting on line 266 describes how the correlation between SpO2 measurements improves when SpO2 values below 97% are removed. It is unclear what this threshold of 97% is referring to. Is it the Masimo SpO2 measurement? The Nervotec SpO2 measurement? Or the average of the two measurements? The source of the 97% threshold should be stated.

Reply 4. This threshold refers to the oximetry Masimo measurement (standard of care).

Changes in the text 4. We have modified the text as follows:
See Page 9, Lines 317-319. Although rPPG SpO2 did not have strongly positive correlation with oximetry SpO2 in the older children, it was observed that above a threshold of 97% (based on oximetry), the agreement between the 2 sets of readings improved.

Heart rates are quoted using mixed units. Specifically, Table 1 uses "BPM" but in line 277 the units are referred to as "/min". I suggest editing the sentence on line 277 so the units are consistent with Table 1.

Reply 5. We have standardized the units to BPM as advised.

Changes in the text 5. We have modified the text as follows:
See Page 9, Lines 328-330. For younger children below 12 years of age with baseline heart rates above 100 beats per minute (BPM)/min, the current algorithm used in our study would require more work and refinement to accurately assess HR and SpO2.

In Table 1. SpO2 accuracy is quoted as "+/-2% ARMS". A Root Mean Square Error analysis is only capable of generating a positive numerical value, so there is no need to include the "+/-" before the "2%".

Reply 6. The +/- has been removed as advised.

Changes in the text 6. We have modified Table 1 as advised.

Why was Root Mean Square Error used to assess the accuracy of the SpO2 measurements but not for HR and RR? The authors should justify why Root Mean Square Error (a calculation requiring multiple measurements as input) was used as the "Regulatory Standard" for SpO2 but used a different method for HR and RR.

Reply 7. The choice to use Root Mean Square Error (RMSE) for assessing the accuracy of SpO2 measurements while employing a different method for HR and RR stems from the distinct regulatory standards associated with each vital sign.
For SpO2, we used ISO Standard 80601-2-61:2019 which specifies an accuracy requirement of RMSE ≤ 2% for pulse oximeter equipment.
For HR and RR, we opted for Mean Absolute Error (MAE) to ensure consistency and alignment with established standards.
We have added justifications in the text on the use of the various standards, including Root Mean Square Error for SpO2. In addition, for clarity, we have changed the term Accuracy Root Mean Square (ARMS) to Root Mean Square Error (RMSE) in the text and modified Table 1 accordingly.

Changes in the text 7. We have modified the text as follows:

See Page 6, Lines 191-201. We also looked at the accuracy of rPPG-derived vital signs for the different ages (Table 1). Regulatory standards from the American Standards National Institute ANSI/AAMI EC13-2002 (20) were utilised to assess the clinical accuracy of rPPG technology for HR (21). The decision to reference ANSI/AAMI EC13-2002 was made in the context of comparing our rPPG estimates with established standards for Electrocardiograph (ECG) devices. As there are no widely recognized regulatory standards for measurements of RR, we referenced clinical guidelines (22) to derive a conservative range of ±4 Respirations per minute (RPM). For accuracy metrics for HR and RR, we used Mean Absolute Error (MAE) to ensure consistency and alignment with established standards. For SpO2, we used the BS EN International Standard for Organization (ISO) 80601-2-61:2019 standard which states that the root mean square error (RMSE) must not exceed 2% of the SpO2 range (23).

We have modified Table 1 to change ARMS to RMSE.

The authors quote ANSI/AAMI EC13:2002 in reference #17 as the source for the "Regulatory Standards'' listed in Table 1. This document is a standard relating to "Electrocardiograph (ECG) heart rate and waveform monitors". The Nervotec system being tested does not fit within the class of devices this standard was designed to cover. The authors should justify why they elected to use this document as the source for regulatory standards used in Table 1.

Reply 8. Our intent in referencing ANSI/AAMI EC13:2002 was to align our accuracy benchmarks with established standards for ECG devices. While acknowledging the distinct nature of our rPPG software, we sought to demonstrate a level of precision akin to ECG devices.

Changes in the text 8. We have modified the text as per the previous comment in Reply 7.

The ANSI standard referred to in reference #17 states that "Cardiac monitors labeled for use with neonatal/pediatric patients shall have an extended heart rate range of at least 250 bpm", however Table 1 of the manuscript appears to have adopted the HR ranges of 30 to 200 BPM for adult patients. This should be amended as this manuscript specifically describes the assessment of the rPPG approach in a pediatric setting.

Reply 9. Thank you for this observation. We have modified Table 1 as advised.

Changes in the text 9. We have modified Table 1 as advised.

The ANSI standard in reference #17 does not contain any regulatory guidance for measurement of SpO2 or RR. The authors should describe how the regulatory standards for these signals were chosen.

Reply 10. We acknowledge the limitation in referencing ANSI/AAMI EC13:2002, which lacks guidance for SpO2 and RR measurements. We will clarify the reference for regulatory standards for SpO2 and RR. While there isn't a widely recognized standard for RR, we referenced established clinical guidelines and for accuracy metric, we used Mean Absolute Error (MAE). A conservative range of ±4rpm was devised. This acknowledges the variability in manual counting methodologies. For SpO2, we used the (ISO) 80601-2-61:2019 standard which recommends an RMSE of ±2% for accurate SpO2 measurements.

Changes in the text 10. We have modified our text as per the previous comment in Reply 7.

Figure 2 shows that for HR and RR there is a tendency for the Nervotec device to under-report high measurements and over-report low measurements. The authors may wish to acknowledge this observation in their discussion section and speculate on possible causes for this systematic bias.

Reply 11. We thank the reviewer for this observation. One contributing factor to the observed bias is that our algorithm was primarily trained on an adult population, where lower heart rates and respiratory rates are more prevalent compared to the pediatric population. The specific signal processing techniques employed, including screening filters, were initially designed to extract information within a certain range. This design posed a challenge when applied to pediatric subjects who exhibited higher heart rates and respiratory rates. The inherent bias towards under-reporting high measurements and over-reporting low measurements was recognized as an area for improvement.

Changes in the text 11. We have modified the text as follows:

See Page 9, Lines 339-346. We also observed that the rPPG device in our study tended to under-report high measurements and over-report low measurements. One contributing factor to the observed systematic bias is that our algorithms were primarily trained on adult populations, where lower heart rates and respiratory rates are more prevalent compared to the paediatric population. The specific signal processing techniques employed, including screening filters, were initially designed to extract information within a certain range. This design posed a challenge when applied to paediatric subjects. Subsequent adjustments will be made by our collaborator,

including the use of different screening filters to mitigate this bias and further improve accuracy results.

Reference #14 from the manuscript (Pologe and Menschik 2012) is a letter to the editor in response to Causey et al. (2011) "Validation of noninvasive hemoglobin measurements using the Masimo Radical-7 SpHb Station." The accuracy of hemoglobin measurements made using the Masimo device is not relevant to the study described in the submitted manuscript. A more relevant reference should be used here.

Reply 12. This reference has been removed.

Changes in the text 12. We have modified the numbering of references in the text.

Major issues:
The study uses correlation (R values) to quantify the level of agreement between the two measurement approaches. However, highly correlated measurements do not necessarily indicate agreement. Several references cited in the manuscript also make this point. For example, Reference #19 (Schober et. al. 2018) states that "Correlations ... do not describe the strength of agreement between 2 variables" and that "two variables can exhibit a high degree of correlation but can at the same time disagree substantially". Schober et. al. conclude their article by saying, " ... correlation is unsuited for analyses of agreement". Reference #14 (Pologe and Menschik 2012) is also highly critical of the use of correlation for assessment of measurement agreement. Pologe and Menschik conclude that "a correlation (or correlation coefficient) can be quite good even when the instrument under evaluation is so inaccurate as to be clinically without value. The correlation coefficient indicates only the degree to which the data ... demonstrate a linear relationship. However, this statistic provides little insight as to the clinical usefulness of a measurement device."
Correlation, or R values, are erroneously used to imply agreement/accuracy throughout the manuscript. These comments appear in the results section of the abstract, in the text on lines 163, 227-230, 237, 239 and 266, and in point 2 under "Key Findings" in the Highlight Box. The authors should clarify the limitations of using R values to assess measurement agreement and/or select an alternative method of assessing agreement between the two devices.

Reply 13. We would like to clarify an error in our manuscript. Spearman's correlation was used for the correlation analyses in our study, and the R values refer to Spearman's correlation coefficent (not Pearson's). For clarity, this has been amended to **Rs** values in our revised manuscript.

We thank the reviewer for his advice that correlation does not necessarily

indicate agreement, and have modified the manuscript to avoid concluding accuracy of the rPPG device based on correlation results. We have also added a limitation of our study on the use of correlation coefficient results to assess measurement agreement.

Changes in the text 13.
In the text, we have replaced “Pearson’s” with “Spearman’s” and “R” with “Rs”
See:
Page 2, Lines 69-72
Page 6, Lines 201-203
Page 7, Lines 254-256
Page 8, Lines 285-286

We have added in a new reference for the use of Spearman’s correlation as follows:
28. Zar JH. Spearman rank correlation: overview. Wiley StatsRef: Statistics Reference Online. 2014 Apr 14

We modified the text in the Abstract as follows:
See Page 2, Lines 75-77. Our study showed that rPPG is acceptable and feasible for neonates and children aged 5-16 years, and HR values in children aged 12 to 16 years correlated well with the current standard.

We removed Lines 286-287 in Page 8.

We modified the text as follows:
See Pages 9-10, Lines 348-353. One limitation of our study was the use of Rs values to quantify the level of association or agreement between rPPG readings and the standard of care. There are several pitfalls to using the correlation coefficient alone to conclude agreement between 2 variables, and a high degree of correlation may not necessarily equate to true agreement (38). As such, we concurrently performed Bland-Altman analyses to further assess the strengths of the agreements and provide some conclusion on the overall accuracy of rPPG in comparison to the current standard of care.

We modified the text as follows:
See Page 10, Lines 377-382. For measurement of HR, values obtained by rPPG correlated well with the current standard of care in children aged 12 to 16 years. This result is promising, and future studies should expand further on the clinical accuracy of this technology for assessment of HR in older children. However, more work is required to refine the rPPG algorithms for younger children, and for obtaining RR and SpO2 in all children.

On line 238 the authors state that assessment of HR by rPPG was "clinically

accurate compared to the established standard". A similar sentence also appears on line 75. These sentences should be removed or amended unless additional analysis is performed and reported. The correlation results reported in the manuscript do not imply accuracy, nor was the Massimo device an "established standard" as its measurements also contain uncertainties. For example see Blanchet, Marie-Anne, et al. "Accuracy of Multiple Pulse Oximeters in Stable Critically Ill Patients." Respiratory Care 68.5 (2023): 565-574.

Reply 14. We have removed the above line (previously Line 238) as advised.

Changes in the text 14. We have modified the text by removing the above sentence (See Page 8, Line 286-287).

Line 177 states "We used the regulatory standards specified in Table 1 to assess clinical accuracy of rPPG-derived vital signs ... ", however I did not find the results of this assessment in the manuscript. The authors should either remove this sentence, or report the results of their assessment. For example, these results could be reported by describing what percentage of HR, RR, and SpO2 measurements fell inside/outside the regulatory standards listed in Table 1.

Reply 15. The line has been removed as advised.

Changes in the text 15. We modified the text as follows:
See Page 6-7, Lines 215-217. We performed correlation and Bland-Altman analyses to compare rPPG-derived vital signs with the standard of care.

The regression slope of the measurement comparisons was assumed to be 1 (line 172) however this assumption was not tested. Based on a cursory visual analysis of Figure 2, the regression slope for HR appears to be around 0.75. The authors should report the slope of the regression lines used to calculate the R values. The implications of these slopes on the calculation of required sample sizes and confidence intervals should be recalculated (lines 169-172, 214-216). If these results are impactful they should be reported as limitations or biases in the manuscript.

Reply 16. We have tested the hypothesis used under the sample size calculation and the regression slope of HR (beta) is 0.92 (90% CI 0.80, 1.04). Since this is a prospective study, the sample size calculated was prior to recruitment, and post-hoc sample size calculation is not recommended in prospective studies. From the results of HR, it turns out that the study is adequately powered to test the hypothesis of slope (HR) = 1 as the 90% CI of the regression slope includes 1.

Changes in the text 16. We have modified the text as follows:

See Page 8, Lines 262-264. We tested the hypothesis used under the sample size calculation and the regression slope for HR (beta) was 0.92 (90% CI 0.80, 1.04).

See Page 9, Lines 308-310. Sample size analysis showed that our study was adequately powered to assess rPPG HR (based on the test of hypothesis of slope = 1, as the 90% CI of the regression slope included 1).

In addition, we modified Lines 256-257 in Page 7 as follows: Values for RR and SpO2 were 0.07 (95% CI -0.012, 0.1605) and -0.03 (95% CI -0.1202, 0.0516) respectively.