



Assessment of neonatal respiratory rate variability

Jesse Coleman^{1,7} · Amy Sarah Ginsburg² · William M. Macharia³ · Roseline Ochieng³ · Dorothy Chomba³ · Guohai Zhou⁴ · Dustin Dunsmuir⁵ · Walter Karlen⁶ · J. Mark Ansermino⁵

Received: 18 August 2021 / Accepted: 2 March 2022 / Published online: 25 March 2022
© The Author(s) 2022

Abstract

Accurate measurement of respiratory rate (RR) in neonates is challenging due to high neonatal RR variability (RRV). There is growing evidence that RRV measurement could inform and guide neonatal care. We sought to quantify neonatal RRV during a clinical study in which we compared multiparameter continuous physiological monitoring (MCPM) devices. Measurements of capnography-recorded exhaled carbon dioxide across 60-s epochs were collected from neonates admitted to the neonatal unit at Aga Khan University-Nairobi hospital. Breaths were manually counted from capnograms and using an automated signal detection algorithm which also calculated mean and median RR for each epoch. Outcome measures were between- and within-neonate RRV, between- and within-epoch RRV, and 95% limits of agreement, bias, and root-mean-square deviation. Twenty-seven neonates were included, with 130 epochs analysed. Mean manual breath count (MBC) was 48 breaths per minute. Median RRV ranged from 11.5% (interquartile range (IQR) 6.8–18.9%) to 28.1% (IQR 23.5–36.7%). Bias and limits of agreement for MBC vs algorithm-derived breath count, MBC vs algorithm-derived median breath rate, MBC vs algorithm-derived mean breath rate were -0.5 ($-2.7, 1.66$), -3.16 ($-12.12, 5.8$), and -3.99 ($-11.3, 3.32$), respectively. The marked RRV highlights the challenge of performing accurate RR measurements in neonates. More research is required to optimize the use of RRV to improve care. When evaluating MCPM devices, accuracy thresholds should be less stringent in newborns due to increased RRV. Lastly, median RR, which discounts the impact of extreme outliers, may be more reflective of the underlying physiological control of breathing.

Keywords Child health · Critical care · Delivery of health care · Diagnosis · Patient care

✉ Jesse Coleman
denots@gmail.com

¹ Evaluation of Technologies for Neonates in Africa (ETNA), Nairobi, Kenya

² Clinical Trials Unit, University of Washington, Seattle, WA, USA

³ Department of Pediatrics, Aga Khan University, Nairobi, Kenya

⁴ Center for Clinical Investigation, Brigham and Women's Hospital, Boston, MA, USA

⁵ Anesthesiology, Pharmacology & Therapeutics, The University of British Columbia, Vancouver, BC, Canada

⁶ Mobile Health Systems Lab, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland

⁷ Centre for International Child Health, 305 – 4088 Cambie Street, Vancouver, BC V5Z 2X8, Canada

1 Introduction

Accurate measurement of respiratory rate (RR) in children is particularly important in low-resource settings where illness and deterioration are typically diagnosed based on a child's clinical signs and symptoms [1–3]. To measure RR, the World Health Organization (WHO) recommends observing and counting chest and abdominal movements over a full 60 s [4]. In practice, this recommendation is frequently modified to counting respirations over a shorter duration of time (e.g., counting breaths for 10 s and multiplying by six). This modification can result in up to a 50% discrepancy compared to WHO recommendations [5]. In addition to inaccuracy, manual breath counting can be influenced significantly by counter bias and lacks reproducibility.

In neonates, measuring RR accurately is especially challenging given high RR and the within-neonate RR variability (RRV) [6]. Control of breathing remains immature until after the first month of life. Neonates often demonstrate periodic

breathing, a benign, regular manifestation of irregular respiratory patterns, consisting of contiguous periods of alternating breaths and respiratory pauses. Healthy neonates also may exhibit benign irregularly irregular breathing patterns with short periods of apnea, similar to the disrupted respiratory control and ataxic breathing which is seen with opioid administration [7, 8]. These irregular breathing patterns are distinct from the regular rhythms of periodic breathing or the apnea of prematurity, though there appears to be some amount of overlap [9, 10]. In contrast to older infants and children, neonates have a marked degree of breath-to-breath variability and are also more likely to exhibit episodes of respiratory pause associated with stimulation and in response to hypoxia [11]. This individual-level, within-neonate RRV, adds to the complexity of identifying and reliably measuring true RR [12]. Perturbations, including sighs, swallows, and coughs, can affect the duration of individual breaths, the resulting RR, and the within-neonate RRV. Given the high RRV in neonates, quantifying RR by observing chest and abdominal movements as recommended by the WHO is fraught with potential inaccuracy [13].

There will always be some level of uncertainty when performing manual or automated measurement of RR. All potential sources for uncertainty in RR measurement should be considered (Table 1). The size and acceptability of the resulting uncertainty depend on the conditions and context of the measurement. In some clinical situations, high accuracy may not be necessary. However, in the emergency department or intensive care unit, when it is crucial to determine if a patient has crossed a diagnostic threshold, accurate RR measurement can enable early identification and expedited management of patient deterioration [14–17]. In research laboratories, accuracy and precision are essential for determining how a new device or method compares with the reference method.

There is growing evidence which suggests including RRV in clinical scoring systems may be beneficial for guiding escalation and de-escalation of care [18, 19]. Heart rate variability (HRV), unlike RRV, has been extensively studied and

is commonly used as a marker of risk for mortality [20–23]. HRV is considered to be central in the clinical assessment of diverse conditions that include neurological and sleep disorders, muscular dystrophy, and diabetes in adults, and sepsis in neonates [24, 25]. Like HRV, the identification of changes in RRV could be used as an indicator of underlying physiological disturbances [18, 19, 26].

Numerous innovative RR monitoring methods and devices for both adults and neonates using non-contact video, sensors embedded in bedding, motion sensors, nanoparticles, and temperature-based methods have been reported previously [27–31]. The variety of monitoring methods has resulted in various different evaluation methods and difficulty when trying to compare results across studies. A recent systematic review of RR monitoring systems suggested standardizing validation frameworks to directly compare different RR monitoring methods and systems [32]. Fortunately, detailed verification and validation recommendations have been made [33, 34]. If followed, these recommendations may result in future cross-comparable research of neonatal RRV. The current research is a result of data analyzed during systematic verification conducted within a device comparison study.

In our study comparing neonatal multiparameter continuous physiological monitoring (MCPM) devices in Nairobi, Kenya we sought to quantify RRV between- and within-neonates, as well as between- and within-epochs to identify the best methods for device comparison studies. We believe this RRV quantification will inform management of uncertainty and RRV when designing, developing, and comparing RR monitoring devices in neonates.

2 Methods

2.1 Setting and participants

We conducted a clinical verification phase of the reference RR monitoring device while studying low-cost neonatal

Table 1 Potential sources of uncertainty in respiratory rate measurement, approximate distribution, and potential solutions

Uncertainty	Source	Distribution		Possible solution
Breath-to-breath variation	Neonate	Normal	Random	Extended within-neonate observation
Perturbation (e.g., sigh, swallow, cough)	Neonate	Multimodal	Non-random	Exclude outliers
Device uncertainty due to inaccurate timing	Device	Normal	Random	Repeatability with multiple simultaneous devices
Device uncertainty due to missed breath(s)	Device	Multimodal	Non-random	Repeatability with multiple simultaneous devices
Observer error due to inaccurate timing	Observer	Normal	Random	Repeatability with multiple simultaneous observers
Observer error due to missed breath(s)	Observer	Multimodal	Non-random	Repeatability with multiple simultaneous observers
Rounding error due to counting breaths in a time interval	Analysis	Normal	Random	Measure breath interval (not count)

The nature of the distributions is based on knowledge of the underlying processes and has not been validated

Table 2 Study definitions and eligibility criteria

Study definitions	
Epoch	A 60-s period of time
Breath	One cycle of neonate-initiated inhalation and exhalation (Table 3)
Breath duration	Length of time from the start to the end of a single breath
Breath start	End of a waveform trough (low point) where the carbon dioxide level starts to ascend
Respiratory rate	Number of breaths initiated within an epoch
Breath rate	The number of breaths within an epoch based on the median or mean of the breath duration
Respiratory rate variability (RRV)	The dispersion of respiratory rate or breath duration (a reciprocal of the respiratory rate) within an epoch or between epochs, calculated as the standard deviation expressed as a percentage of the mean; epochs for comparison are defined below
Within-subject RRV	RRV_{bd} = Breath duration variability (average within each epoch) RRV_{bm} = Minute-to-minute variability (average between epochs) RRV_{10bd}/RRV_{60bd} = Breath duration variability between epochs measured at 10- or 60-min intervals
Between-subject RRV	RRV_{bs} = Between subject minute-to-minute variability
Neonatal eligibility criteria	
Inclusion criteria	<ul style="list-style-type: none"> • Male or female neonate, corrected age of <28 days • Willingness and ability of neonate's caregiver to provide informed consent and to be available for follow-up for the planned duration of the study
Exclusion criteria	<ul style="list-style-type: none"> • Receiving mechanical ventilation or continuous positive airway pressure • Skin abnormalities in the nasopharynx and/or oropharynx • Contraindication to the application of skin sensors • Known arrhythmia • Any medical or psychosocial condition or circumstance that, in the opinion of the investigators, would interfere with the conduct of the study or for which study participation might jeopardize the neonate's health

MCPM devices in Nairobi, Kenya [35]. Study participants were spontaneously breathing neonates admitted for observation and care in the maternity ward, neonatal intensive care, and the neonatal high dependency units at Aga Khan University-Nairobi (AKU-N) Hospital. Caregivers were approached, recruited, and sequentially screened for enrollment by trained study staff during routine intake procedures. Final eligibility determination was based on medical history, physical examination, appropriate understanding of the study by the caregiver, and completion of the written informed consent process (Table 2).

2.2 Study procedures and data collection

Detailed study procedures are described in the published protocol [35]. In brief, term and preterm neonates were enrolled in a MCPM accuracy and feasibility evaluation. Male or female neonates were eligible if they had a corrected age of <28 days and the caregiver was willing and able to provide informed consent and to be available for follow-up for the planned duration of the study. Neonates were excluded if they were receiving mechanical ventilation or continuous positive airway pressure, had skin abnormalities in the nasopharynx and/or oropharynx or a contraindication to the application of skin sensors, a known arrhythmia or any medical or psychosocial condition or circumstance that, in the opinion of the investigators, would interfere with the conduct of the study or for which study participation might jeopardize the neonate's health. Solely for the purposes of the study, we used the Masimo Rad-97 with NormoLine

capnography as a reference device to record and measure RR using exhaled carbon dioxide (CO_2) levels. The collected continuous capnography data were digitized at approximately 20 Hz using asynchronous communication with a custom software application. Capnography readings were collected for a minimum of one hour and continued until the neonate was discharged (range 1–6.25 h; median: 3.75 h). Demographic and capnography data were entered and stored on a secure AKU-N-hosted REDCap server [36].

Sixty-second epochs of capnography data were extracted at predetermined time intervals and converted to capnogram waveform tracings. Intervals between epochs were predetermined and based on study-related clinical observations: at 10-min intervals throughout the first hour of participation followed by 60-min intervals starting at the second hour [35]. The resulting capnogram tracings included a total of 64 s (Fig. 1); two seconds were added at the beginning and end of each epoch to facilitate manual breath counting of the epoch.

One of the authors (JMA, an anesthesiologist) reviewed all capnogram tracings for quality control; difficult-to-count plots were discarded ($n = 164$; Fig. 2). All remaining epochs were included, and breaths were manually counted from capnograms and identified using an automated signal detection algorithm. For the manual counting, capnographs were provided to two trained observers to count all breaths within each epoch independently, and the results were averaged. A breath was identified using a set of predefined rules created by the investigators (Table 3). If the number of breaths counted by the two observers varied by more than three

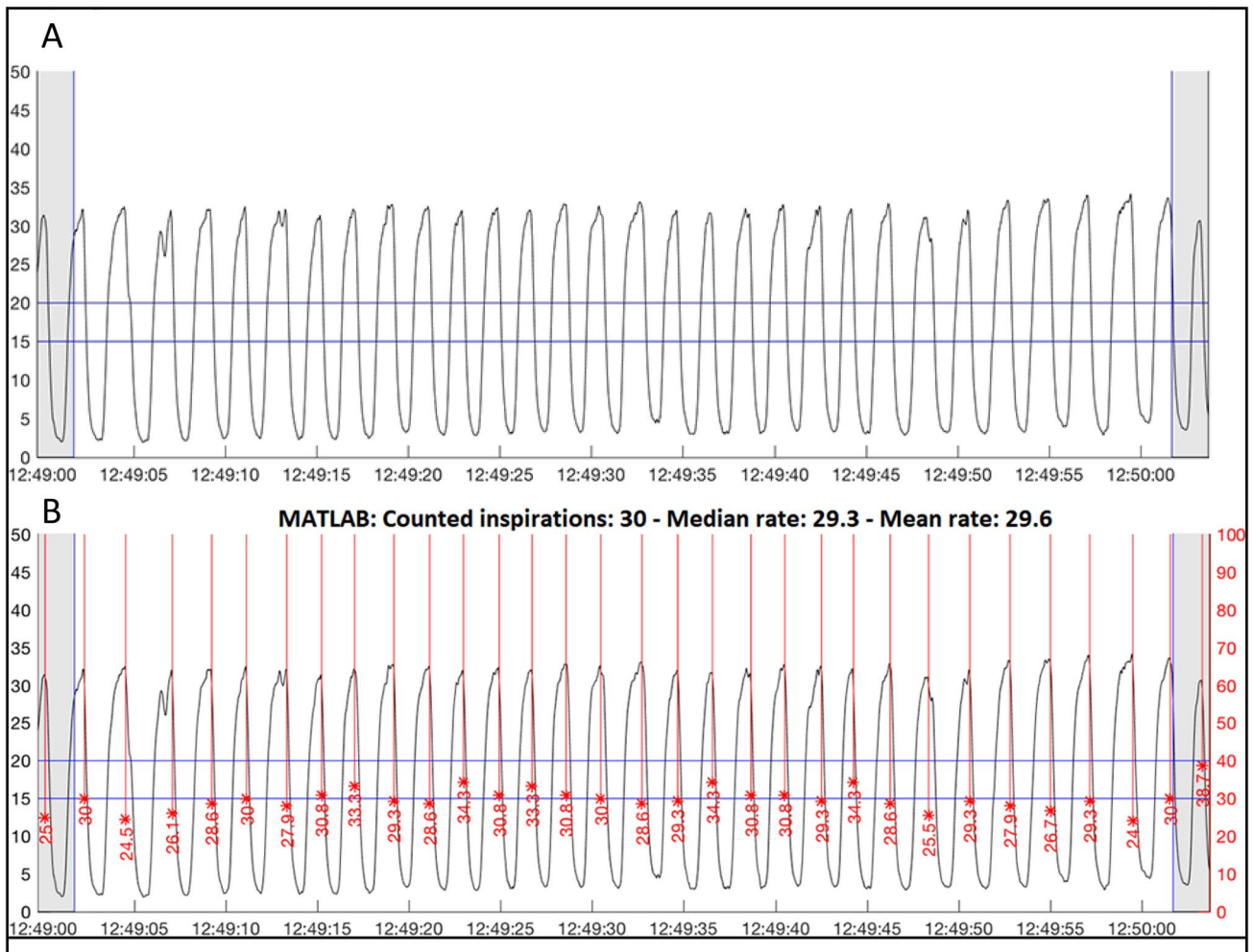


Fig. 1 Example capnograms (carbon dioxide (CO_2) waveform plots) before (**A**) and after (**B**) algorithm processing. The plotted CO_2 waveform shows the breathing pattern of a neonate and algorithm-derived identification of breaths (red, vertical lines). Only peaks on the white background were included; peaks that fell within the grey zone were ignored as they were outside the 60-s epoch. **A** Plotted waveform from example epoch before processing by the algorithm. Each peak

within the 60-s epoch was counted by one to three trained observers. The horizontal blue 15 and 20 lines were used to assist observers during irregular or incomplete breaths (not shown). **B** Plotted waveform after processing by algorithm. The red vertical lines show identified peaks, with the length and label of the red line representing the calculated breath rate based on the breath duration

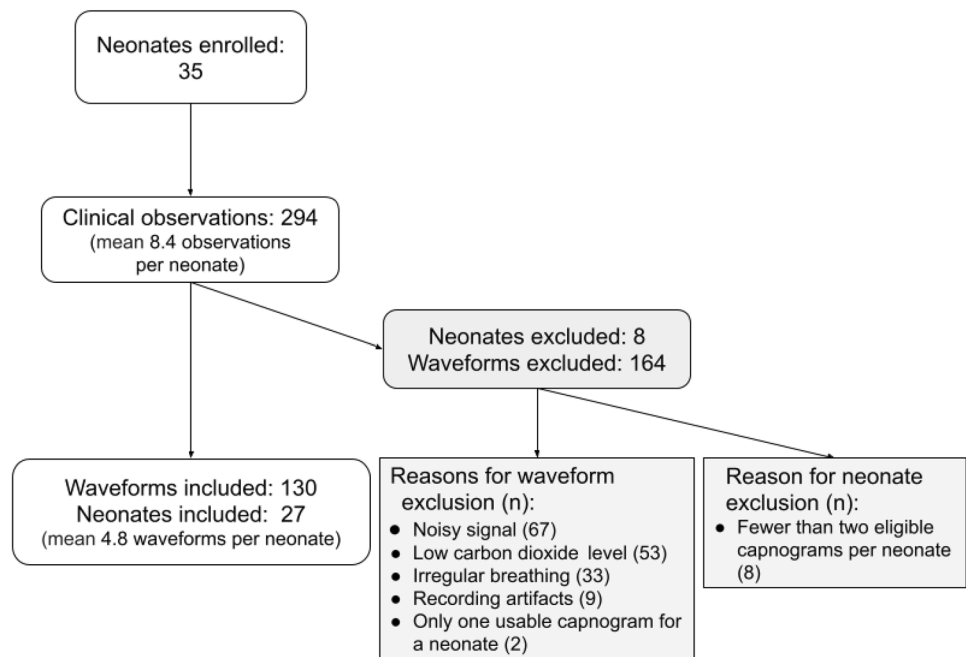
breaths per epoch, a third trained observer independently counted the plot, and the two closest counts were averaged.

The automatic breath detection method was based on a time-domain algorithm that identified regular patterns in physiological waveforms [37].

The algorithm was modified to identify unique breaths in the CO_2 signal by dividing the waveform in time and identifying changes in direction to define segments. Specifically, an incremental-merge algorithm divided the waveform into geometrically similar segments by establishing a line between equidistant samples and iteratively merging adjacent lines that had the same slope sign into single, longer lines. The resulting line segments of alternating slope signs defined inhalation and exhalation components, as well as

artifacts interrupting this sequence. An adaptive threshold was applied to the length of these segments to separate artifacts and double breaths from regular breathing components [37].

The breath duration was calculated between the beginning of two adjacent regular inhalation components that were interrupted by at least one exhalation component and no artifact. To investigate the effect of length of time between epochs on breath duration variability (RRV_{bd}), epochs were grouped into 10- and 60-min intervals for subgroup analysis.

Fig. 2 Recruitment flow diagram**Table 3** Rules for identifying breaths based on graphical waveform plots

1. Count peaks (tops) of the waveform that are within the white background. Ignore peaks that are within the grey background on either side of the image
2. A peak should be counted as a breath when the peak of the waveform is above 15 mmHg, the lower horizontal blue line
3. If the peak does not reach the lower horizontal blue line at 15 mmHg, to be counted as a breath, the peak should reach at least 50% of the mean peak
4. The waveform should dip down to the normal baseline (either below 15 mmHg, the lower horizontal blue line, or based on other breaths). If the waveform does not reach below this point, then this is considered part of the same (double) peak and only counted as a breath once

2.3 Data analysis

Results from the manual and automatic breath detection methods were analyzed using STATA 13 and R [38, 39]. The coefficient of variation, the standard deviation (SD) expressed as a percentage of the mean, was reported as RRV between- and within-neonates, as well as between- and within-epochs (see Table 2 for definitions).

Agreement between the manual breath counts and the algorithm-derived breaths was assessed using the methods described by Bland and Altman's Sect. 5.3 on replicated data pairs [40]. Agreement was reported as a mean bias with 95% confidence intervals (CIs) and 95% limits of agreement (LOA) and root-mean-square deviation (RMSD) [40].

2.4 Sample size

Sample size estimates for method comparison studies typically depend on the CI required around the LOA, and sample sizes of 100 to 200 provide tight CIs [40]. We estimated that

20 neonates with ten replications each would give the 95% CI of LOA between the first and second methods to be ± 0.76 times the SD of their differences. The study team aimed for a sample size of at least 30 neonates to ensure a diverse population and sufficient replications for tight CIs.

3 Results

Between June and August 2019, 35 neonates were enrolled, and 294 clinical observations were completed. We included 130 (44.2%) CO₂ waveform plots in this analysis (Fig. 2) across 27 neonates, 23 at term (range of gestational age 32–42 weeks). Four preterm neonates, born before 37 weeks gestation, were included. Three of the four preterm neonates received caffeine during their admission. There were on average 4.8 (range 2–9) epochs per subject.

The mean manual breath count was 48 breaths per minute (bpm) (95% CI 31–71) and the median RRV_{bm} was 25.8% (interquartile range (IQR) 22–31.7%; Table 4; Fig. 3A).

When grouped by neonate, the mean between-neonate manual breath count was also 48 bpm, while the median RRV_{bm} showed a narrower distribution (12.3%; IQR 9.8–19.4%). The median between-epoch algorithm-derived RRV_{bd} of 25.1% (IQR 21.1–30.8%) was marginally lower than the median manual breath count RRV_{bm} .

Within neonate RRV increased when observations were taken further apart. The median RRV_{60bd} was 4.6% higher compared with the RRV_{10bd} (28.1% (IQR 23.5–36.7%) vs 23.5% (IQR 20.5–28.5%)). A graphical representation of the within-neonate RRV_{10bd} and RRV_{60bd} trends over time showed a marked downward trend in RRV_{60bd} between the baseline and one-hour epochs (Fig. 3C); other time periods did not show this trend.

Manual breath count and the algorithm-derived breath count showed minimal bias (-0.52) and strong agreement (95% lower limit of agreement (LLA) -2.7 , 95% upper limit of agreement (ULA) 1.77 , RMSD 1.2 ; Table 5). However, the manual breath count had a larger bias (at least -3 bpm), and a larger normalized spread (95% LLA 37.2% and 95% ULA 30.4%) compared with both the algorithm-derived median and mean breath rates and a small bias and tighter spread of 95% LLA and 95% ULA compared with algorithm-derived breath counts (Fig. 4).

LOA were calculated based on log-transformed data which were found to be normally distributed as evidenced by the histogram and the Shapiro–Wilk test (S1). We also calculated a non-parametric version of the LOA (S2), which were the values outside which 5% of the observations fell, based on a nonparametric approach for comparing methods [40] which allows the use of dependent data. All these LOA provided qualitatively similar conclusions.

4 Discussion

The results, which showed a range of RRV from 11.5 to 25.8% within 60-s epochs, were consistent with previously published research quantifying RRV in neonates, children, and adults, supporting the external validity of the current data while also highlighting challenges when performing device comparison studies [12, 33, 41–43]. This range also suggested the RRV was within a normal range for short-term neonatal monitoring in relatively healthy neonates; no neonates had a deterioration of their condition during or immediately following participation in the study.

Manually counting breaths from a capnogram is a labor-intensive process commonly used in clinical RR measurement. However, due to its limited temporal precision, manual breath counting does not provide for the precise breath durations required when estimating within-epoch mean or median RRV , or breath rate variability. Human observers are not precise enough to identify the exact breath duration differences required to objectively discern accurate variation. An accurate breath identification algorithm can both identify individual breaths and breath lengths from raw waveform data extracted from the capnogram and can be automated. The algorithm used in the current study showed a high level of accuracy for the algorithm-derived breath count as compared to the manual breath count, evidenced by the small bias, tight 95% LLA and ULAs, and a small RMSD across all epochs, and was confirmed by plotting each breath in the capnograms. These results suggested that this algorithm correctly identified individual breaths and could identify breath duration and RRV . Each source of uncertainty (Table 1)

Table 4 Respiratory rate (RR) median and median coefficient of variation between- and within-neonates, as well as between- and within-epochs

		Epochs (n)	Median respiratory or breath rate (interquartile range)	Median respiratory or breath rate coefficient of variation (interquartile range)
Between-neonates	Manual count	130	46.7 (43.1–52.0)	12.3% (9.8–19.4%)
	Algorithm-derived	130	51.2 (45.1–61.8)	17.6% (10.7–24.2%)
Within-neonate	Manual count	130	46.8 (42.5–55.0)	11.5% (6.8–18.9%)
	Algorithm-derived	130	51.0 (45.1–59.3)	17.5% (8.8–24.1%)
Between-epochs	Manual count	130	47.0 (39.0–56.0)	25.8% (22.0–31.7%)
	Algorithm-derived	130	47.5 (40.0–56.0)	25.1% (21.1–30.8%)
Within-epoch	Manual count	N/A	N/A	N/A
	Algorithm-derived	130	50 (41.4–57.1)	20.8% (13.6–27.3%)
<i>Sub-group</i>				
Epochs at 10-min intervals	Algorithm-derived (18 neonates)	63	46.1 (38–53)	23.5% (20.5–28.5%)
Epochs at 60-min intervals	Algorithm-derived (19 neonates)	52	50.2 (38.5–60)	28.1% (23.5–36.7%)

Sub-group analysis includes algorithm-derived RR grouped by length of time between epochs. Sub-group analysis excludes neonates with fewer than two time-relevant epochs

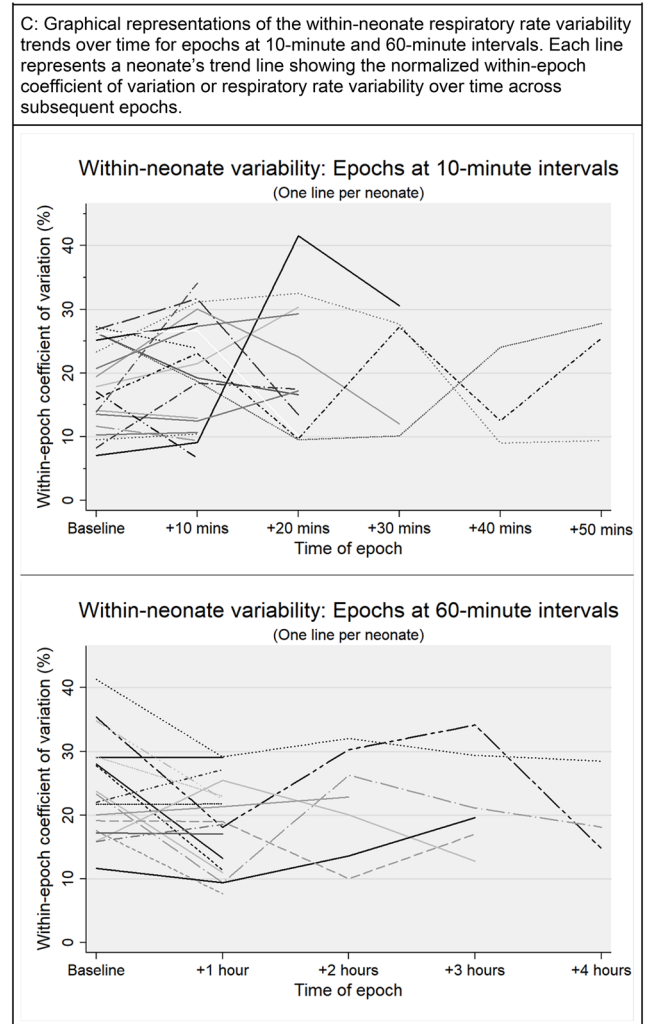
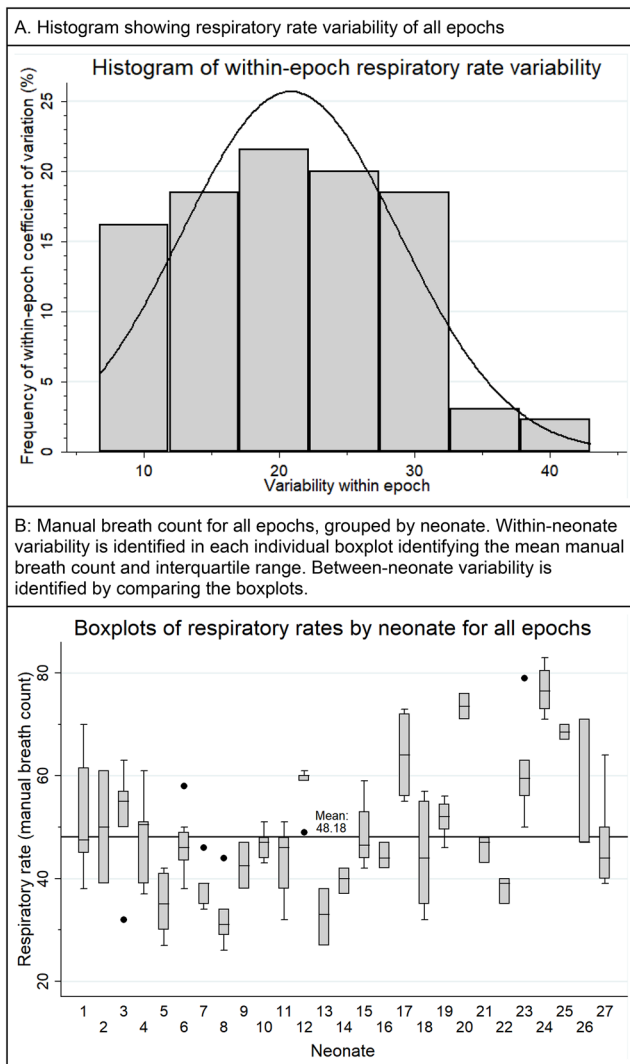


Fig. 3 Graphic representations of respiratory rate variability in all epochs (n = 130). **A** Histogram showing respiratory rate variability of all epochs. **B** Manual breath count for all epochs, grouped by neonate. Within-neonate variability is identified in each individual boxplot identifying the mean manual breath count and interquartile range. Between-neonate variability is identified by comparing the boxplots.

C Graphical representations of the within-neonate respiratory rate variability trends over time for epochs at 10-min and 60-min intervals. Each line represents a neonate’s trend line showing the normalized within-epoch coefficient of variation or respiratory rate variability over time across subsequent epochs

Table 5 Bland–Altman analysis results comparing manual breath count with algorithm-derived breath counts, median and mean breath rates

	Bias (normalized)	95% Upper/lower limits of agreement	Spread of upper and lower 95% limits of agreement (normalized)	RMSD (normalized)
Manual breath count vs algorithm-derived breath count	-0.52 (-1.1%)	-2.7/1.66	4.37 (9.1%)	1.2 (2.5%)
Manual breath count vs algorithm-derived median breath rate	-3.16 (-6.6%)	-12.12/5.8	17.92 (37.2%)	5.5 (11.4%)
Manual breath count vs algorithm-derived mean breath rate	-3.99 (-8.3%)	-11.3/3.32	14.62 (30.4%)	5.5 (11.4%)
Algorithm-derived breaths vs algorithm-derived median breath rate	-2.64 (-5.4%)	-11.54/6.27	17.82 (36.6%)	5.2 (10.7%)

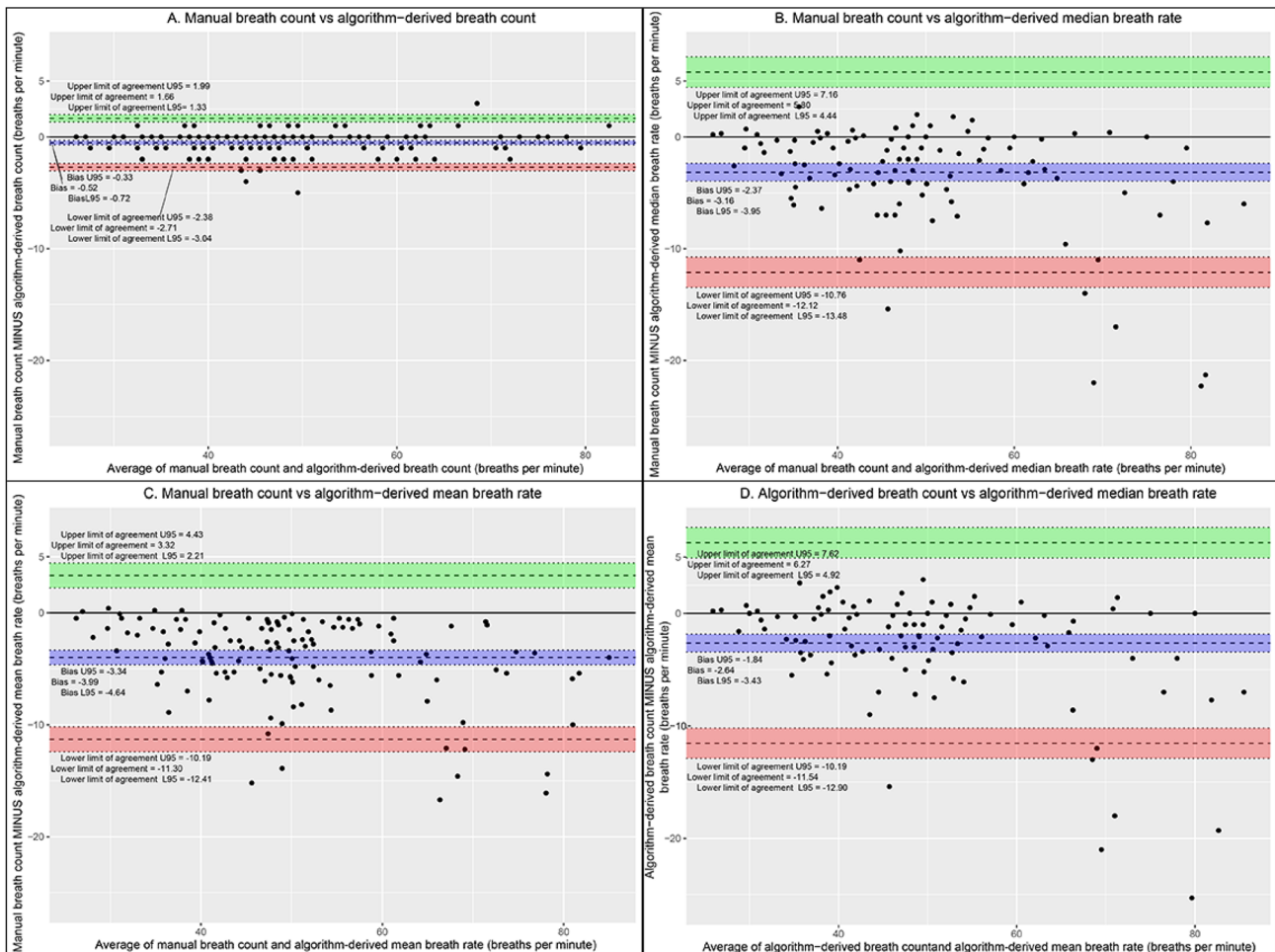


Fig. 4 Bland–Altman plots comparing manual breath count vs algorithm-derived breath count (**A**), manual breath count vs algorithm-derived median breath rate (**B**), manual breath count vs algorithm-

derived mean breath rate (**C**), and algorithm-derived breath count vs algorithm-derived median breath rate (**D**)

increased the challenge in estimating the true RR, and RRV should be considered when comparing MCPM devices.

The marked within-epoch RRV_{bd} that was identified highlights the challenge of performing accurate clinical RR measurements in neonates. Both RR and RRV will be significantly affected by the timing of the start of the epoch analysed. RRV also needs to be reflected in typical clinical decision-making thresholds. A longer measurement period, for example, 60 or 120 s compared to 15 or 30 s, is likely to make the RR more accurate due to the marked short-term, breath-by-breath variability. It is also critical when performing device comparison studies to use the exact same breaths. This requires a high degree of device time synchronization.

Healthy adults have mean RRs that range between 12 and 20 per minute, a 67% difference between lower and upper ‘normal’ values. Healthy neonates have a mean RR of 30 to 60 bpm, a 100% difference [41]. Some healthy neonates have an upper RR range as high as 72 bpm [42]. The substantial

neonatal RRV identified in these results has significant implications for the use of guidelines, setting clinical thresholds, and when comparing RR measurement devices. The United Nations International Children's Emergency Fund (UNICEF) recommended a maximum RRV of no more than two bpm in accuracy for diagnostic device comparisons of acute respiratory infection is not appropriate for neonates [44]. For a neonate with a RR of 70 bpm, this recommendation equates to 2.9% variability, somewhat stricter than any of the within-neonate or between-neonate RRV identified in our results. In an adult, a two bpm difference at ten bpm could be an important difference (20%), but a two bpm difference in a neonate typically breathing at 60 bpm (3.3%) would be less clinically relevant.

Furthermore, there is substantial RRV across individual neonates over time. The algorithm-derived breath counts identified changes in neonatal RRV_{10bd} and RRV_{60bd} and across all epochs. RRV was higher within RRV_{60bd} epochs

compared with RRV_{10bd} . Previous research looking at pediatric populations has suggested RRV could be used to diagnose sleep apnea–hypopnea syndrome [19]. In adults, within-subject RRV might be useful as a predictor of subsequent intensive care unit admission [18]. In neonates, RRV may be used to improve clinical care, but more research is needed.

A common dilemma encountered when performing studies of RR measurement is the definition of a ‘true’ breath [45]. Various methods have been used to estimate RR across devices that result in the measurement of different respiratory events as breaths. Small ineffectual respiratory efforts are commonly seen in neonates and may not be consistently accepted or rejected as breaths. Capnograms show many different patterns, including small-amplitude ineffective breaths, double breaths (two peaks in a single breath), subsequent breaths starting before the waveform reaches the trough or baseline value, pauses, catch-up rapid breathing, sighs, and sharp rapid cycles as seen with coughs (Fig. 1; Table 3). Further investigation into the identification and inclusion of abnormal respiratory events during respiratory device accuracy comparison studies are needed.

The clinical implication of this high degree of RRV, even in neonates with regular breathing, is that clinicians should be aware of the inherent uncertainty of clinical decisions made based on selected threshold values. It would be advisable to use repeated observations before making critical clinical decisions and ideally to use continuous monitoring devices and values summarized over more than one minute.

When conducting device comparison studies, accuracy thresholds should be adapted to a neonate’s baseline RR. Therefore, we suggest that an accuracy threshold should be normalized as a percentage of the baseline value and not more restrictive than the within-neonate RRV. This aligns with a previous proposal to use a percentage error threshold for LOA to determine the acceptability of a new technique in cardiac measurement and is also relevant when comparing RR measurement technologies [46]. When conducting RR accuracy testing, precise synchronization between investigational and reference devices will ensure that the same breaths are compared between devices. Measuring RR in a calm child, as recommended by WHO, will also minimize variability. However, following the WHO recommendations for RR measurement results in a rounding-down to the nearest breath and assumes the mean breathing rate is the most important clinical variable [47]. Instead, a median RR, unaffected by cough or pause, may be more reflective of the underlying physiological control of breathing and more clinically relevant than a mean breath count over 60 s.

We did not study the full range of RRV in real-world settings, particularly among critically ill neonates. The RRV identified in our study likely under-represents the true RRV present in neonates given the data selection used only

capnograms with easy-to-identify breaths in the manual breath count and algorithm-derived breath count processes. Epochs that were excluded from analysis were not evenly distributed across neonates and poor data quality was the most common reason for exclusion. Selecting good quality capnograms likely increased the observed agreement within the Bland–Altman analysis. Expansion of the data quality thresholds for data inclusion would likely result in wider CIs and increased RRV.

The capnogram CO_2 sampling rate was approximately 20 times per second or 20 Hz which is likely sufficient for an adult breathing at 10 to 20 breaths per minute. However, sampling frequency inaccuracies are more apparent at higher breath rates, such as those seen in distressed neonates (which may exceed 80 breaths per minute). When working with neonates or other populations expected to have high breath rates, sampling rates of 100 Hz, and even as high as > 200 Hz, are suggested [48]. These higher sampling rates would avoid any aliasing effects, enable oversampling to accommodate filtering to remove artifacts, and ensure precision in RRV_{bd} estimation.

In the current study, the resulting algorithm-derived breath count had closer agreement with the manual breath count than either the algorithm-derived mean or median breath rates. This finding highlights the impact that smoothing, averaging, normalizing, or other post-processing procedures may have on RR measurement. Most devices will provide a processed result rather than a count, so consideration as to the impact these post-processing decisions are critically important when evaluating automated devices. The critical question that remains as yet unresolved, is the clinical importance of count or the mean/median as a representation of disease severity? RR is often averaged across multiple breaths, resulting in additional uncertainty when making clinical decisions and when comparing devices.

These results identify the range and sources of RRV found between- and within-neonates, as well as between- and within-epochs. RR is traditionally measured by counting the number of breaths within 60 s. While manual counting may seem to be a practical clinical approach, it has limitations, especially when compared to RR measurement with digital devices. Large within-neonate RRV will also impact the application of RR thresholds in MCPM devices and their clinical applications. For devices that estimate RR, we propose a median value of inter-breath intervals within 60 s to remove any extreme outliers and to minimize the effect of rounding.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10877-022-00840-2>.

Acknowledgements We thank the dedicated staff at Aga Khan University-Nairobi Hospital for providing patient care, and the study participants, their caregivers, and the local community in Nairobi, Kenya,

for their participation. Additionally, we would like to acknowledge Prof. Guy A Dumont from UBC's Electrical and Computer Engineering department for his support with the terminology and concepts contained within the manuscript.

Author contributions ASG and JMA conceived of the study and ASG obtained the funding. WM, ASG, and JMA designed and provided oversight for the administration and implementation of the study. JC contributed to implementation of the study and DD wrote the application used to collect data from the reference device, WK wrote the application used to analyse some of the processed data. WM, RO and DC collected the data. JC, JMA, and DD verified the underlying data. JC and GZ performed data analysis and WK and GZ provided statistical analysis guidance. JC, ASG, and JMA wrote the draft of the manuscript with critical input from all co-authors. All authors edited and reviewed the final manuscript.

Funding This study was funded by Bill & Melinda Gates Foundation grant OPP1196617.

Data availability Data will be made available on completion of the secondary analyses of the data. We have ethics approval to deposit the data in an open-access repository upon completion of the study.

Code availability N/A.

Declarations

Conflict of interest All authors declare no competing interests.

Ethical approval The study was conducted per the International Conference on Harmonisation Good Clinical Practice and the Declaration of Helsinki 2008. The protocol and other relevant study documents were approved by Western Institutional Review Board (20191102; Puyallup, Washington, USA), Aga Khan University Nairobi Research Ethics Committee (2019/REC-02 v2; Nairobi, Kenya), Kenyan Pharmacy and Poisons Board (19/05/02/2019(078)), and Kenyan National Commission for Science, Technology and Innovation (NACOSTI/P/19/68024/30253).

Consent to participate Written informed consent was obtained in English or Swahili by trained study staff from each neonate's caregiver according to a checklist that included ascertainment of caregiver comprehension.

Consent for publication N/A

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Palafox M, Guiscafré H, Reyes H, Munoz O, Martínez H. Diagnostic value of tachypnoea in pneumonia defined radiologically. *Arch Dis Child*. 2000;82:41–5. <https://doi.org/10.1136/adc.82.1.41>.
2. Amirav I, Lavie M. Rethink respiratory rate for diagnosing childhood pneumonia. *EClinicalMedicine*. 2019;12:6–7. <https://doi.org/10.1016/j.eclinm.2019.06.013>.
3. Fiesemann JF, Hendryx MS, Helms CM, Wakefield DS. Respiratory rate predicts cardiopulmonary arrest for internal medicine inpatients. *J Gen Intern Med*. 1993;8:354–60. <https://doi.org/10.1007/BF02600071>.
4. World Health Organization. Integrated management of childhood illness: caring for newborns and children in the community; 2011.
5. Simoes EA, Roark R, Berman S, Esler LL, Murphy J. Respiratory rate: measurement of variability over time and accuracy at different counting periods. *Arch Dis Child*. 1991;66:1199–203. <https://doi.org/10.1136/adc.66.10.1199>.
6. Daw W. Medical devices for measuring respiratory rate in children: a review. *J Adv Biomed Eng Technol*. 2016. <https://doi.org/10.15379/2409-3394.2016.03.01.04>.
7. Walker JM, Farney RJ, Rhondeau SM, Boyle KM, Valentine K, Cloward TV, et al. Chronic opioid use is a risk factor for the development of central sleep apnea and ataxic breathing. *J Clin Sleep Med*. 2007;3:455–61.
8. Del Negro CA, Funk GD, Feldman JL. Breathing matters. *Nat Rev Neurosci*. 2018;19:351–67. <https://doi.org/10.1038/s41583-018-0003-6>.
9. Eichenwald EC, Committee on Fetus and Newborn, American Academy of Pediatrics. Apnea of prematurity. *Pediatrics*. 2016. <https://doi.org/10.1542/peds.2015-3757>.
10. Mohr MA, Fairchild KD, Patel M, Sinkin RA, Clark MT, Randall Moorman J, et al. Quantification of periodic breathing in premature infants. *Physiol Meas*. 2015;36:1415–27. <https://doi.org/10.1088/0967-3334/36/7/1415>.
11. Al-Hathlol K, Idiong N, Hussain A, Kwiatkowski K, Alvaro RE, Weintraub Z, et al. A study of breathing pattern and ventilation in newborn infants and adult subjects. *Acta Paediatr*. 2007;89:1420–5. <https://doi.org/10.1111/j.1651-2227.2000.tb02769.x>.
12. Tobin MJ, Mador MJ, Guenther SM, Lodato RF, Sackner MA. Variability of resting respiratory drive and timing in healthy subjects. *J Appl Physiol*. 1988;65:309–17. <https://doi.org/10.1152/jappl.1988.65.1.309>.
13. Marjanovic N, Mimoz O, Guenezan J. An easy and accurate respiratory rate monitor is necessary. *J Clin Monit Comput*. 2020;34:221–2. <https://doi.org/10.1007/s10877-019-00357-1>.
14. Saria S, Rajani AK, Gould J, Koller D, Penn AA. Integration of early physiological responses predicts later illness severity in preterm infants. *Sci Transl Med*. 2010. <https://doi.org/10.1126/scitranslmed.3001304>.
15. Warburton A, Monga R, Sampath V, Kumar N. Continuous pulse oximetry and respiratory rate trends predict short-term respiratory and growth outcomes in premature infants. *Pediatr Res*. 2019;85:494–501. <https://doi.org/10.1038/s41390-018-0269-4>.
16. Jain D, Bancalari E. Neonatal monitoring during delivery room emergencies. *Semin Fetal Neonatal Med*. 2019;24:101040. <https://doi.org/10.1016/j.siny.2019.101040>.
17. Chung HU, Rwei AY, Hourlier-Fargette A, Xu S, Lee K, Dunne EC, et al. Skin-interfaced biosensors for advanced wireless physiological monitoring in neonatal and pediatric intensive-care units. *Nat Med*. 2020;26:418–29. <https://doi.org/10.1038/s41591-020-0792-9>.
18. Garrido D, Assioun JJ, Keshishyan A, Sanchez-Gonzalez MA, Goubran B. Respiratory rate variability as a prognostic factor in

- hospitalized patients transferred to the intensive care unit. *Cureus*. 2018;10:e2100. <https://doi.org/10.7759/cureus.2100>.
19. Barroso-Garcia V, Gutierrez-Tobal GC, Kheirandish-Gozal L, Alvarez D, Vaquerizo-Villar F, Del Campo F, et al. Usefulness of spectral analysis of respiratory rate variability to help in pediatric sleep apnea-hypopnea syndrome diagnosis. *Conf Proc IEEE Eng Med Biol Soc*. 2019;2019:4580–3. <https://doi.org/10.1109/EMBC.2019.8857719>.
 20. Tsuji H, Venditti FJ Jr, Manders ES, Evans JC, Larson MG, Feldman CL, et al. Reduced heart rate variability and mortality risk in an elderly cohort. The Framingham Heart Study. *Circulation*. 1994;90:878–83. <https://doi.org/10.1161/01.cir.90.2.878>.
 21. Dekker JM, Crow RS, Folsom AR, Hannan PJ, Liao D, Swenne CA, et al. Low heart rate variability in a 2-minute rhythm strip predicts risk of coronary heart disease and mortality from several causes. *Circulation*. 2000;102:1239–44. <https://doi.org/10.1161/01.CIR.102.11.1239>.
 22. Fairchild K, Aschner. HeRO monitoring to reduce mortality in NICU patients. RRN. 2012. <https://doi.org/10.2147/RRN.S32570>.
 23. Malik M, Bigger JT, Camm AJ, Kleiger RE, Malliani A, Moss AJ, et al. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *Eur Heart J*. 1996;17:354–81.
 24. Cygankiewicz I, Zareba W. Heart rate variability. *Handb Clin Neurol*. 2013;117:379–93. <https://doi.org/10.1016/B978-0-444-53491-0.00031-6>.
 25. Fairchild KD, Schelonka RL, Kaufman DA, Carlo WA, Kattwinkel J, Porcelli PJ, et al. Septicemia mortality reduction in neonates in a heart rate characteristics monitoring trial. *Pediatr Res*. 2013;74:570–5. <https://doi.org/10.1038/pr.2013.136>.
 26. Kumar N, Akangire G, Sullivan B, Fairchild K, Sampath V. Continuous vital sign analysis for predicting and preventing neonatal diseases in the twenty-first century: big data to the forefront. *Pediatr Res*. 2020;87:210–20. <https://doi.org/10.1038/s41390-019-0527-0>.
 27. Massaroni C, Nicolò A, Lo Presti D, Sacchetti M, Silvestri S, Schena E. Contact-based methods for measuring respiratory rate. *Sensors*. 2019;19:908. <https://doi.org/10.3390/s19040908>.
 28. Al-Halhouli A, Al-Ghussain L, El Bouri S, Liu H, Zheng D. Fabrication and evaluation of a novel non-invasive stretchable and wearable respiratory rate sensor based on silver nanoparticles using inkjet printing technology. *Polymers*. 2019;11:1518. <https://doi.org/10.3390/polym11091518>.
 29. Hurtado DE, Abusleme A, Chávez JAP. Non-invasive continuous respiratory monitoring using temperature-based sensors. *J Clin Monit Comput*. 2020;34:223–31. <https://doi.org/10.1007/s10877-019-00329-5>.
 30. Rossol SL, Yang JK, Toney-Noland C, Bergin J, Basavaraju C, Kumar P, et al. Non-contact video-based neonatal respiratory monitoring. *Children*. 2020;7:171. <https://doi.org/10.3390/children7100171>.
 31. Joshi R, Bierling B, Feijs L, van Pul C, Andriessen P. Monitoring the respiratory rate of preterm infants using an ultrathin film sensor embedded in the bedding: a comparative feasibility study. *Physiol Meas*. 2019;40:045003. <https://doi.org/10.1088/1361-6579/ab1595>.
 32. Vanegas E, Igual R, Plaza I. Sensing systems for respiration monitoring: a technical systematic review. *Sensors*. 2020;20:5446. <https://doi.org/10.3390/s20185446>.
 33. Goldsack JC, Coravos A, Bakker JP, Bent B, Dowling AV, Fitzer-Attas C, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). *NPJ Digit Med*. 2020;3:55. <https://doi.org/10.1038/s41746-020-0260-4>.
 34. Godfrey A, Goldsack JC, Tenaerts P, Coravos A, Aranda C, Husain A, et al. BioMeT and algorithm challenges: a proposed digital standardized evaluation framework. *IEEE J Transl Eng Health Med*. 2020;8:0700108. <https://doi.org/10.1109/JTEHM.2020.2996761>.
 35. Ginsburg AS, Nkwopara E, Macharia W, Ochieng R, Waiyego M, Zhou G, et al. Evaluation of non-invasive continuous physiological monitoring devices for neonates in Nairobi, Kenya: a research protocol. *BMJ Open*. 2020;10:e035184. <https://doi.org/10.1136/bmjopen-2019-035184>.
 36. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42:377–81. <https://doi.org/10.1016/j.jbi.2008.08.010>.
 37. Karlen W, Ansermino JM, Dumont G. Adaptive pulse segmentation and artifact detection in photoplethysmography for mobile applications. *Conf Proc IEEE Eng Med Biol Soc*. 2012;2012:3131–4. <https://doi.org/10.1109/EMBC.2012.6346628>.
 38. StataCorp LP. Stata statistical software: Release 13. Statacorp lp College Station, TX; 2013.
 39. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; <https://www.r-project.org/>
 40. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8:135–60. <https://doi.org/10.1177/096228029900800204>.
 41. Sepanski RJ, Godambe SA, Zaritsky AL. Pediatric vital sign distribution derived from a multi-centered emergency department database. *Front Pediatr*. 2018;6:66. <https://doi.org/10.3389/fped.2018.00066>.
 42. Tveiten L, Diep LM, Halvorsen T, Markestad T. Respiratory rate during the first 24 hours of life in healthy term infants. *Pediatrics*. 2016. <https://doi.org/10.1542/peds.2015-2326>.
 43. Bonafide CP, Brady PW, Keren R, Conway PH, Marsolo K, Daymont C. Development of heart and respiratory rate percentile curves for hospitalized children. *Pediatrics*. 2013;131:e1150–7. <https://doi.org/10.1542/peds.2012-2443>.
 44. UNICEF Supply Division. Target Product Profile: Acute Respiratory Infection Diagnostic Aid (ARIDA). 2014 Nov. [https://www.unicef.org/videoaudio/PDFs/ARIDA_-_Target_Product_Profile_\(2\).pdf](https://www.unicef.org/videoaudio/PDFs/ARIDA_-_Target_Product_Profile_(2).pdf)
 45. Jaffe MB. What is a “valid” breath?-Methodological issues. *Anesth Analg*. 2011;113:5.
 46. Critchley LA, Critchley JA. A meta-analysis of studies using bias and precision statistics to compare cardiac output measurement techniques. *J Clin Monit Comput*. 1999;15:85–91. <https://doi.org/10.1023/a:1009982611386>.
 47. World Health Organization. Pocket book of hospital care for children: guidelines for the management of common childhood illnesses. Geneva: World Health Organization; 2013.
 48. Schmalisch G. Current methodological and technical limitations of time and volumetric capnography in newborns. *Biomed Eng Online*. 2016;15:104. <https://doi.org/10.1186/s12938-016-0228-4>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.