Clinical feasibility of an advanced neonatal epidermal multiparameter continuous monitoring technology in a large public maternity hospital in Nairobi, Kenya

Amy Sarah Ginsburg1, Sahar Zandi Nia2, Dorothy Chomba3, Millicent Parsimei3, Dustin Dunsmuir2, Mary Waiyego4, Jesse Coleman5, Roseline Ochieng1, Guohai Zhou6, William M. Macharia3 & J. Mark Ansermino2

Clinically feasible multiparameter continuous physiological monitoring technologies are needed for use in resource-constrained African healthcare facilities to allow for early detection of critical events and timely intervention for major morbidities in high-risk neonates. We conducted a prospective clinical feasibility study of a novel multiparameter continuous physiological monitoring technology in neonates at Pumwani Maternity Hospital in Nairobi, Kenya. To assess feasibility, we compared the performance of Sibel’s Advanced Neonatal Epidermal (ANNE) technology to reference technologies, including Masimo’s Rad-97 pulse CO-oximeter with capnography technology for heart rate (HR), respiratory rate (RR), and oxygen saturation (SpO2) measurements and Spengler’s Tempo Easy non-contact infrared thermometer for temperature measurements. We evaluated key performance criteria such as up-time, clinical event detection performance, and the agreement of measurements compared to those from the reference technologies in an uncontrolled, real-world setting. Between September 15 and December 15, 2020, we collected and analyzed 503 h of ANNE data from 109 enrolled neonates. ANNE’s up-time was 42 (11%) h more for HR, 77 (25%) h more for RR, and 6 (2%) h less for SpO2 compared to the Rad-97. However, ANNE’s ratio of up-time to total attached time was less than Rad-97’s for HR (0.79 vs 0.86), RR (0.68 vs. 0.79), and SpO2 (0.69 vs 0.86). ANNE demonstrated adequate performance in identifying high and low HR and RR and high temperature events; however, showed relatively poor performance for low SpO2 events. The normalized spread of limits of agreement were 8.4% for HR and 52.2% for RR and the normalized root-mean-square deviation was 4.4% for SpO2. Temperature agreement showed a spread of limits of agreement of 2.8 °C. The a priori-identified optimal limits were met for HR and temperature but not for RR or SpO2. ANNE was clinically feasible for HR and temperature but not RR and SpO2 as demonstrated by the technology’s up-time, clinical event detection performance, and the agreement of measurements compared to those from the reference technologies.

In high-income countries, multiparameter continuous monitoring technologies can be essential in the clinical management of high-risk neonates. However, these technologies are frequently unavailable in resource-constrained African settings despite higher neonatal morbidity and mortality1,2. At a large, public, referral hospital in Nairobi, few neonates had vital signs recorded within the first hour of life, and less than half received heart rate...
low HR and RR events, low SpO2 events, and high temperature events. In addition, we evaluated the agreement between the ANNE and reference technologies’ total time attached and up-time, and event detection of high and ensure study participation did not interfere with or unnecessarily delay clinical care of neonates.

Sensor were compared to Tempo Easy’s skin temperature readings from the forehead. Data collected by ANNE and the reference technologies (Table 1). For temperature, readings from ANNE’s chest graph quality index (PO-SQI) data at 62.5 Hz (Hz), and capnography (carbon dioxide (CO2)) waveform data. Data were parsed in C (Dennis Ritchie & Bell Labs, USA) to obtain plethysmograph waveform and plethysmography (Supplementary Fig. S1). Up to 30 h of data can be stored within the sensor and transmitted wirelessly to a central database supported by customized software. The projected cost of commercial acquisition of Sibel’s reusable and rechargeable technology is about $40USD with an additional $0.20USD per unit for adhesives/ consumables lasting 24 to 72 h. In a previous clinical trial of neonates in Kenya, we evaluated the accuracy of ANNE to measure HR, RR, SpO2, and temperature when compared to verified reference technologies. We also completed qualitative assessments of the feasibility, usability, and acceptability among healthcare personnel and caregivers by conducting in-depth interviews and observations.

Novel medical technologies may compare favorably to established reference technologies in more controlled research settings; however, also important to ascertain is their clinical feasibility in uncontrolled, real-world settings. If clinical feasibility performance is not evaluated and the findings incorporated during technology development and refinement, a novel medical technology’s adoption, uptake, scale-up, and use in clinical practice may be impacted. Thus, in this study we evaluated key performance criteria such as up-time (periods of adequate signal quality), clinical event detection performance, and the agreement of measurements compared to those from the reference technologies in an uncontrolled, real-world setting.

Methods

Study design and participants. We conducted a prospective, observational, clinical feasibility study of ANNE at Pumwani Maternity Hospital (PMH) in Nairobi, Kenya, the largest referral public maternity hospital in sub-Saharan Africa. PMH has no neonatal intensive care unit. Caregivers of neonates delivered at or admitted to PMH were approached by trained study staff who obtained written informed consent and assessed the neonate for study eligibility based on the results of the medical history, clinical examination, and appropriate understanding of the study by the caregiver (Table 1). Carried out in accordance with the Declaration of Helsinki and Guideline for Good Clinical Practice/ International Standards Organization (ISO) 14155 to ensure accurate, reliable, and consistent data collection, the study protocol was approved by Western Institutional Review Board (20191102), Aga Khan University Nairobi Research Ethics Committee (2019/REC-02), and Kenya Pharmacy and Poisons Board (EUCTT/19/05/02) and registered with ClinicalTrials.gov, NCT03920761. Effort was made to ensure study participation did not interfere with or unnecessarily delay clinical care of neonates.

Study procedures. Enrolled neonates received local standard of care while additionally being simultaneously monitored with the ANNE multiparameter, continuous monitoring technology (Sibel Inc. IL, USA), the Rad-97 pulse CO-oximeter with capnography (Masimo Corporation, USA), and Spengler’s Tempo Easy non-contact infrared thermometer (SPENGLER HOLTEX Group, Aix-en-Provence, France). The Rad-97 was selected as a reference technology based on its ability to extract and record high resolution data, perform neonatal capnography and pulse oximetry, and its compact design enabling bedside monitoring. Tempo Easy was chosen as a reference technology for temperature monitoring due to its in-country availability. HR, RR, and SpO2 data were collected in real-time for a minimum of one hour from ANNE and Rad-97 and temperature data via hourly spot checks with the Tempo Easy technology. The total opportunity for attaching the technologies was similar. Up-time, clinical event detection performance, and HR, RR, SpO2, and temperature measurements data were collected by ANNE and the reference technologies (Table 1). For temperature, readings from ANNE’s chest sensor were compared to Tempo Easy’s skin temperature readings from the forehead.

Outcomes. To ensure objective measures of clinical feasibility, study outcomes included comparisons between the ANNE and reference technologies’ total time attached and up-time, and event detection of high and low HR and RR events, low SpO2 events, and high temperature events. In addition, we evaluated the agreement between HR, RR, SpO2, and temperature measurements in a real-world setting (Table 1).

Data processing and analysis. The total number of minutes the sensors were attached and the up-time for each technology were calculated. We assessed the quality of the measurements from the ANNE and reference technologies, and designated periods of adequate signal quality data as up-time (Table 1). For ANNE, we obtained HR, RR, and SpO2 measurements with their corresponding proprietary signal quality index every second. Raw data collected in real-time was retrieved from the Rad-97 with a custom Android application. Data were parsed in C (Dennis Ritchie & Bell Labs, USA) to obtain plethysmograph waveform and plethysmography quality index (PO-SQI) data at 62.5 Hz (Hz), and capnography (carbon dioxide (CO2)) waveform data at approximately 20 Hz. CO2 waveform data were analyzed using a breath detection algorithm developed in MATLAB (Math Works, USA) based on adaptive pulse segmentation. We obtained HR and RR from intra-beat and breath duration intervals, respectively. A custom algorithm based on capnography features was utilized to determine the capnography quality index (CO2-SQI). SpO2 values were calculated by the Rad-97 at 1 Hz, and 8-s medians were used in the analysis. To standardize event detection, upper limits for HR and RR were individualized for each neonate and were calculated for HR to be 20% and for RR to be 15% greater than their respective
Eligibility criteria

| Inclusion | Neonate with corrected age of < 28 days requiring admission to the neonatal ward, the post-natal ward, or the neonatal high dependency unit at Pumwani Maternity Hospital for prematurity or other clinical indication(s) based on the attending physician’s assessment. Caregiver(s) willing and able to provide informed consent and available for follow-up for the duration of the study. |
| Exclusion | Receiving continuous positive airway pressure or mechanical ventilation. Skin abnormalities in the nasopharynx and/or oropharynx. Contraindication to skin sensor application. Known arrhythmia. Congenital abnormality requiring major surgical intervention. Any medical or psychosocial condition or circumstance that would interfere with study conduct or for which study participation could put the neonate’s health at risk. |

Study endpoints

Up-time duration of ANNE compared to the reference technologies

Diagnostic performance of ANNE compared to the reference technologies for clinical event detection including sensitivity, specificity, positive predictive value, negative predictive value, and ratio of false negative-to-false positive events

Agreement between ANNE and the reference technologies for heart rate (HR), respiratory rate (RR), oxygen saturation (SpO₂), and temperature

Study definitions

| Total time attached | Measured in minutes as non-zero values recorded by the technology starting 10 min after technology placement and 5 min before disconnection; the 5-min periods before temporary removal and after reconnection to ANNE and the reference technologies were also excluded. |
| Up-time | Measured in minutes as the total time the sensor was attached that met the a priori-identified signal quality limits for each technology. |
| Signal quality | HR and SpO₂ | ANNE—for every second, we evaluated the preceding 59 s in addition to the current second to ensure that all 60 (100%) seconds > 0. Rad-97—for every second, we evaluated the preceding 59 s in addition to the current second to ensure that at least 30 (50%) seconds demonstrated a signal quality index (Masimo SQI) > 150 |
| | RR | ANNE—for every second, we evaluated the preceding 59 s in addition to the current second to ensure that all 60 (100%) seconds > 0. Rad-97—for every second, we evaluated the preceding 59 s in addition to the current second to ensure that at least 30 (50%) seconds demonstrated no capnography exceptions, indicating low RR quality, and a capnography quality score ≥ 2. |
| | Temperature | ANNE—all temperatures > 0. Spengler’s technology—all temperature spot checks. |
| Event second | Any second that contains a high or low HR or RR event (a value above or below the thresholds) for either ANNE or Rad-97. |
| Event window | A 10-min window centered from 5 min before to 5 min after the first event second noted by the reference technology; no overlapping windows are allowed. So, event seconds less than 5 min from the end of the previous event window result in the new event window starting immediately following the previous window. |
| True positive event | A reference technology event window containing at least 1 event second identified by ANNE. |
| False negative event | A reference technology event window containing no event seconds recorded by ANNE. |
| False positive event | An event recorded by ANNE outside all reference technology’s event windows. |
| True negative event | Any 10-min window with no events recorded by either ANNE or the reference technology. |
| Clinically significant event | Any false negative or false positive event that would likely require a clinician to institute a change in clinical practice. |

Table 1. Study eligibility criteria, endpoints, and definitions.

baseline values (based on a review of historical data) once the neonate was settled (approximately 15 min after monitoring was started) but no less than 140 beats/minute for HR and 40 breaths/minute for RR. For all neonates, an upper limit temperature of 37.5 °C was used. Lower limits were static with values of 80 beats/minute for HR, 15 breaths/minute for RR, and 90% for SpO₂ based on the normal physiological range. To identify clinical events, Sibel provided a software parser that processed all of their monitoring data and provided event detection in pseudo- real-time that was finalized prior to data collection. Reference technology data were processed following data collection with a custom algorithm to identify events.

For overlapping periods of up-time from ANNE and Rad-97, clinical events were detected by examining the previous minute of data for both technologies. An event was identified if the one-minute median and the most recent ten-second median both met adequate signal quality and exceeded the threshold for upper or lower alarm values for HR, RR, SpO₂, or temperature. Using a custom algorithm, events were aggregated from the
Event adjudication. Pre- and post-adjudication, there were minimal changes in HR and RR performance parameters with the exception of PPV which improved by 46% post-adjudication for high HR (Table 3). For SpO2, a substantial improvement ranging from 9% (PPV) to 49% (sensitivity) was observed across all performance parameters between pre- and post-adjudication of events. Of the 1140 false negative and false positive events.
events, 3 high HR events, 24 high RR events, and 87 low SpO2 events were interpreted as clinically significant during adjudication.

**Agreement.** HR agreement showed a normalized spread of LOA of 8.4% (Fig. 3a) and RR agreement showed a normalized spread of LOA of 52.2% (Fig. 3b). The a priori-identified limit of 30% for the normalized spread of LOA was met for HR but not RR. SpO2 agreement showed a normalized RMSD of 4.4% (Fig. 3c); the adequate a priori-identified limit of 4.5% for the RMSD was met for SpO2, but the optimal limit of 3.5% was not met. A sensitivity analysis removing cases where there were at least 10 missed SpO2 events revealed a minimal reduction in the RMSD from 4.4 to 4.2% (Supplementary Figure S5). Temperature agreement showed a spread of LOA of 2.8 °C (Fig. 3d) and therefore, met the a priori-identified limit of 4.5 °C.
Discussion

In an evaluation in a large public maternity hospital in Nairobi, Kenya, we found Sibel's ANNE to be clinically feasible for some measurement parameters (HR and temperature) but not others (RR and SpO2) as demonstrated by the technology’s up-time, clinical event detection performance, and the agreement of measurements compared to those from the reference technologies. Compared to Rad-97’s wired and more invasive technology, wireless ANNE’s up-time was longer because ANNE could continue to monitor the neonate during feeding and kangaroo mother care. However, while HR and temperature event detection were optimal and RR event detection was acceptable, SpO2 event detection was not acceptable. Adjudication established that ANNE incorrectly detected RR and SpO2 artifacts as clinically significant events during periods of noise, yielding many false positive high RR and low SpO2 events. Agreement of neonatal measurements between ANNE and the verified reference technologies showed that HR and temperature measurements met the a priori-identified optimal limits, RR measurements did not, and SpO2 measurements met the adequate but not the optimal a priori-identified limits in both the original and sensitivity analyses.

Compared to the findings from a more controlled clinical trial at a better-resourced hospital with a neonatal intensive care unit, there was significant degradation in RR and SpO2 agreement in this real-world clinical environment, limiting the clinical feasibility of ANNE to assess these measurements in routine practice7. RR agreement degradation may have been due to imprecise sensor placement and/or inadequate skin adherence. The low SpO2 missed events were concentrated among a small number of neonates, most likely due to motion artifacts, low perfusion states, and/or inadequate application of the ANNE sensors in individual neonates.

Reliable SpO2 measurement and monitoring can be challenging in smaller neonates and those with darker skin pigmentation. Motion artifacts among neonates induce many false alarms, which in turn can cause additional workload for healthcare providers and stress for the neonates and their caregivers as well as potentially impair response times in real critical situations16–18. Low perfusion states and variations in breathing can be common in neonates and can degrade pulse oximeter performance19–21. Smaller neonatal finger size and a correspondingly smaller pulsatile signal detected by the pulse oximeter sensor may also compromise performance22–24. As melanin is a secondary absorber of near-infrared light, darker skin pigmentation may impact pulse oximeter performance and accuracy25,26.

In two large cohorts of adults, Black patients had almost three times the frequency of occult hypoxemia undetected by pulse oximetry compared to White patients27. However, in a small infant study, no evidence of systematic bias in pulse oximetry measurement based on skin pigmentation was found in 36 patients with hypoxemia28. In another study of 294 neonates less than 32 weeks gestation, SpO2 overestimation measured by mean bias was 2.4-fold greater among Black preterm neonates and resulted in greater occult hypoxemia26. Validation studies of new medical technologies need to include diverse populations29.

Previous studies have evaluated the accuracy of single parameter monitoring technologies under controlled experimental conditions or have performed qualitative feasibility assessments30,31. In our study, we utilized a comprehensive and integrated approach that included real-world accuracy evaluations combined with performance metrics such as up-time and event detection that are often important in critical clinical situations. Solely focusing on the accuracy may not appropriately reflect the poor performance in event detection in individual cases. Regulatory approval typically requires controlled experiments looking at agreement or accuracy but does not consider real-world applications or event detection. Performing well in a controlled accuracy evaluation, while important, does not necessarily ensure a medical technology will perform well in clinical practice and have real-world impact.
Figure 3. Bland–Altman plots of measured (a) heart rate (HR), (b) respiratory rate (RR), (c) oxygen saturation (SpO₂), and (d) temperature as measured by ANNE and the reference technologies. Colors indicate which participant neonate is associated with the measurement pair.
Figure 3. (continued)
Limitations to our evaluation, analyses, and adjudication included the relatively arbitrary a priori signal quality limits we selected to mitigate confusion that could have arisen if poor quality signals were used for event detection. To ensure comparable event detection, we evaluated the technologies in real-time even though we analyzed the data later. Limiting the generalizability of our evaluation was a relatively healthy study population with few critically ill neonates included and infrequent life-threatening events recorded. While comparing ANNE’s electrocardiogram-derived HR to Rad-97’s photoplethysmogram-derived HR may contribute to increased uncertainty in these comparisons, HR estimation has been shown to be well preserved from the photoplethysmogram. We limited our comparison to the detection of low RR and did not evaluate the detection of apnea since ANNE did not have an apnea algorithm. Overall, clinically significant events were relatively uncommon despite the sustained duration of monitoring, and we did not permit ANNE or the reference technologies to generate alarms that may have impacted clinical outcomes. ANNE would benefit from being further evaluated in different populations (e.g., preterm and more critically ill neonates) and settings with assessment of clinical outcomes and/or impact. To be clinically feasible, safe, and effective in improving quality of neonatal care, a monitoring technology needs to avoid excessive alarms (and alarm fatigue) resulting from false positive events and to suppress transient artifacts without missing clinically significant true positive events. Typically, this could be accomplished by employing certain false alarm limiting strategies such as an appropriate delay before alarm generation. Exploring threshold and adaptive alerts provided by the technology, assessing barriers and facilitators to adoption and use, and conducting costing studies will be key to ANNE’s development and if successful, to its uptake and scale-up.

There were additional technology constraints. ANNE experienced intermittent difficulty turning on after more than one week without use and a software bug resulted in some corruption of data and data loss. The wireless Bluetooth technology required daily charging of batteries, with regular checks to mitigate against data loss. Furthermore, the battery life degraded over the study period. Another potential limitation of wireless technology may include electrical interference from other devices in the vicinity; however, this was not experienced in this study. For temperature measurement, both ANNE and Tempo Easy’s technology measured skin surface temperature, which is dependent on the environmental room temperature, and thus, may not be accurate in detecting hypothermia. The measurement of core temperature is critical in neonates due to their large surface area which accelerates heat loss. Hypothermia is a common presenting sign for other severe physiological disturbances such as sepsis. For this reason, we did not set an event threshold for low temperature, but we acknowledge that monitoring core body temperature in neonates may be critically important. The Rad-97 also had constraints, including that it was relatively heavy with cumbersome sensory cables and cords, and prone to overheating when the air vent was covered. Furthermore, its sensors were not reusable and the nasal cannula was more frequently removed, limiting the technology’s up-time.

Despite the limitations and technology constraints, the study results from this and a previous accuracy evaluation in Kenya indicate that ANNE appears to be clinically feasible for wirelessly monitoring HR and temperature but not RR and SpO2 when compared to the more invasive Rad-97 reference technology. Studies of monitoring technologies in neonates have largely been limited to high-income countries and would benefit from diversification of study populations. Established methods for continuous monitoring exist, but factors such as invasiveness, time-consuming application, and high cost have contributed to feasibility concerns in resource-constrained settings. Novel neonatal multiparameter continuous monitoring technologies such as ANNE may be able to address these concerns; however, more work needs to be done regarding ANNE’s RR and SpO2 measurements.

Data availability
Access to data will be provided to researchers subject to submission of a research proposal and signing a Data Use Agreement. Interested researchers can request access to the data at https://doi.org/10.25934/PR00007550 by creating a Vivli account and submitting a request. Data will be available from September 2022.

Received: 25 January 2022; Accepted: 4 July 2022
Published online: 09 July 2022

References
1. United Nations Inter-agency Group for Child Mortality Estimation. Levels & Trends in Child Mortality: Report 2019, Estimates Developed by the United Nations Inter-agency Group for Child Mortality Estimation (United Nations Fund, 2019).
2. World Health Organization. Newborns: Improving Survival and Well-Being. https://www.who.int/news-room/fact-sheets/detail/newborns-reducing-mortality. Accessed 22 January 2021 (2020).
3. Mitchell, E. J. et al. Feasibility of using an Early Warning Score for preterm or low birthweight infants in a low-resource setting: Results of a mixed-methods study at a national referral hospital in Kenya. BMJ Open 10, e039061 (2020).
4. Gathara, D. et al. Missed nursing care in newborns: A cross-sectional direct observational study. BMJ Qual. Saf. 29, 19–30 (2020).
5. Zhu, Z., Liu, T., Li, G., Li, T. & Inoue, Y. Wearable sensor systems for infants. Sensors (Basel) 15, 3721–3749 (2015).
6. Sahni, R. Continuous noninvasive monitoring in the neonatal ICU. Curr. Opin. Pediatr. 29, 141–148 (2017).
7. Coleman, J. et al. Evaluation of Sibelfs advanced neonatal epidermal (ANNE) wireless continuous physiological monitor in Nairobi, Kenya. PLoS One (2022) (accepted).
8. Ginsburg, A. S. et al. Multiparameter continuous physiological monitoring technologies in neonates among health care providers and caregivers at a Private Tertiary Hospital in Nairobi, Kenya: Feasibility, usability, and acceptability study. J. Med. Internet Res. 23, e29755 (2021).
9. Kinsella, M. W. et al. Qualitative study exploring the feasibility, usability and acceptability of neonatal continuous monitoring technologies at a public tertiary hospital in Nairobi, Kenya. BMJ Open 12, e053486 (2022).
10. Ginsburg, A. S. et al. Evaluation of non-invasive continuous physiological monitoring devices for neonates in Nairobi, Kenya: A research protocol. BMJ Open 10, e035164 (2020).
Karel, W., Ansermino, J. M. & Dumont, G. Adaptive pulse segmentation and artifact detection in photoplethysmography for mobile applications. *Ann. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2012, 3131–3134 (2012).

Ginsburg, A. S. et al. Clinical feasibility of a contactless multiparameter continuous monitoring technology for neonates in a large public maternity hospital in Nairobi, Kenya. *Sci. Rep.* 12, 3097 (2022).

Bland, J. M. & Altman, D. G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1, 307–310 (1986).

Coleman, J. et al. Identification of thresholds for accuracy comparisons of heart rate and respiratory rate in neonates. *Gates Open Res.* 5, 93 (2021).

ISO 80601-2-61. (International Organization for Standardization (ISO) Central Secretariat, 2017).

Sendelbach, S. & Funk, M. Alarm fatigue: A patient safety concern. *AANCA Adv. Crit. Care* 24, 378–386 (2013) ([quiz 387–378]).

Johnson, K. R., Hagadorn, J. I. & Sink, D. W. Alarm safety and alarm fatigue. *Clin. Perinatol.* 44, 713–728 (2017).

Li, T. et al. Epidemiology of patient monitoring alarms in the neonatal intensive care unit. *J. Perinatol.* 38, 1030–1038 (2018).

Poets, C. F. & Souttell, D. P. Noninvasive monitoring of oxygenation in infants and children: Practical considerations and areas of concern. *Pediatrics* 93, 737–746 (1994).

Fouzas, S., Priftis, K. N. & Antracopoulus, M. B. Pulse oximetry in pediatric practice. *Pediatrics* 128, 740–752 (2011).

Okunlola, O., et al. Pulse oximeter performance, racial inequity, and the work ahead. *Respir. Care* 67(2), 252–257 (2022).

Barker, S. J. & Tremper, K. K. Pulse oximetry: Applications and limitations. *Int. Anesthesiol. Clin.* 25, 155–175 (1987).

Wukitsch, M. W., Pettersson, M. T., Tobler, D. R. & Pologe, J. A. Pulse oximetry: Analysis of theory, technology, and practice. *J. Clin. Monit. 4*, 290–301 (1988).

Kelleher, J. F. & Ruff, R. H. The penumbra effect: Vasomotion-dependent pulse oximeter artifact due to probe malposition. *Anesthesiology* 71, 787–791 (1989).

Norton, H. L. Variation in pulse oximetry readings: Melanin, not ethnicity, is the appropriate variable to use when investigating bias. *Anesthesia* 77(3), 354–355 (2022).

Vesoulis, Z., Tims, A., Lodhi, H., Lados, N. & Whitehead, H. Racial discrepancy in pulse oximeter accuracy in preterm infants. *J. Perinatol.* 42(1), 79–85 (2022).

Sjöding, M. W., Dickson, R. P., Iwashyna, T. J., Gay, S. E. & Valley, T. S. Racial bias in pulse oximetry measurement. *N. Engl. J. Med.* 383, 2477–2478 (2020).

Foglia, E. E. et al. The effect of skin pigmentation on the accuracy of pulse oximetry in infants with hypoxemia. *J. Pediatr.* 182, 375–377.e372 (2017).

Knight, M. J., Subbe, C. P. & Inada-Kim, M. Racial discrepancies in oximetry: Where do we stand? *Anaesthesia* 77, 129–131 (2022).

Browne, S.H., Bernstein, M. & Bickler, P.E. Accuracy of samsung smartphone integrated pulse oximetry meets full FDA clearance standards for clinical use. *medRxiv* [https://doi.org/10.1101/2021.02.17.21249739] (2021).

Huhn, S. et al. The impact of wearable technologies in health research: Scoping review. *JMIR Mhealth Uhealth* 10, e34384 (2022).

Deckordi, P. et al. Pulse rate variability compared with heart rate variability in children with and without sleep disordered breathing. *Ann. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2013, 6563–6566 (2013).

Acknowledgements

We thank Steve Xu, Brandon Kim, Jong Yoon Lee, and Dennis Ryu at Sibel for their support in accessing the data from ANNE. We are grateful for the many hours of time provided by Nikita Jethwa, Ikol Kasede, and Annette Metho for their expert opinions and adjudication of clinical events. We also thank the dedicated staff at Pumwani Maternity Hospital for providing patient care, and the neonatal participants, their caregivers, and the local community in Nairobi, Kenya, for their participation and support.

Author contributions

A.S.G. and J.M.A. conceived of the study and A.S.G. obtained the funding. A.S.G., W.M.M., and J.M.A. designed and provided oversight for the administration and implementation of the study. D.D. wrote the application used to collect data from the reference device and processed data. D.C., M.P., M.W., R.O., and W.M.M. collected the data and J.C. supported implementation of the study. S.Z., D.D., and J.M.A. verified the underlying data, and S.Z. and D.D. performed data analysis and visualization. G.Z. provided statistical analysis guidance. A.S.G., S.Z., and J.M.A. wrote the draft of the manuscript with critical input from the co-authors. All authors edited and reviewed the final manuscript.

Funding

This study was funded by Bill & Melinda Gates Foundation (Grant No. OPP1203136).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-16051-3.

Correspondence and requests for materials should be addressed to A.S.G.

Reprints and permissions information is available at www.nature.com/reprints.

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