

# BMJ Paediatrics Open

BMJ Paediatrics Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Paediatrics Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjpaedsopen.bmj.com>).

If you have any questions on BMJ Paediatrics Open's open peer review process please email [info.bmjpo@bmj.com](mailto:info.bmjpo@bmj.com)

# BMJ Paediatrics Open

## Multi-center paired non-inferiority study of the cardiorespiratory monitoring performance of the wireless and non-adhesive Bambi® belt measuring diaphragm activity in neonates: Study Protocol

Journal:	<i>BMJ Paediatrics Open</i>
Manuscript ID	bmjpo-2022-001430
Article Type:	Protocol
Date Submitted by the Author:	27-Jan-2022
Complete List of Authors:	Scholten, Anouk; Amsterdam UMC Locatie AMC, Neonatology; Amsterdam UMC Locatie AMC, Amsterdam Reproduction & Development research institute Zhuozhao, Zhan; Eindhoven University of Technology, Department of Mathematics and Computer Science Niemarkt, Hendrik; Máxima Medisch Centrum, Neonatology Vervoorn, Marieke; Máxima Medisch Centrum, Neonatology van Leuteren, Ruud; Amsterdam UMC Locatie AMC, Neonatology; Amsterdam UMC Locatie AMC, Amsterdam Reproduction & Development research institute de Jongh, Frans; Amsterdam UMC Locatie AMC, Neonatology; University of Twente, Faculty of Science and Technology van Kaam, Anton; Amsterdam UMC Locatie AMC, Neonatology; Amsterdam UMC Locatie VUmc, Neonatology van den Heuvel, Edwin; Eindhoven University of Technology, Department of Mathematics and Computer Science Hutten, G. Jeroen; Amsterdam UMC Locatie AMC, Neonatology; Amsterdam UMC Locatie VUmc, Neonatology
Keywords:	Neonatology, Technology

SCHOLARONE™  
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1  
2  
3 **MULTI-CENTER PAIRED NON-INFERIORITY STUDY OF THE**  
4  
5 **CARDIORESPIRATORY MONITORING PERFORMANCE OF THE WIRELESS**  
6  
7 **AND NON-ADHESIVE BAMBI® BELT MEASURING DIAPHRAGM ACTIVITY IN**  
8  
9 **NEONATES: STUDY PROTOCOL**  
10

11  
12  
13  
14 Anouk W.J. Scholten<sup>1,2\*†</sup>, Zhan Zhuozhao<sup>3†</sup>, Hendrik J. Niemarkt<sup>4</sup>, Marieke Vervoorn<sup>4</sup>, Ruud  
15  
16 W. van Leuteren<sup>1,2</sup>, Frans H. de Jongh<sup>1,5</sup>, Anton H. van Kaam<sup>1,6</sup>, Edwin R. van den Heuvel<sup>3</sup>, G.  
17  
18 Jeroen Hutten<sup>1,6</sup>  
19

20  
21 *<sup>1</sup>Department of neonatology, Emma Children's Hospital, AmsterdamUMC, University of*  
22  
23 *Amsterdam, Amsterdam, The Netherlands*  
24

25  
26 *<sup>2</sup>Amsterdam Reproduction & Development research institute, Amsterdam, The Netherlands*  
27

28  
29 *<sup>3</sup>Department of Mathematics and Computer Science, Eindhoven University of Technology,*  
30  
31 *Eindhoven, The Netherlands*  
32

33  
34 *<sup>4</sup>Department of neonatology, Maxis Medical Center, Veldhoven, The Netherlands*  
35

36  
37 *<sup>5</sup>Faculty of Science and Technology, University of Twente, Enschede, The Netherlands*  
38

39  
40 *<sup>6</sup>Department of neonatology, Emma Children's Hospital, AmsterdamUMC, Vrije Universiteit,*  
41  
42 *Amsterdam, the Netherlands*  
43

44  
45 † These authors have contributed equally to this work and share first authorship

46  
47 **CORRESPONDING AUTHOR:**

48  
49 Anouk W.J. Scholten

50  
51 Department of Neonatology

52  
53 Emma Children's Hospital, AmsterdamUMC, location AMC

54  
55 Meibergdreef 9, 1105AZ Amsterdam, The Netherlands

56  
57 Phone: +31 20 5663675

58  
59 Email: a.w.j.scholten@amsterdamumc.nl  
60

## ABSTRACT

**Introduction:** Cardiorespiratory monitoring is used in the Neonatal Intensive Care Unit (NICU) to assess the clinical status of newborn infants and detect critical deteriorations in cardiorespiratory function. Currently, heart rate is monitored by electrocardiography (ECG) and respiration by chest impedance (CI). Disadvantages of current monitoring techniques are usage of wired adhesive electrodes which may damage the skin and hinder care. The Bambi® belt is a wireless and non-adhesive alternative that enables cardiorespiratory monitoring by measuring electrical activity of the diaphragm via transcutaneous electromyography (dEMG). A previous study showed feasibility of the Bambi belt and this study compares the belt performance to ECG and CI.

**Methods and analysis:** This multi-center non-inferiority paired study will be performed in the NICU of the Máxima Medical Center (MMC) in Veldhoven and the Emma Children's Hospital AmsterdamUMC in Amsterdam, the Netherlands. 39 infants in different postmenstrual age groups (minimally 10 infants <30 weeks, between 30-32 weeks and >32 weeks) will be recruited. These infants will be monitored with the Bambi® belt in addition to standard ECG and CI for 24 h. The primary outcome is the heart rate (HR), studied with three criteria: 1) the agreement in HR measurements between the belt and standard ECG, 2) the detection of cardiac events consisting of bradycardia and tachycardia and 3) the quality of HR-monitoring. The secondary outcome is the RR, studied with the criteria 1) agreement in respiratory rate (RR) trend monitoring, 2) apnea and tachypnea detection, and 3) reliable registrations.

**Ethics and dissemination:** This protocol was approved by the Medical Ethical Committee of the Máxima Medical Center and the Central Committee for Human Research (CCMO). The MMC started patient recruitment in July and the AmsterdamUMC in August 2021. The results will be presented at conferences and published in peer-reviewed journals.

**Trial registration number:** NL9480 ([www.trialregister.nl](http://www.trialregister.nl))

## INTRODUCTION

In the Neonatal Intensive Care Unit (NICU), cardiorespiratory monitoring is crucial to assess clinical condition and to timely detect and treat frequently occurring cardio-respiratory events to prevent morbidity and mortality.(1, 2) To date, this is performed by measuring the electrocardiogram (ECG) and chest impedance (CI) with three wired adhesive electrodes. CI measures variation in electrical impedance across the chest during respiration caused by changes in lung aeration and chest wall movement. These techniques provide continuous monitoring of heart rate (HR), respiratory rate (RR), and breathing pattern. However, as CI measures respiration indirectly, adequate detection of breathing cycles and apnea may not always be optimal.(3)

With transcutaneous electromyography of the diaphragm (dEMG) breathing effort can be recorded directly by measuring the electrical activity of this main respiratory muscle. To date, this technique also uses three adhesive electrodes and provides information on respiration and HR. Studies have shown its feasibility in the NICU-setting.(4)

The use and especially removal of adhesive electrodes can lead to epidermal stripping in vulnerable preterm infants, resulting in an increased risk of infection and pain.(5, 6) Furthermore, the wires attached to the electrodes restrict movements of the infant and may hinder both nursing and kangaroo care. Restrictions in kangaroo care may impact patient outcome as it has been associated with beneficial effects such as decreased mortality, decreased risk of severe infection/sepsis and hypothermia, and increased likelihood of exclusive breast feeding.(7, 8) Therefore, it is important to find alternatives for using wired adhesive electrodes.

1  
2  
3 In the past years, several wireless wearable sensors have been developed to measure various  
4 parameters in neonates such as ECG, HR, RR, peripheral oxygen saturation and (skin)  
5 temperature.(9-15) Recently, a novel wireless and non-adhesive sensor belt (Bambi® belt,  
6 Bambi B.V., Eindhoven, the Netherlands) was developed for neonatal use that measures ECG  
7 and respiration based on the dEMG technique. A recent pilot study showed that measuring HR  
8 and RR with this belt in preterm infants is feasible and that the measured HR and RR trend was  
9 similar to ECG and CI.(submitted for publication, NICU AmsterdamUMC, 2021) However,  
10 before replacing the current techniques using adhesive wired electrodes with the non-adhesive  
11 sensor belt, a larger study is required to demonstrate the non-inferiority of this belt as an  
12 alternative cardiorespiratory monitor. In this study, we compare the monitoring performance of  
13 the Bambi® belt to ECG and CI and hypothesize that the performance of the belt is non-inferior  
14 to the current monitoring techniques.  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32

## 33 **METHODS**

### 34 **Study design**

35 This multi-center paired non-inferiority study will be performed in the NICU of Máxima  
36 Medical Center (MMC) in Veldhoven and the Emma Children's Hospital of the Amsterdam  
37 University Medical Centre (AmsterdamUMC), both located in the Netherlands. Each patient  
38 will be simultaneously measured with the belt and ECG/CI (paired design). To compare the  
39 devices, a non-inferiority/equivalence framework will be used. Here, equivalence is defined as  
40 the limit of agreement of the HR/RR between the belt and ECG/CI being within prespecified  
41 clinically accepted margins. Non-inferiority is defined as the performance of clinical event  
42 detection and quality criteria not being worse than prespecified clinically accepted margins.  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

### Study population

Preterm and term infants being routinely monitored with the standard cardiorespiratory monitor (Intellivue MP90, Philips Healthcare, Eindhoven, The Netherlands) are included in the study.

To ensure a representative sample of the target population, infants in different age groups will be included. Infants with chest skin lesions, congenital anomalies, and other scenario's preventing belt placement, such as (effects of) surgery or wrap for therapeutic hypothermia, will be excluded.

### Primary outcome

As HR-monitoring is clinically most relied upon and both ECG and dEMG provide the HR by measuring cardiac electrical activity, while CI and dEMG measure respiration with a different technique, the HR is considered the primary outcome.(3, 16) This will be studied with three criteria. 1) Reliable monitoring performance through second-to-second HR measurement agreement between the belt and the ECG-CI monitoring. 2) The detection of a composite cardiac event consisting of bradycardia (HR < 100 beats per minute for at least five seconds)(17) and tachycardia (HR > 180 beats per minute for at least ten seconds)(18) between the belt and the ECG measured with adhesive electrodes. 3) Non-inferior quality (percentage of time with HR recordings without data loss).

Moreover, we will perform subgroup analyses to investigate whether the HR measurement performance is consistent under different clinical activities (e.g. kangaroo care, feeding) and in the different age groups.

### Secondary outcomes

The secondary outcome is the measured RR. This will be studied using the following three criteria:



- 1  
2  
3 1. Comparing the trend (10-minute moving average) in RR values provided by the belt and  
4  
5 CI, as the trend is the primary clinical usage of respiratory monitoring.(3) Since CI is  
6  
7 widely used for neonatal respiratory monitoring, it is used as the reference technique.  
8  
9
- 10 2. Comparing the ability to detect apnea and tachypnea. Clinically relevant apneas are  
11  
12 considered when indicated by a RR < 20 breaths per minute (to capture all periods of  
13  
14 low breathing frequency) measured with CI for at least 10 seconds, associated with a  
15  
16 desaturation (oxygen saturation (SpO<sub>2</sub>) <80% for at least 10 seconds) and/or  
17  
18 bradycardia (HR <100 beats per minute for at least five seconds) (objective apnea  
19  
20 measurement).(17)  
21  
22  
23 Tachypnea is defined as a prolonged period of RR >60 breaths per minute and >100  
24  
25 breaths per minute (approximately two times the average normal RR).(19) To cover  
26  
27 short and long periods of tachypnea, 3 different durations are studied (30 seconds, 60  
28  
29 seconds, and 10 minutes).  
30  
31  
32
- 33 3. Calculating the percentage of time with reliable respiratory monitoring (without data  
34  
35 loss and with an acceptable signal-to-noise ratio).  
36  
37  
38  
39

### 40 **Data collection**

41 The following basic characteristics and demographic information will be collected at the  
42  
43 baseline of the study: gestational age, birth weight, gender, age and weight at day of  
44  
45 measurement, relevant medical status (respiratory support, medication and underlying illness  
46  
47 during measurement), chest circumference, nipple distance, skin type at study start by visual  
48  
49 inspection (normal, dry, flaky, oily, moist, other).  
50  
51  
52  
53  
54

### 55 **Sample size calculation**

56 A power calculation is performed for the primary outcome using data collected in a previous  
57  
58 study.(submitted for publication, NICU AmsterdamUMC, 2021) Among the three criteria,  
59  
60

1  
2  
3 criteria 1 needs the largest sample size and is used for our study. This resulted in 39 required  
4  
5 infants to achieve 80% power with an overall 5% type I error with a Bonferroni correction  
6  
7 (details in the Statistical Analysis Plan (SAP) in the online supplement).  
8  
9

10 In addition, an interim analysis will be performed as the power calculation was based on the  
11  
12 previous study and recruitment of infants without being able to answer research questions is  
13  
14 unethical.<sup>(20)</sup> This will be performed after including 1/3th of the infants for sample size  
15  
16 adaption using the method of Mehta and Pocock.<sup>(21)</sup> If the conditional power falls within the  
17  
18 pre-defined “promising zone”, the sample size will be increased to an upper limit of 52 infants.  
19  
20  
21 Otherwise, the study will proceed with the original sample size.  
22  
23  
24  
25

### 26 **Study procedures**

27  
28 The Bambi® Belt System is a non CE-certified medical device, designed for wireless  
29  
30 cardiorespiratory monitoring of (pre)term infants in a hospital environment. All included  
31  
32 infants will be monitored with the belt in addition to standard ECG/CI for 24 hours to obtain  
33  
34 representative clinical scenarios throughout the entire day. The measurement set-up is  
35  
36 visualised in Figure 1 and consists of 1) dEMG measurement with the belt and 2) the extraction  
37  
38 of patient monitor data.  
39  
40  
41  
42  
43

44 In the belt, three dry electrodes are incorporated. When placing the belt at the height of the  
45  
46 diaphragm, the outer two electrodes are in the nipple line and the middle electrode is in line  
47  
48 with the sternum. The three ECG/CI electrodes are attached at the original location without  
49  
50 hindering belt placement. The measured electrical signal of the diaphragm with the belt is  
51  
52 wirelessly transmitted to the Receiver Module (REM) by the Sensor Module (SEM). The REM  
53  
54 processes the dEMG signal to obtain the ECG and respiration signal (averaged diaphragmatic  
55  
56 activity). An inbuilt algorithm provides the HR and RR out of the ECG and respiration signal  
57  
58 respectively. This data is transported to a bedside computer. The data from the patient monitor  
59  
60

1  
2  
3 (ECG, HR, RR, and SpO<sub>2</sub>) is extracted from the bedside monitor using an isolated cable and is  
4  
5 also transported to the bedside computer.  
6  
7  
8  
9

10 The belt data from the REM and patient monitor are recorded and synchronised using a  
11  
12 dedicated software package (Polybench, Applied Biosignals, Weener, Germany) on a personal  
13  
14 bedside computer. Data is recorded at a sample rate of 1 to 500 Hz for rate and waveform data  
15  
16 respectively. The bedside software also provides the possibility to make measurements  
17  
18 annotations during data recording, such as re-positioning of the infant, nursing and kangaroo  
19  
20 care.  
21  
22  
23  
24  
25

26 During the study, daily routine care proceeds as usual. The location of the belt is regularly  
27  
28 checked and if necessary repositioned (similar to the clinical practice). Notifications are  
29  
30 visualised when contact between skin and the belt is lost (Leads off) or when there is no  
31  
32 connection between the SEM and REM (Bluetooth Loss Error). In case of the first notification,  
33  
34 the belt may be repositioned, while in case of Bluetooth loss the battery level of the SEM or  
35  
36 blocking of this sensor (e.g. by an arm) are checked.  
37  
38  
39  
40  
41

42 Preferably, the belt stays in place during the study. However, the belt can be removed during  
43  
44 diagnostic imaging, patient handling, or in case of skin irritation at the belt location. The reason  
45  
46 for removal will be annotated. If the belt is removed, the medical staff, parents and one of the  
47  
48 dedicated researchers will decide together if the belt can be re-applied.  
49  
50  
51  
52

### 53 **Recruitment**

54  
55 Parents of all eligible infants are approached for consent to obtain a sample as heterogeneous  
56  
57 and representative as possible. Preferably, infants are included as soon as possible after birth.  
58  
59  
60

1  
2  
3 During the 24 hours, the study can be terminated if requested by parents or the treating  
4 physicians. In case of withdrawal of a subject, an extra subject will be included.  
5  
6  
7  
8  
9

### 10 **Safety**

11 Being a medical device study, this study was classified as a moderate risk.(22) A specified  
12 monitor plan for the study is made based on risk-classification.  
13  
14  
15  
16  
17  
18

### 19 **STATISTICAL ANALYSIS**

20 A detailed SAP can be found in the online supplement. Unless otherwise specified, all  
21 hypothesis tests are two-sided with a significance level of 0.05. All statistical analyses will be  
22 performed using R version 4.0 (the R Foundation for Statistical Computing; Vienna, Austria)  
23 and SAS software version 9.4 (SAS Institute Inc., Cary, NC, USA).  
24  
25  
26  
27  
28  
29  
30  
31

32 The non-inferiority/equivalence margins based on expert opinions (survey send to  
33 neonatologists of different NICU's in the Netherlands) and literature (4, 23, 24) are described  
34 in Table 1. In the different subparagraphs we refer to this table.  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 1** - The non-inferiority/equivalence margins for the primary and secondary outcomes

<i>Endpoints</i>	<i>acceptance margins<sup>#</sup></i>
<i>LOA of second-to-second HR differences</i>	$\pm 8$ bpm
<i>LOA of RR trend differences</i>	$\pm 15$ brpm
<i>Sensitivity of brady-/tachycardia detection</i>	90%
<i>PPV of brady-/tachycardia detection</i>	90%
<i>Sensitivity of apnea/tachypnea detection</i>	70%
<i>PPV of apnea/tachypnea alarms</i>	0-100%*
<i>Data loss percentage</i>	5%
<i>Robust data percentage (HR)</i>	90%
<i>Robust data percentage (RR)</i>	70%

LOA: limits of agreement, HR: heart rate, RR: respiratory rate, PPV: positive predictive value.

Data loss is defined as the percentage of data without “Leads off” or “Bluetooth Loss Error” in the belt.

<sup>#</sup> The acceptance margins are compared to confidence intervals with corresponding confidence levels (see SAP for more details).

\*Since the reference devices for apnea detection in the clinical practice are the peripheral oxygen saturation (SpO<sub>2</sub>) and electrocardiogram instead of the respiration signal and the performance for Chest Impedance to detect tachypnea is unsatisfactory due to the presence of cardiac interference, all values for PPV for apnea/tachypnea are acceptable. Interpretations will be made based on the results.

### Summary and descriptive statistics

Categorical data will be summarized by numbers of counts and percentages. Continuous data will be summarized by mean, standard deviation if data is normal and median, interquartile range (IQR) if data is skewed. Minimum and maximum values will also be presented for continuous data when appropriate.

### Statistical analysis of the primary outcome

#### Criteria 1: agreement in HR

To investigate the equivalence of HR measurement between the belt and ECG, we will fit a linear mixed model to the second-to-second HR difference between both. With this model, the 95% limits of agreement (Bland-Altman analysis) will be derived. The two-one-sided tests (TOST) with a multiplicity corrected alpha of 0.0167 and the prespecified margin ( $\pm 8$ bpm) will test equivalence between the two devices. In addition, based on a bivariate heteroscedastic model fitted to HR segments of a prespecified length, additional performance measures will be calculated as sensitivity analyses (details in SAP).

#### Criteria 2: cardiac event detection

For HR monitoring, we also consider the detection of bradycardia and tachycardia. We will estimate the sensitivity and the positive predictive value (PPV) of the belt using the patient monitor data as the ground truth and perform a non-inferiority test with an alpha of 0.0167. The non-inferiority margin for the sensitivity and PPV are listed in Table 1. In case of missed bradycardias, one independent expert per center will qualify the severity and acceptability of each missing event with the use of the discrepancy in HR and contextual information.

### Criteria 3: signal quality

The quality of the investigational device will be quantified based on the percentage of time during the 24-hour period it produces any reading (percentage without data loss due to “Leads off” or “Bluetooth Loss Error”) and the percentage in time it produces a good-quality-reading (percentage of robust data) for the HR and RR, respectively. For the HR non-robust data can be caused by bad connection (suboptimal Bluetooth or skin-electrode connection). These criteria are built-in in the belt algorithm and therefore this data is automatically labeled. Hypothesis testing will be used to establish the non-inferiority of this “uptime” percentage (percentage without data loss and percentage of robust data) of the belt.

For the RR, the uptime percentage is also categorized as a) data readings without data loss and b) robust data readings, i.e. good quality readings with acceptable signal-to-noise ratio’s undisturbed by among others suboptimal connection and patient handling. Again, the prespecified margins for the HR and RR are described in Table 1.

### Statistical analysis of secondary outcomes

Secondary analyses, based on the same statistical methods for the criteria of the primary outcome, include all secondary endpoints (apnea and tachypnea detection, RR trend analysis (see SAP)) and evaluation during different scenarios.

### ETHICS AND DISSEMINATION

The Medical Ethical Committee of the MMC (W21.042) and the Central Committee for Human Research in the Netherlands (CCMO, CCMO21/0167/PP) approved the study protocol (Version 2, 19<sup>th</sup> of May 2021). Local feasibility at the AmsterdamUMC was approved by the Medical Ethical Committee of the AMC (2021\_146). This study was registered in the Dutch Trial Register (<https://www.trialregister.nl>, NL9480). Regarding patient safety, no belt related events were observed in the pilot study and are therefore unexpected. Moreover, as every patient is

1  
2  
3 monitored with ECG/CI and the belt, safety is guaranteed in case of missing belt data. The SAP  
4 will be used for the analyses. The results will be published in peer-reviewed journals and  
5  
6 presented at future congresses.  
7  
8  
9

10 The MMC started patient recruitment in July and the AmsterdamUMC in August 2021. The  
11 duration of this study will be approximately seven months.  
12  
13  
14  
15  
16  
17  
18

### 19 **PATIENT AND PUBLIC INVOLVEMENT STATEMENT**

20  
21 Patients were included in this study after obtaining parental informed consent. The patients  
22 could not be involved in the design, recruitment, conduction and dissemination of results of this  
23 study. Neither could we ask the burden of the study. The outcome measures were developed by  
24 combining clinical and statistical knowledge to ensure a SAP that enables confirmation of non-  
25 inferiority of the belt compared to ECG/CI.  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36

### 37 **REFERENCES**

- 38  
39 1. Joshi R, van Straaten H, van de Mortel H, et al. Does the architectural layout of a NICU  
40 affect alarm pressure? A comparative clinical audit of a single-family room and an open bay  
41 area NICU using a retrospective study design. *BMJ Open* 2018;8(6):e022813. doi:  
42 10.1136/bmjopen-2018-022813  
43  
44  
45  
46  
47  
48 2. Bouwstra S, Chen W, Feijs LMG, et al. Designing the parents-to-infant bonding  
49 experience in the NICU. *IARIA* 2012:7-13.  
50  
51  
52  
53  
54 3. Di Fiore JM. Neonatal cardiorespiratory monitoring techniques. *Semin Neonatol*  
55 2004;9(3):195-203. doi: 10.1016/j.siny.2003.11.009  
56  
57  
58  
59  
60



- 1  
2  
3 4. Kraaijenga JV, Hutten GJ, de Jongh FH, et al. Transcutaneous electromyography of the  
4 diaphragm: A cardio-respiratory monitor for preterm infants. *Pediatr Pulmonol*  
5 2015;50(9):889-95. doi: 10.1002/ppul.23116  
6  
7  
8  
9
- 10  
11 5. Csoma ZR, Meszes A, Abraham R, et al. Iatrogenic Skin Disorders and Related Factors  
12 in Newborn Infants. *Pediatr Dermatol* 2016;33(5):543-8. doi: 10.1111/pde.12960  
13  
14
- 15  
16 6. Sekar KC. Iatrogenic complications in the neonatal intensive care unit. *J Perinatol*  
17 2010;30 (Suppl):P51-6. doi: 10.1038/jp.2010.102  
18  
19  
20
- 21  
22 7. Boundy EO, Dastjerdi R, Spiegelman D, et al. Kangaroo Mother Care and Neonatal  
23 Outcomes: A Meta-analysis. *Pediatrics* 2016;137(1). doi: 10.1542/peds.2015-2238  
24  
25
- 26  
27 8. Conde-Agudelo A, Diaz-Rossello JL. Kangaroo mother care to reduce morbidity and  
28 mortality in low birthweight infants. *Cochrane Database Syst Rev* 2016(8). doi:  
29 <https://doi.org/10.1002/14651858.CD002771.pub4>  
30  
31  
32
- 33  
34 9. Chung HU, Rwei AY, Hourlier-Fargette A, et al. Skin-interfaced biosensors for  
35 advanced wireless physiological monitoring in neonatal and pediatric intensive-care units. *Nat*  
36 *Med* 2020;26(3):418-29. doi: <https://doi.org/10.1038/s41591-020-0792-9>  
37  
38  
39
- 40  
41 10. Henry C, Shipley L, Ward C, et al. Accurate neonatal heart rate monitoring using a new  
42 wireless, cap mounted device. *Acta Paediatr* 2021;110(1):72-8. doi:  
43 <https://doi.org/10.1111/apa.15303>  
44  
45  
46  
47  
48
- 49  
50 11. Rimet Y, Brusquet Y, Ronayette D, et al. Surveillance of infants at risk of apparent life  
51 threatening events (ALTE) with the BBA bootie: a wearable multiparameter monitor. *Annu Int*  
52 *Conf IEEE Eng Med Biol Soc* 2007;2007:4997-5000. doi: 10.1109/IEMBS.2007.4353462  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 12. Chen H, Bao S, Lu C, et al. Design of an Integrated Wearable Multi-Sensor Platform  
4 Based on Flexible Materials for Neonatal Monitoring. *IEEE* 2020:23732-47. doi:  
5 10.1109/ACCESS.2020.2970469  
6  
7  
8  
9  
10 13. Chen W, Dols S, Bambang Oetomo S, et al. Monitoring body temperature of newborn  
11 infants at neonatal intensive care units using wearable sensors. *BodyNets '10* 2010. doi:  
12 <https://doi.org/10.1145/2221924.2221960>  
13  
14  
15  
16  
17 14. Linti C, Horter H, Osterreicher P, et al. Sensory baby vest for the monitoring of infants.  
18 *IEEE* 2006. doi: 10.1109/BSN.2006.49  
19  
20  
21  
22  
23 15. Pike H, Eilevstjonn J, Bjorland P, et al. Heart rate detection properties of dry-electrode  
24 ECG compared to conventional 3-lead gel-electrode ECG in newborns. *BMC Res Notes*  
25 2021;14(1):166. doi: 10.1186/s13104-021-05576-x  
26  
27  
28  
29  
30 16. van Leuteren RW, Hutten GJ, de Waal CG, et al. Processing transcutaneous  
31 electromyography measurements of respiratory muscles, a review of analysis techniques.  
32 *Journal of Electromyography and Kinesiology* 2019:176-86. doi:  
33 10.1016/j.jelekin.2019.07.014  
34  
35  
36  
37  
38  
39  
40 17. Lee H, Rusin CG, Lake DE, et al. A new algorithm for detecting central apnea in  
41 neonates. *Physiol Meas* 2012;33(1):1-17. doi: 10.1088/0967-3334/33/1/1  
42  
43  
44  
45 18. Li T, Matsushima M, Timpson W, et al. Epidemiology of patient monitoring alarms in  
46 the neonatal intensive care unit. *J Perinatol* 2018;38(8):1030-8. doi: 10.1038/s41372-018-  
47 0095-x  
48  
49  
50  
51  
52  
53 19. Reuter S, Moser C, Baack M. Respiratory distress in the newborn. *Pediatr Rev*  
54 2014;35(10):417-28. doi: 10.1542/pir.35-10-417  
55  
56  
57  
58  
59  
60

- 1  
2  
3 20. European Medicines Agency. ICH Topic E 9 – Statistical Principles for Clinical Trials.  
4  
5 2006:21-3.  
6  
7  
8  
9 21. Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are  
10 promising: a practical guide with examples. *Stat Med* 2011;30(28):3267-84. doi:  
11 <https://doi.org/10.1002/sim.4102>  
12  
13  
14  
15 22. Central Committee on Research Involving Human Subjects.  
16  
17  
18 23. Anton O, Fernandez R, Rendon-Morales E, et al. Heart Rate Monitoring in Newborn  
19 Babies: A Systematic Review. *Neonatology* 2019;116(3):199-210. doi: 10.1159/000499675  
20  
21  
22  
23 24. Breteler MJM, KleinJan EJ, Dohmen DAJ, et al. Vital Signs Monitoring with Wearable  
24 Sensors in High-risk Surgical Patients: A Clinical Validation Study. *Anesthesiology*  
25  
26 2020;132(3):424-39. doi: 10.1097/ALN.0000000000003029  
27  
28  
29  
30  
31  
32  
33

#### 34 **AUTHOR CONTRIBUTIONS**

35  
36 AS, HN, MV, RL, FJ, AK, JH conceptualized the study. ZZ and EH made the statistical analysis  
37 plan, which was reviewed by all authors. AS wrote the first version of this manuscript. All  
38 authors contributed to the final draft of the manuscript.  
39  
40  
41  
42  
43  
44

#### 45 **FUNDING STATEMENT**

46 This work was supported by the Louise Vehmeijer foundation.  
47  
48  
49  
50  
51

#### 52 **COMPETING INTERESTS STATEMENT**

53  
54 The authors declare that the research was conducted in the absence of any commercial or  
55 financial relationships that could be construed as a potential conflict of interest.  
56  
57  
58  
59  
60

**WHAT IS KNOWN ABOUT THE SUBJECT:**

- Disadvantages of the cardiorespiratory monitoring technique in neonates are indirect measurements of respiration, usage of adhesive electrodes and hindering wires.
- With transcutaneous electromyography of the diaphragm, respiratory activity is measured directly by recording the activity of the main respiratory muscle.
- The Bambi® belt is a novel wireless and non-adhesive belt that enables cardiorespiratory monitoring by measuring diaphragm activity with dry electrodes.

**WHAT THIS STUDY HOPES TO ADD:**

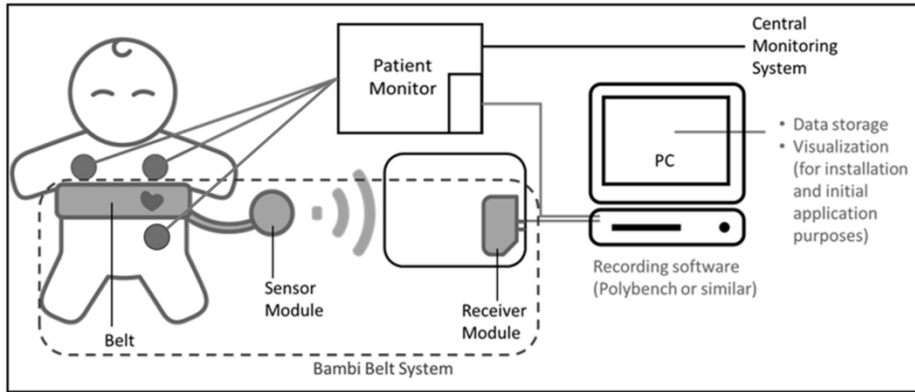
- Demonstration of the non-inferiority of the Bambi® belt compared to the electrocardiogram and chest impedance for cardiorespiratory monitoring in preterm and term infants.

**HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY**

- When non-inferiority of the Bambi® belt compared to the current cardiorespiratory monitor is confirmed, the belt could be used as a wireless and skin-friendly alternative.

## FIGURES

**FIGURE 1** - The measurement set-up. The adhesive electrodes used for standard cardiorespiratory monitoring are attached at the original location, visualised by the three blue dots. The diaphragm activity measured with the Bambi® belt is wirelessly transmitted with the Sensor Module to the Receiver Module where the data is processed to obtain an electrocardiogram and respiration waveform (and heart rate and respiratory rate). This data and the data measured with the patient monitor are transported to a personal bedside computer with Polybench software to synchronise and record these signals.



170x80mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## ONLINE SUPPLEMENT – STATISTICAL ANALYSIS PLAN

### PURPOSE OF THE STATISTICAL ANALYSIS PLAN

The purpose of this statistical analysis plan (SAP) is to document technical and detailed specification for the analysis of data collected in the Bambi Belt monitoring performance study.

The SAP has been written based on information contained in study protocol, dated 12th April 2021 before any data collection had taken place. It is prepared in compliance with the International Council on Harmonization (ICH) E9.

This SAP will be the guiding document for the analyses that will be conducted. Results of the analyses described in this SAP will be included in the Clinical Study Report (CSR). Any post hoc or unplanned analyses performed to provide results for inclusion in the CSR, but not identified in the prospective SAP will be identified in the given report. Additionally, the planned analyses of the primary aims will be included in future manuscripts. All the aims and research questions will be presented as an addendum as well.

### OVERVIEW AND DESCRIPTION OF THE STUDY

#### Study design

The study is a multicenter, paired design, clinical monitoring device measurement comparison study. The *investigational device* under consideration is the Bambi® Belt monitoring system (using dry electrodes). The current standard device of cardiorespiratory monitoring through adhesive electrodes is considered as the clinical reference standard and thereafter referred to as the *reference device/method*. The Bambi Belt® monitoring system will be used on infants by trained nurses in the neonatal intensive care units (NICU's) for continuous 24 hours monitoring in addition to the routine monitoring with the reference device on the same patients. Infants admitted to NICU's of the the Emma Children's Hospital of the Amsterdam University Medical

Centre (Amsterdam UMC) or Maxima Medical Center (MMC) will be measured at the earliest suitable moment for clinical practice without interfering with infants' routine cycles.

### Randomization and blinding

No randomization is required for the paired design since both monitoring devices will be used on the same patient at the same time. Blinding is also not possible since both the measurement protocol and algorithmic characteristic differ substantially.

### Framework

The goal of this study is to establish the agreement between the investigational device and the reference device. Unlike the traditional difference-based tests, non-inferiority and equivalence techniques provide a better alternative for demonstrating the similarity between the two measurement methods. Thus, we have adopted the non-inferiority/equivalence trial framework for this primary objective of this study. This study considers three hypotheses ( $H_0$  denotes the null hypothesis and  $H_A$  denotes the alternative hypothesis) for the first two primary outcomes:

1. Primary outcome, criterion 1: Heart rate measurement (second-by-second measurement)

$H_0$ : The absolute difference between the investigational device and the reference device is larger than the prespecified equivalence margin.

$H_A$ : The absolute difference between the investigational device and the reference device is within the prespecified equivalence margin.



1  
2  
3 2. Primary outcome, criterion 2: Brady-/tachy-cardia event detection  
4

5  $H_0$ : The composite cardiac event detection performances in terms of sensitivity and positive  
6 predictive value (PPV) based on the investigational device with respect to the reference device  
7  
8 is less than the prespecified non-inferiority margin.  
9

10  
11  
12  $H_A$ : The composite cardiac event detection performances in terms of sensitivity and PPV based  
13 on the investigational device with respect to the reference device is greater or equal to the  
14 prespecified non-inferiority margin.  
15  
16  
17

18  
19  
20  
21 3. Primary outcome, criterion 3: Reliable reading (percentage of the time)  
22

23  
24  $H_0$ : The percentage of the time the investigational device produces reliable readings is less than  
25 the prespecified non-inferiority margin.  
26

27  
28  $H_A$ : The percentage of the time the investigational device produces reliable readings is greater  
29 or equal to the prespecified non-inferiority margin.  
30  
31  
32

33  
34  
35 **Statistical interim analysis and stopping guidance**  
36

37 One interim analysis for sample size adaptation will be performed. That is, we will start with a  
38 certain sample size commitment which will be increased at the interim analysis in case the  
39 results obtained are reasonably promising. The interim analysis will be conducted after the  
40 prospectively recruited participant's number reaches one-third of the planned sample size.  
41  
42 Conditional power will be calculated for the analyses of the primary endpoints and compared  
43 to the boundary values of the conditional power for the promising zones (1, 2). In case the  
44 conditional power calculated at the interim analysis does fall inside the promising zone, the  
45 sample size will be increased to a pre-determined limit. On the other hand, if the calculated  
46 conditional power is outside the promising zone, the study will proceed with the original sample  
47 size. Therefore, no early stopping rule is entailed in this study. Furthermore, a conventional  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 final analysis will be used without altering the level of type I error, since the promising zone is  
4  
5 defined as a set that ensures the type I error to be preserved conservatively for the final analysis.  
6  
7  
8  
9

## 10 **Study data**

11  
12 The following patient characteristics will be collected at baseline:

- 13
- 14
- 15 • Gestational age
- 16
- 17 • Postmenstrual age
- 18
- 19 • Gender
- 20
- 21
- 22 • Weight at enrollment
- 23
- 24 • Ethnicity
- 25
- 26 • Chest circumference
- 27
- 28 • Nipple distance
- 29
- 30
- 31 • Skin condition and abnormality
- 32

33  
34 During the monitor study period, the following information will be measured:

- 35
- 36 • Clinical event
- 37
- 38 • SpO<sub>2</sub>: Blood oxygen saturation level
- 39
- 40 • Medical status:
- 41
- 42
  - 43 ○ Ventilation support
  - 44
  - 45 ○ Reports of medication and illness during the measurement
  - 46
- 47
- 48 • Lead status: Indicates whether at least one lead was off
- 49
- 50 • Bluetooth link quality
- 51
- 52 • Activities
- 53
- 54
  - 55 ○ Kangaroo care
  - 56
  - 57 ○ Nurse care
  - 58
  - 59 ○ Feeding
  - 60

- Medical Procedure
- Belt status
  - Moved: the belt is being moved
  - Open: the belt is removed from the patient
- Patient position
  - Unknown
  - Lying prone
  - Lying supine
  - Lying on the left side
  - Lying on the right side

## STATISTICAL ANALYSIS

Based on the collected information described above, the following total of variables will be derived:

- 10, 30, and 60 minutes moving average of the heart rate, and respiratory rate measured by both the investigational device and the reference device.
- Premature birth:
  - Premature (gestational age < 37 weeks)
  - Normal (gestational age  $\geq$  37 weeks)
- Desaturation: SpO<sub>2</sub> < 80% for at least 10 consecutive seconds
- Heart rate status (investigational and reference device):
  - Normal
  - Tachycardia (heart rate > 180 for at least 10 consecutive seconds)
  - Bradycardia (heart rate < 100 for at least 5 consecutive seconds)
- Respiration status (investigational and reference device):

- Apnea (according to standard clinical definitions)
- Tachypnea (respiratory rate >60 and >100 for 30 seconds, 1 minute, and 10 consecutive minutes in stationary signal)
- Measurement quality:
  - No anomalies
  - Poor data link: Bluetooth link is poor but data is still received
  - Unreliable data: One or more lead off, or no Bluetooth connection (Bluetooth Loss Error, BLE)

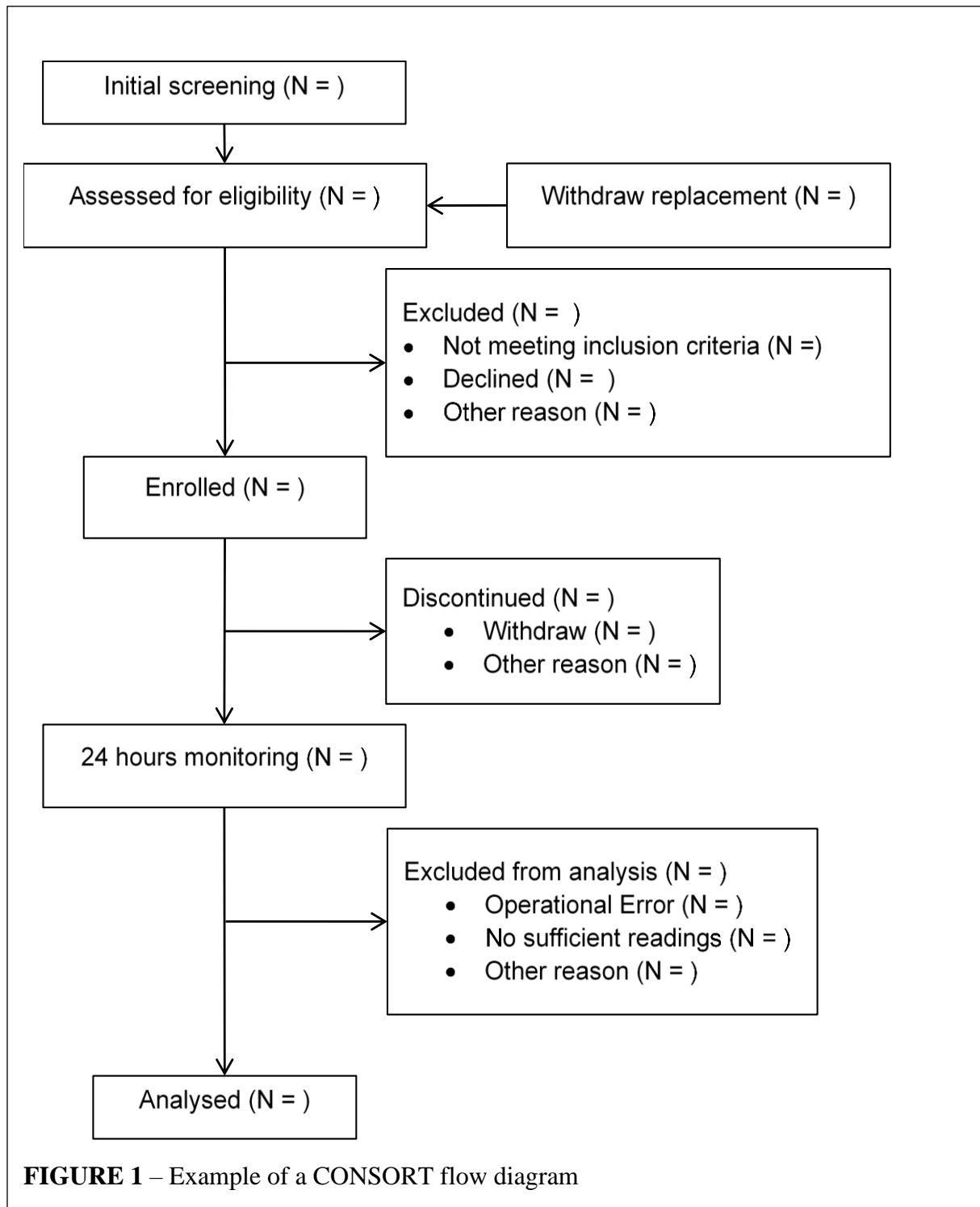
### Summary and descriptive statistics

Categorical data will be summarized by numbers and percentages. Continuous data will be summarized by mean, standard deviation if data are normal and median, interquartile range (IQR) if data are skewed. Minimum and maximum values will also be presented for continuous data. Tests of statistical significance will not be undertaken for baseline characteristics; rather the clinical importance of any imbalance will be noted.

A CONSORT flow diagram (example in Figure 1) will be used to summarize the number of infants who were:

- Assessed for eligibility at the screening
  - Eligible at screening
  - Ineligible at screening (with reasons)
- Eligible and enrolled
- Eligible but not enrolled
- Enrolled but did not receive any / sufficient measurements
  - Discontinued

- Included in the analysis
- Excluded in the analysis



## Analysis methods

Primary outcome, criterion 1: Heart rate measurement

To investigate and verify the equivalence of heart rate measurement between the investigational device and the reference device, we will fit a linear mixed model to the second-to-second heart rate difference between the two devices. Based on the estimates of the model, we will derive the 95% limits of agreement (3) as our main performance measure, known as the Bland-Altman analysis. The endpoints of the Bland-Altman 95% limits of agreement are the 2.5th percentile and 97.5th percentile for the distribution of the difference between paired measurements. We will calculate the  $(1 - \alpha/2)100\%$  confidence intervals of the percentiles according to Shieh (4), and conduct the two-one-sided t-tests (TOST) procedure with the prespecified equivalence margins (Table 1).

In addition, we will calculate the following performance measures to supplement the main analysis as sensitivity analyses to assess the agreement between the two devices from different aspects:

- The concordance correlation coefficient (5) and its variants
- Probability of Agreement (6) and Total Deviation Index (7)
- Coefficient of individual agreement (8)

These performance measures will be based on a bivariate heteroscedastic linear mixed-effects model fitted to each segment of the readings of a prespecified length from both devices. We will assume that measurements made between the two devices at the same time are correlated. Therefore, investigating the correlation between the two devices leads to the quantification of the degrees of agreement between them. Furthermore, we will consider the temporal correlations between measurements obtained with the same devices and the variabilities between different infants. Besides, we will start with a heteroscedastic model which does not assume equal variances for the two devices (namely, the measurement errors are not assumed

1  
2  
3 to be equal) and investigate the homogeneity of the measurement variabilities between the two  
4  
5 devices. Baseline characteristics of the infants and records of activities (listed in the study data  
6  
7 section) will be used as covariates in the model to partly explain the variabilities between the  
8  
9 infants. We will use the stepwise model selection procedure based on the Bayesian Information  
10  
11 Criteria (BIC) goodness-of-fit criteria.  
12  
13

#### 14 15 16 17 Primary outcome, criterion 2: Brady-/tachycardia event detection

18  
19 For brady-/tachycardia, the clinical event periods will be identified based on pre-specified  
20  
21 clinical definitions. We will investigate the non-inferiority of sensitivity and positive predictive  
22  
23 values (PPV) of the event detected by the investigation device assuming that the reference  
24  
25 device is the predicate device and compare both values to the prespecified non-inferiority  
26  
27 margins (Table 1). For the calculation of the sensitivity, when the event period identified based  
28  
29 on the investigational device overlaps with the event period identified by the reference device,  
30  
31 it will be counted as a true positive case. This is to prevent the repeated signaling of events from  
32  
33 the investigational device during a positive period identified by the reference device to inflate  
34  
35 the number of true positives. The same applies to the reference device when it comes to the  
36  
37 calculation of the PPV. That is, during an event period identified by the investigational device,  
38  
39 multiple event periods identified by the reference device will only be counted as one true-  
40  
41 positive case. Note that the true negative is ill-defined and will not be reported. Since true  
42  
43 negatives are used in the calculation of specificity, specificity will not be reported either.  
44  
45  
46  
47  
48  
49  
50

#### 51 Primary outcome, criterion 3: Safety and Quality

52  
53 *Safety:* The investigation of safety and tolerability is a multidimensional problem. Although we  
54  
55 don't anticipate any specific adverse effects for the investigational device, new and  
56  
57 unforeseeable effects are always possible. This background underlies the statistical difficulties  
58  
59  
60

1  
2  
3 associated with the analytical evaluation of the safety and tolerability of the device. We will  
4 address the safety and tolerability implications by applying descriptive statistical methods to  
5 the data, supplemented by calculation of confidence intervals whenever this aids interpretation  
6 and make use of graphical presentations in which patterns of adverse events are displayed.  
7  
8  
9  
10  
11

12  
13  
14 *Quality:* The quality of the investigational device will be quantified in terms of the point  
15 estimate and 95% confidence intervals based on the estimated percentages in time during the  
16 24-hour period it produces reliable readings for heart rate and respiratory rate, respectively.  
17 Reliable readings are defined in the study protocol. The uptime percentages are the percentage  
18 of data loss and the percentage of robust data readings. For each outcome, hypothesis testing  
19 will be used to establish the non-inferiority of the uptime percentages of the investigational  
20 device considering a non-inferiority margin specified in Table 1. The uptime percentages will  
21 be estimated based on a Generalized Estimation Equations (GEE) model.  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34

### 35 Missing data

36  
37 To get an idea about the complexity of the missing data problem in the data and information  
38 about the location of the missing values, the missing data pattern will be evaluated and reported.  
39  
40 We expect missing data in the primary outcomes measured by the investigational device to be  
41 the results of external causes such as the movement of the belt, signal losses, poor Bluetooth  
42 link qualities and so on. Therefore, it will be reasonable to assume that data are missing  
43 completely at random (MCAR). Formally, we will investigate the validity of such an  
44 assumption using Little's MCAR test. Furthermore, the availability of the data from the  
45 reference device (since it depends on a separate measurement system) provides us the  
46 opportunity to investigate whether the missingness is related to the underlying measurand. That  
47 is, whether the missing data mechanism is missing not at random (MNAR). This is rarely  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



possible in other types of studies. Nevertheless, considering the pair of bivariate measurements from the investigational and the reference device, we will investigate the assumption using the covariate-dependent missing (CDM) test proposed in Li (9). Note that CDM is usually considered as missing at random (MAR), we here simply exploit the advantage of the data from the reference device to test the dependencies between the missingness and the underlying measurand. Furthermore, we will use the CDM test on other covariates (excluding the reference device data) as well to test if the missingness is MAR.

In the case of MAR (i.e., CDM without measurements from reference device), list wise deletion can still be unbiased and will be used if the percentage of missingness is less than 5%. Otherwise, multiple imputations (MI) will be considered. We will not use the measurements from the reference device for the MI to avoid biasing the results towards the equivalence of the two devices. On the other hand, if the missingness is related to the measurand after taking into account all covariates, this indicates a potential problem of the measurement device, and a separate analysis will be carried out to investigate the associations between the missingness and the measurand.

For multiple imputations, we will use the fully conditional specification method. Unrealistic values (e.g., negative values for strictly positive variable) will be checked and corrected (e.g., using truncations). The imputation will be repeated at least 5 times and Rubin's rule will be used to combine estimates and standard errors from the imputed data.

### Secondary analyses

If the sample size permits, we will perform subset analyses to explore the performances of the investigational device under different scenarios.

### Subset analyses: primary endpoints

For each of the primary endpoints, we will consider additional exploratory analyses on the following subsets:

- During periods of a clinical event (e.g., apnea, bradycardia)
- During activities (e.g., Kangaroo care, feeding)
- During periods where the reference device's readings are stable
- Gestational age (e.g., preterm birth)
- Respiratory support (e.g., mechanical ventilation)

For these subsets, we will use the same model as the primary outcome to investigate the performances of the investigational device under various scenarios/activities of the infants. In case the subset does not contain enough data to fit the same model as the primary one, we will resort to a simpler model for case-by-case analyses.

### Respiratory rate analysis

It is known the reference device does not provide point-by-point accurate measurement resulting in large variabilities (measurement errors) in the measured respiratory rates. The intended clinical use of the readings in the NICU thus consists of two different aspects:

1. The trend of the respiratory rates over time;
2. Signaling of potentially respiratory related clinical events (i.e. apnea related desaturation and/or bradycardia, and potentially disease related tachypnea);

For the first usage, we will apply the same analysis method as the one used for heart rate on the moving average of the respiratory rate. We will primarily focus on the 10 minutes moving average for the respiratory rate. Analysis of the 1 minute and 5 minutes moving averages will

1  
2  
3 be used as a sensitivity analysis to establish the robustness of the conclusions made for the 10  
4  
5 minutes moving average.  
6

7 For apnea and tachypnea, respectively, the clinical event periods will be identified based on  
8  
9 clinical definitions and the same methods as the brady-/tachycardia event detection will be used  
10  
11 to compare the sensitivity and PPV to the prespecified non-inferiority limits (Table 1).  
12  
13 However, it should be noted that since the reference device is known to have an unsatisfactory  
14  
15 performance of apnea/tachypnea detection, cautions are needed to interpret the sensitivity and  
16  
17 PPV as if the reference device is the truth.  
18  
19  
20  
21  
22

#### 23 Statistical software

24 All statistical analyses will be performed using R version 4.0 (the R Foundation for Statistical  
25  
26 Computing; Vienna, Austria) and SAS software version 9.4 (SAS Institute Inc., Cary, NC,  
27  
28 USA).  
29  
30  
31  
32  
33  
34

#### 35 Non-inferiority/equivalence criteria

36 In Table 1 the non-inferiority/equivalence criteria for the primary and secondary outcomes are  
37  
38 visualized.  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 1** - The non-inferiority/equivalence margins for the primary and secondary outcomes

<i>Endpoints</i>	<i>acceptance margins<sup>#</sup></i>
<i>LOA of second-to-second HR differences</i>	$\pm 8$ bpm
<i>LOA of RR trend differences</i>	$\pm 15$ brpm
<i>Sensitivity of brady-/tachycardia detection</i>	90%
<i>PPV of brady-/tachycardia detection</i>	90%
<i>Sensitivity of apnea/tachypnea detection</i>	70%
<i>PPV of apnea/tachypnea alarms</i>	0-100%*
<i>Data loss percentage</i>	5%
<i>Robust data percentage (HR)</i>	90%
<i>Robust data percentage (RR)</i>	70%

LOA: limits of agreement, HR: heart rate, RR: respiratory rate, PPV: positive predictive value.

Data loss is defined as the percentage of data without “Leads off” or “Bluetooth Loss Error” in the belt.

<sup>#</sup> The acceptance margins are compared to confidence intervals with corresponding confidence levels (see SAP for more details).

\*Since the reference devices for apnea detection in the clinical practice are the peripheral oxygen saturation (SpO<sub>2</sub>) and electrocardiogram instead of the respiration signal and the performance for Chest Impedance to detect tachypnea is unsatisfactory due to the presence of cardiac interference, all values for PPV for apnea/tachypnea are acceptable. Interpretations will be made

## Sample size

Based on preliminary analysis of data collected in a feasibility study on a total of 13 infants with measurements from both the investigational device and the reference device, we were able to obtain preliminary information with regards to the characteristics of the primary endpoints

1  
2  
3 upon which we have formulated our sample size calculation.(submitted for publication, NICU  
4  
5 AmsterdamUMC, 2021)  
6  
7  
8  
9

10 A detailed specification of the sample size calculation can be found in the sections below. In  
11  
12 summary, for the monitor performance, 39 infants are needed to achieve 80% power with a 5%  
13  
14 overall type I error with a Bonferroni correction for multiplicity. It is worth noting that no  
15  
16 dropout was assumed during the sample size calculation. This is because we plan to include an  
17  
18 extra infant in case of withdrawal of an infant to fulfil the required sample size. Infants who  
19  
20 withdraw from the study will be followed up by one of the investigators and responsible medical  
21  
22 staff to obtain detailed reasons behind the withdraw. Dropout rate for the monitor performance  
23  
24 study is expected to be low, between 0-5%.  
25  
26  
27  
28  
29

30 While the pre-planned sample size is 39 infants, we will include an adaptive sample size re-  
31  
32 estimation procedure as per the “promising zone” methodology of Mehta and Pocock (2) using  
33  
34 the data from the first 1/3 infants. This procedure involves the evaluation of conditional power  
35  
36 in the interim analysis, and if it were to fall in the pre-specified “promising zone”, the sample  
37  
38 size will be increased, subject to a pre-determined upper limit (52 infants) to increase the  
39  
40 conditional power to 80%. The boundary of the conditional power for the “promising zone” is  
41  
42 0.36 and 0.8.  
43  
44  
45  
46  
47  
48

49 Monitoring study: Primary endpoints: Heart rate

50  
51 For the sample size calculation, we will assume the measured heart rate difference  $D_{ij}$  between  
52  
53 the investigational device and the reference device at time point  $j$  ( $j = 1, \dots, m$ ) on infant  $i$  ( $i =$   
54  
55  $1, \dots, n$ ) can be modelled as:  
56  
57

$$D_{ij} = d + a_i + e_{ij}$$

where  $d$  is the overall difference,  $a_i$  is a random effect with  $a_i \sim N(0, \sigma_a^2)$ , and  $e_{ij} \sim N(0, \sigma_e^2)$  is the random error independent of  $a_i$ . Though, we considered a bivariate mixed-effects model for our analysis, the variance component model for the difference can be derived from the bivariate mixed-effects model, therefore we will use this variance component model for the sample size calculation. The variance of the difference will be estimated from the aforementioned model via  $\hat{\sigma}_d^2 = \hat{\sigma}_a^2 + \hat{\sigma}_e^2$ . Here  $\hat{\sigma}_a^2$  and  $\hat{\sigma}_e^2$  is the estimator of the between-subject variability  $\sigma_a^2$  and residual variability  $\sigma_e^2$ , respectively. The 95% limit of agreement (LOA) can be estimated as  $\text{LOA} = \hat{d} \pm 1.96 \hat{\sigma}_d$  with  $\hat{d}$  and  $\hat{\sigma}_d$  denotes the estimator of  $d$  and  $\sigma_d$ , respectively. The variance of the LOA estimator is  $\text{var}(\hat{d} \pm 1.96 \hat{\sigma}_d) = \text{var}(\hat{d}) + 1.96^2 \text{var}(\hat{\sigma}_d)$  ( $\hat{d}$  and  $\hat{\sigma}_d$  is asymptotically independent). Since for  $\hat{\sigma}_d^2 = \hat{\sigma}_a^2 + \hat{\sigma}_e^2$ , we have  $\text{var}(\hat{\sigma}_d^2) = \text{var}(\hat{\sigma}_a^2) + \text{var}(\hat{\sigma}_e^2) + \text{cov}(\hat{\sigma}_a^2, \hat{\sigma}_e^2)$ . Furthermore, each term on the right-hand side (assuming  $m$  is large) is given by:

$$\text{var}(\hat{\sigma}_a^2) = \frac{2}{m^2} \left[ \frac{(m\sigma_a^2 + \sigma_e^2)^2}{n-1} + \frac{\sigma_e^4}{n(m-1)} \right] \approx \frac{2\sigma_a^4}{n-1}, \text{var}(\hat{\sigma}_e^2) = \frac{2\sigma_e^4}{n(m-1)+2} \approx 0,$$

$$\text{cov}(\hat{\sigma}_a^2, \hat{\sigma}_e^2) = -\frac{2\sigma_e^4}{nm(m-1)} \approx 0;$$

This leads to  $\text{var}(\hat{\sigma}_d^2) \approx 2\sigma_a^4/(n-1)$ . Therefore, by the delta method, we have  $\text{var}(\hat{\sigma}_d) = \frac{1}{4\sigma_d^2} \text{var}(\hat{\sigma}_d^2) = \frac{\sigma_a^4}{2(n-1)\sigma_d^2}$ . According to Lu et al. (10), the power for the TOST is given by:

$$1 - \beta = 1 - T_{n-1} \left( t_{1-\frac{\alpha}{2}}, \frac{\delta - d - 1.96\sigma_d}{se_{LOA}} \right) - T_{n-1} \left( t_{1-\frac{\alpha}{2}}, \frac{\delta + d - 1.96\sigma_d}{se_{LOA}} \right)$$

where  $\alpha$ ,  $\beta$  denotes type I and type II error respectively,  $\delta$  is the pre-defined acceptance limit that is clinically acceptable,  $se_{LOA} \approx \sqrt{\frac{\sigma_d^2}{n} + \frac{1.96^2 \sigma_a^4}{2(n-1)\sigma_d^2}}$  is the standard error of the LOA estimate calculated according to the variance component model, and  $T_{n-1}(\cdot, \tau)$  denotes the cumulative distribution function of a non-central Student's t-distribution with  $n-1$  degrees of freedom, and non-centrality parameter  $\tau$ .

For a 5% overall type I error rate, with a multiplicity correction factor of 3, and 80% power, the minimum sample size required is calculated at  $n = 39$ , for  $d = -0.5$ ,  $\sigma_a = 0.3$ , and  $\sigma_d = 3$ .

Primary endpoints: Brady-/tachycardia event detection

Suppose the total number of true events is  $M$  and are 100% detected by the reference device. Assuming the true sensitivity is 95% for the investigational device, then a non-inferiority test using Z-test with normal approximation to the binomial distribution leads to a required  $M$  of 271 for a power of 80% and  $\alpha = 0.05/3 \approx 0.01667$  assuming the detection between each event (conditioning on the event itself) is independent. Considering the incidence of bradycardia to be 1 event per hour per infant according to the preliminary analysis of data from the feasibility study, at least 12 infants are needed to satisfy the required  $M$  (assuming each infant is measured for 24 hours long). The calculation is the same for PPV if we assume the investigational device is the truth. Assuming an incidence rate of 1.5 per hour per infant according to the preliminary analysis of data from the feasibility study, the required sample size is 8. Note that in the aforementioned calculation, we assume that the event-detection

performance of the investigational device is homogeneous (or independent) among infants. A sensitivity/robustness investigation regarding the sample size for infant-specific heterogeneous performances was performed, with results from which we can see that with  $n = 39$ , we have more than 90% power to detect a heterogeneous performance scenario where 15% of the population would have sensitivities between 80% - 90% and less than 5% of the population have sensitivities less than 80%.

Primary endpoints: Safety and quality

Based on preliminary analysis of data collected in the feasibility study, we will assume that the overall probability of producing an erroneous reading at any time  $p_e$  is 2% and is constant across all participants. We will consider a non-inferiority test using normal approximation and a Z-test with the null hypothesis of  $H_0: p_e > 0.05$  and the alternative hypothesis of  $H_A: p_e \leq 0.05$ . The required number of observations for a given type I error of 1.667% ( $\approx 5\%/3$ ) to achieve 80% power is 376. Here the sample size 376 refers to 376 independent observations. Considering the large numbers of repeated measurements (more than 376) within each participant, we will have sufficient power for this non-inferiority test even with 1 participant. However, the assumption of independence can be too strong in the setting of our study. Therefore, if we would assume an AR(1)-type dependency with correlation parameter  $\rho = 0.8$  between two measurements within a participant, the variance inflation factor (VIF) for the asymptotic variance of the GEE estimator  $\hat{p}_e$  according to Pan is approximately (with the number of repeats  $m = 376$ ):

$$1 + \frac{2\rho}{1 - \rho} = 9$$

in the case of identity working correlation matrix when the true correlation has an AR(1) structure. To achieve the same power as the independent case calculated before, we need



$$\text{var}(\hat{p}_e) := \text{VIF} \frac{p_e(1-p_e)}{nm} = \frac{p_e(1-p_e)}{m}$$

Thus, we can conclude that at least  $n = \text{VIF} = 9$  participants will be needed to provide enough power for the non-inferiority test based on the GEE estimator using the identity working correlation matrix using the inverse proportionality between the required sample size and the variance of the estimator used in the Z-test. The same calculation can be carried for the robust data percentages. It can be seen that only the number of repeats  $m$  will differ when the probabilities and the non-inferiority margins change while the VIF remains the same for the same value of the correlation parameter  $\rho$ . Among all settings, the largest  $m$  needed will be 718 when we assume the probability of producing robust data for respiratory rate is 75% with the corresponding non-inferiority margin equals to 70%. This number of repeats is still fully covered by the high-frequency measurements found in the study.

### Protocol deviations and analysis sets

#### Definition of protocol deviations

Protocol deviations (PD) occurring during the study will be determined for all enrolled infants, mainly from the clinical database by either clinical and/or medical review processes.

The mapping of the protocol deviations from the clinical database to analysis will be performed as per Table 2:

Table 2 – The influence of protocol deviations on the statistical analysis plan (SAP)

<i>Database label</i>	<i>SAP</i>
<i>Minor</i>	Not required
<i>Major</i>	Important
<i>Critical</i>	Important
<i>Clinical (a subset of Critical)</i>	Important

Important protocol deviations are protocol deviations that might significantly affect the completeness, accuracy, and/or reliability of the study data or that might significantly affect a subject's rights, safety, or well-being.

Important protocol deviations may also be recorded as "Major" protocol deviations in the database, but will be presented only as important in the analysis output.

Important protocol deviations include:

- Infants that are included in the study despite not satisfying the eligibility criteria;
- Infants that develop exclusion criteria while on the study but not withdrawn;
- Infants being measured with operational human errors;
- Deviation from Good Clinical Practice (ICE E6)

Clinically Important protocol deviations are a subset of important protocol deviations which lead to the exclusion of a subject from the analysis set.

The following deviations will be identified and confirmed before the partial database lock for the final analysis.

- Important protocol deviations including
  - Deviations from the inclusion and exclusion criteria
  - Deviations post inclusion

1  
2  
3 Protocol deviations may be identified by the data managers, clinical and medical staff either by  
4  
5 programmed validation checks or data listings/reports or manual verification of data sources.  
6  
7  
8 Some important/major protocol deviation criteria may be identified in the clinical database via  
9  
10 biostatistical programs. Every important protocol deviation will be documented in the database  
11  
12 whether identified through sites monitoring, medical review or programming.  
13  
14  
15  
16

## 17 REFERENCES

- 18  
19  
20 1. European Medicines Agency. ICH Topic E 9 – Statistical Principles for Clinical Trials.  
21  
22 2006:21-3.  
23  
24
- 25 2. Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are  
26  
27 promising: a practical guide with examples. *Stat Med* 2011;30(28):3267-84. doi:  
28  
29 <https://doi.org/10.1002/sim.4102>  
30  
31
- 32 3. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat*  
33  
34 *Methods Med Res* 1999;8(2):135-60. doi: 10.1177/096228029900800204  
35  
36  
37
- 38 4. Shieh G. The appropriateness of Bland-Altman's approximate confidence intervals for  
39  
40 limits of agreement. *BMC Med Res Methodol* 2018;18(1):45. doi:  
41  
42 <https://doi.org/10.1186/s12874-018-0505-y>  
43  
44
- 45 5. Lin L. I.-K. A Concordance Correlation Coefficient to Evaluate Reproducibility.  
46  
47 *Biometrics* 1989:255-68.  
48  
49
- 50 6. Stevens NT, Steiner SH, MacKay RJ. Assessing agreement between two measurement  
51  
52 systems: An alternative to the limits of agreement approach. *Stat Methods Med Res*  
53  
54 2017;26(6):2487-504. doi: 10.1177/0962280215601133  
55  
56  
57  
58  
59  
60

- 1  
2  
3 7. Lin L. I.-K. Total deviation index for measuring individual agreement with applications  
4 in laboratory performance and bioequivalence. *Statistics in Medicine* 2000;255-70.  
5  
6  
7
- 8 8. Haber M, Barnhart HX. A general approach to evaluating agreement between two  
9 observers or methods of measurement from quantitative data with replicated measurements.  
10  
11 *Stat Methods Med Res* 2008;17(2):151-69.  
12  
13
- 14 9. Li C. Little's Test of Missing Completely at Random. *The Stata Journal* 2013. p. 795-  
15 809. doi: 10.1177/1536867X1301300407  
16  
17  
18  
19
- 20 10. Lu MJ, Zhong WH, Liu YX, et al. Sample Size for Assessing Agreement between Two  
21 Methods of Measurement by Bland-Altman Method. *Int J Biostat* 2016;12(2). doi: 10.1515/ijb-  
22 2015-0039  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

# BMJ Paediatrics Open

## Multi-center paired non-inferiority study of the cardiorespiratory monitoring performance of the wireless and non-adhesive Bambi® belt measuring diaphragm activity in neonates: Study Protocol

Journal:	<i>BMJ Paediatrics Open</i>
Manuscript ID	bmjpo-2022-001430.R1
Article Type:	Protocol
Date Submitted by the Author:	21-Mar-2022
Complete List of Authors:	Scholten, Anouk; Amsterdam UMC Locatie AMC, Neonatology; Amsterdam UMC Locatie AMC, Amsterdam Reproduction & Development research institute Zhuozhao, Zhan; Eindhoven University of Technology, Department of Mathematics and Computer Science Niemarkt, Hendrik; Máxima Medisch Centrum, Neonatology Vervoorn, Marieke; Máxima Medisch Centrum, Neonatology van Leuteren, Ruud; Amsterdam UMC Locatie AMC, Neonatology; Amsterdam UMC Locatie AMC, Amsterdam Reproduction & Development research institute de Jongh, Frans; Amsterdam UMC Locatie AMC, Neonatology; University of Twente, Faculty of Science and Technology van Kaam, Anton; Amsterdam UMC Locatie AMC, Neonatology; Amsterdam UMC Locatie VUmc, Neonatology van den Heuvel, Edwin; Eindhoven University of Technology, Department of Mathematics and Computer Science Hutten, G. Jeroen; Amsterdam UMC Locatie AMC, Neonatology; Amsterdam UMC Locatie VUmc, Neonatology
Keywords:	Neonatology, Technology

SCHOLARONE™  
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1  
2  
3 **MULTI-CENTER PAIRED NON-INFERIORITY STUDY OF THE**  
4  
5 **CARDIORESPIRATORY MONITORING PERFORMANCE OF THE WIRELESS**  
6  
7 **AND NON-ADHESIVE BAMBI® BELT MEASURING DIAPHRAGM ACTIVITY IN**  
8  
9 **NEONATES: STUDY PROTOCOL**  
10

11  
12  
13  
14 Anouk W.J. Scholten<sup>1,2\*†</sup>, Zhan Zhuozhao<sup>3†</sup>, Hendrik J. Niemarkt<sup>4</sup>, Marieke Vervoorn<sup>4</sup>, Ruud  
15  
16 W. van Leuteren<sup>1,2</sup>, Frans H. de Jongh<sup>1,5</sup>, Anton H. van Kaam<sup>1,6</sup>, Edwin R. van den Heuvel<sup>3</sup>, G.  
17  
18 Jeroen Hutten<sup>1,6</sup>  
19

20  
21 *<sup>1</sup>Department of neonatology, Emma Children's Hospital, AmsterdamUMC, University of*  
22  
23 *Amsterdam, Amsterdam, The Netherlands*  
24

25  
26 *<sup>2</sup>Amsterdam Reproduction & Development research institute, Amsterdam, The Netherlands*  
27

28  
29 *<sup>3</sup>Department of Mathematics and Computer Science, Eindhoven University of Technology,*  
30  
31 *Eindhoven, The Netherlands*  
32

33  
34 *<sup>4</sup>Department of neonatology, Maxis Medical Center, Veldhoven, The Netherlands*  
35

36  
37 *<sup>5</sup>Faculty of Science and Technology, University of Twente, Enschede, The Netherlands*  
38

39  
40 *<sup>6</sup>Department of neonatology, Emma Children's Hospital, AmsterdamUMC, Vrije Universiteit,*  
41  
42 *Amsterdam, the Netherlands*  
43

44  
45 † These authors have contributed equally to this work and share first authorship

46  
47 **CORRESPONDING AUTHOR:**

48  
49 Anouk W.J. Scholten

50  
51 Department of Neonatology

52  
53 Emma Children's Hospital, AmsterdamUMC, location AMC

54  
55 Meibergdreef 9, 1105AZ Amsterdam, The Netherlands

56  
57 Phone: +31 20 5663675

58  
59 Email: a.w.j.scholten@amsterdamumc.nl  
60

## ABSTRACT

**Introduction:** Cardiorespiratory monitoring is used in the Neonatal Intensive Care Unit (NICU) to assess the clinical status of newborn infants and detect critical deteriorations in cardiorespiratory function. Currently, heart rate is monitored by electrocardiography (ECG) and respiration by chest impedance (CI). Disadvantages of current monitoring techniques are usage of wired adhesive electrodes which may damage the skin and hinder care. The Bambi® belt is a wireless and non-adhesive alternative that enables cardiorespiratory monitoring by measuring electrical activity of the diaphragm via transcutaneous electromyography (dEMG). A previous study showed feasibility of the Bambi belt and this study compares the belt performance to ECG and CI.

**Methods and analysis:** This multi-center non-inferiority paired study will be performed in the NICU of the Máxima Medical Center (MMC) in Veldhoven and the Emma Children's Hospital AmsterdamUMC in Amsterdam, the Netherlands. 39 infants in different postmenstrual age groups (minimally 10 infants <30 weeks, between 30-32 weeks and >32 weeks) will be recruited. These infants will be monitored with the Bambi® belt in addition to standard ECG and CI for 24 h. The primary outcome is the heart rate (HR), studied with three criteria: 1) the agreement in limits of agreement of the HR measurements in terms of the second-to-second difference in the HR between the belt and standard ECG, 2) the detection of cardiac events consisting of bradycardia and tachycardia and 3) the quality of HR-monitoring. The secondary outcome is the respiratory rate (RR), studied with the criteria 1) agreement in respiratory rate (RR)-RR trend monitoring, 2) apnea and tachypnea detection, and 3) reliable registrations.

**Ethics and dissemination:** This protocol was approved by the Medical Ethical Committee of the Máxima Medical Center and the Central Committee for Human Research (CCMO). The



1  
2  
3 MMC started patient recruitment in July and the AmsterdamUMC in August 2021. The results  
4  
5 will be presented at conferences and published in peer-reviewed journals.  
6  
7

8 **Trial registration number:** NL9480 (www.trialregister.nl)  
9

## 10 11 12 13 INTRODUCTION

14  
15 In the Neonatal Intensive Care Unit (NICU), cardiorespiratory monitoring is crucial to assess  
16  
17 clinical condition and to timely detect and treat frequently occurring cardio-respiratory events  
18  
19 to prevent morbidity and mortality.(1, 2) To date, this is performed by measuring the  
20  
21 electrocardiogram (ECG) and chest impedance (CI) with three wired adhesive electrodes. CI  
22  
23 measures variation in electrical impedance across the chest during respiration caused by  
24  
25 changes in lung aeration and chest wall movement. These techniques provide continuous  
26  
27 monitoring of heart rate (HR), respiratory rate (RR), and breathing pattern. However, as CI  
28  
29 measures respiration indirectly, adequate detection of breathing cycles and apnea may not  
30  
31 always be optimal.(3)  
32  
33  
34  
35  
36  
37

38 With transcutaneous electromyography of the diaphragm (dEMG) breathing effort can be  
39  
40 recorded directly by measuring the electrical activity of this main respiratory muscle. To date,  
41  
42 this technique also uses three adhesive electrodes and provides information on respiration and  
43  
44 HR. Studies have shown its feasibility in the NICU-setting.(4)  
45  
46  
47  
48

49 The use ~~and especially removal~~ of adhesive electrodes is restricted in infants with a  
50  
51 postmenstrual age <26 weeks in fear of skin damage. electrodes can lead to epidermal stripping  
52  
53 in vulnerable preterm infants, resulting in an increased risk of infection and pain.(5) Moreover,  
54  
55 electrode removal may cause discomfort. Furthermore, the wires attached to the electrodes  
56  
57 restrict movements of the infant and may hinder parent-infant interaction, both nursing and  
58  
59  
60

1  
2  
3 kangaroo care. Restrictions in kangaroo care may impact patient outcome as it has been  
4 associated with beneficial effects such as decreased mortality, decreased risk of severe  
5 infection/sepsis and hypothermia, and increased likelihood of exclusive breast feeding.(6, 7)  
6  
7

8 All things considered, Therefore, it is important to find alternatives for using wired adhesive  
9 electrodes.  
10  
11  
12  
13  
14  
15  
16

17 In the past years, several wireless wearable sensors have been developed to measure various  
18 parameters in neonates such as ECG, HR, RR, peripheral oxygen saturation and (skin)  
19 temperature.(8-14) Recently, a novel wireless and non-adhesive sensor belt (Bambi® belt,  
20 Bambi B.V., Eindhoven, the Netherlands) was developed for neonatal use that measures ECG  
21 and respiration based on the dEMG technique. A recent pilot study showed that measuring HR  
22 and RR with this belt in preterm infants is feasible and that the measured HR and RR trend was  
23 similar to ECG and CI.(submitted for publication, NICU AmsterdamUMC, 2021) However,  
24 before replacing the current techniques using adhesive wired electrodes with the non-adhesive  
25 sensor belt, a larger study is required to demonstrate the non-inferiority of this belt as an  
26 alternative cardiorespiratory monitor. In this study, we compare the monitoring performance of  
27 the Bambi® belt to ECG and CI and hypothesize that the performance of the belt is non-inferior  
28 to the current monitoring techniques.  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46

## 47 **METHODS**

### 48 **Study design**

49 This multi-center paired non-inferiority study will be performed in the NICU of Máxima  
50 Medical Center (MMC) in Veldhoven and the Emma Children's Hospital of the Amsterdam  
51 University Medical Centre (AmsterdamUMC), both located in the Netherlands. Each patient  
52 will be simultaneously measured with the belt and ECG/CI (paired design). To compare the  
53 devices, a non-inferiority/equivalence framework will be used. Here, equivalence is defined as  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 the limit of agreement of the HR/RR between the belt and ECG/CI being within prespecified  
4  
5 ~~clinically accepted~~ margins (see Table 1 for the margins). Non-inferiority is defined as the  
6  
7 performance of clinical event detection and quality criteria not being worse than prespecified  
8  
9  
10 ~~clinically accepted~~ margins.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Confidential: For Review Only

## Study population

Preterm and term infants being routinely monitored with the standard cardiorespiratory monitor (Intellivue MP90, Philips Healthcare, Eindhoven, The Netherlands) are included in the study.

To ensure a representative sample of the target population, infants in different age groups will be included. Infants with chest skin lesions, congenital anomalies, and other scenario's preventing belt placement, such as (effects of) surgery or wrap for therapeutic hypothermia, will be excluded.

## Primary outcome

As HR-monitoring is clinically most relied upon and both ECG and dEMG provide the HR by measuring cardiac electrical activity, while CI and dEMG measure respiration with a different technique, the HR is considered the primary outcome.(3, 15) This will be studied with three criteria, which will be compared to the prespecified margins in Table 1. 1) Reliable monitoring performance through second-to-second HR measurement agreement in terms of differences in measured HR -between the belt and the ECG/-CI monitoring. 2) The detection of a composite cardiac event consisting of bradycardia (HR < 100 beats per minute for at least five seconds)(16) and tachycardia (HR > 180 beats per minute for at least ten seconds)(17) between the belt and the ECG measured with adhesive electrodes. The minimal duration of a bradycardia or tachycardia will prevent the inclusion of technical errors (short drops or increases in the HR) in our analysis and is lower for bradycardia compared to tachycardia as bradycardias are shorter events.(1) The thresholds are empirically chosen to detect all low and high HR-values. 3) Non-inferior quality (percentage of time with HR recordings without data loss).

Moreover, we will perform subgroup analyses to investigate whether the HR measurement performance is consistent under different clinical activities (e.g. kangaroo care, feeding) and in the different age groups.

## Secondary outcomes

The secondary outcome is the measured RR. This will be studied using the following three criteria, which will be compared to the prespecified margins in Table 1:

- 1) Comparing the trend (~~10-minute moving average~~) in RR values provided by the belt and CI, based on the difference in the 10-minute moving averages. The RR-trend is studied as this as the trend is used in the clinical practice to detect for example increases in RR over time as a marker of clinical deterioration of a patient is the primary clinical usage of respiratory monitoring.(3) Since CI is widely used for neonatal respiratory monitoring, it is used as the reference technique.
- 2) Next to comparing the RR-trend, Comparing the ability to detect apnea and tachypnea is studied as the detection of these critical respiratory events based on RR is another purpose of the respiratory monitoring. Clinically relevant apneas are considered when indicated by a RR < 20 breaths per minute (~~to capture all periods of low breathing frequency~~) measured with CI for at least 10 seconds, associated with a desaturation (arterial oxygen saturation as measured by pulse oximetry (SpO<sub>2</sub>) <80% for at least 10 seconds) and/or bradycardia (HR <100 beats per minute for at least five seconds) (objective apnea measurement).(16) A RR<20 breaths per minute is chosen for the apnea definition as we solely use the numerical RR-values, because despite the two different measurement techniques this endpoint is equal, and to capture all periods of low breathing frequency.

Tachypnea is defined as a prolonged period of the averaged (moving average with a window size of 10 minutes) RR >60 breaths per minute and >100 breaths per minute (approximately two times the average normal RR).(18) To cover short and long periods of tachypnea, 3 different durations are studied (30 seconds, 60 seconds, and 10 minutes).

- 1  
2  
3 3) Calculating the percentage of time with reliable respiratory monitoring (without data  
4  
5 loss and with an acceptable signal-to-noise ratio).  
6  
7  
8  
9

### 10 **Data collection**

11 The following basic characteristics and demographic information will be collected at the  
12 baseline of the study: gestational age, birth weight, gender, age and weight at day of  
13 measurement, relevant medical status (respiratory support, medication and underlying illness  
14 during measurement), chest circumference, nipple distance, skin type at study start by visual  
15 inspection (normal, dry, flaky, oily, moist, other).  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25

### 26 **Sample size calculation**

27 A power calculation is performed for the primary outcome using data collected in a previous  
28 study.(submitted for publication, NICU AmsterdamUMC, 2021) Among the three criteria,  
29 criteria 1 needs the largest sample size and is used for our study. This resulted in 39 required  
30 infants to achieve 80% power with an overall 5% type I error with a Bonferroni correction  
31 (details in the Statistical Analysis Plan (SAP) in the online supplement).  
32  
33  
34  
35  
36  
37  
38

39 In addition, an interim analysis will be performed as the power calculation was based on the  
40 previous study and recruitment of infants without being able to answer research questions is  
41 unethical.(19) This will be performed after including 1/3th of the infants for sample size  
42 adaption using the method of Mehta and Pocock.(20) If the conditional power falls within the  
43 pre-defined “promising zone”, the sample size will be increased to an upper limit of 52 infants.  
44  
45  
46  
47  
48  
49  
50  
51 Otherwise, the study will proceed with the original sample size.

52  
53 To ensure that a representative sample of the age distribution of infants at a NICU, infants in  
54 different postmenstrual age groups will be recruited with the same proportions as in the target  
55 population (minimally 10 infants <30 weeks, between 30-32 weeks and >32 weeks).  
56  
57  
58  
59  
60

## Study procedures

The Bambi® **bBelt sSystem** is a non CE-certified medical device, designed for wireless cardiorespiratory monitoring of (pre)term infants in a hospital environment. All included infants will be monitored with the belt in addition to standard ECG/CI for 24 hours to obtain representative clinical scenarios throughout the entire day. The measurement set-up is visualised in Figure 1 and consists of 1) dEMG measurement with the belt and 2) the extraction of patient monitor data.

In the belt, three dry electrodes are incorporated. When placing the belt at the height of the diaphragm, the outer two electrodes are in the nipple line and the middle electrode is in line with the sternum. The three ECG/CI electrodes are attached at the original location without hindering belt placement. The measured electrical signal of the diaphragm with the belt is wirelessly transmitted to the Receiver Module (REM) by the Sensor Module (SEM). The REM processes the dEMG signal to obtain the ECG and respiration signal (averaged diaphragmatic activity). An inbuilt algorithm provides the HR and RR out of the ECG and respiration signal respectively. This data is transported to a bedside computer. The data from the patient monitor (ECG, HR, RR, and SpO<sub>2</sub>) is extracted from the bedside monitor using an isolated cable and is also transported to the bedside computer.

The belt data from the REM and patient monitor are recorded and synchronised using a dedicated software package (Polybench, Applied Biosignals, Weener, Germany) on a personal bedside computer. Data is recorded at a sample rate of 1 to 500 Hz for rate and waveform data respectively. The bedside software also provides the possibility to make measurements annotations by nurses and researchers during data recording, such as re-positioning of the infant, nursing and kangaroo care.

1  
2  
3 During the study, daily routine care proceeds as usual. The location of the belt is regularly  
4 checked and if necessary repositioned (similar to the clinical practice). Notifications are  
5 visualised when contact between skin and the belt is lost (Leads off) or when there is no  
6 connection between the SEM and REM (Bluetooth Loss Error). In case of the first notification,  
7 the belt may be repositioned, while in case of Bluetooth loss the battery level of the SEM or  
8 blocking of this sensor (e.g. by an arm) are checked.  
9

10  
11  
12  
13  
14  
15  
16  
17  
18  
19 Preferably, the belt stays in place during the study. However, the belt can be removed during  
20 diagnostic imaging, patient handling, or in case of skin irritation at the belt location. The reason  
21 for removal will be annotated. If the belt is removed, the medical staff, parents and one of the  
22 dedicated researchers will decide together if the belt can be re-applied.  
23  
24  
25  
26  
27  
28  
29

### 30 **Recruitment**

31  
32 Parents of all eligible infants are approached for consent to obtain a sample as heterogeneous  
33 and representative as possible. Preferably, infants are included as soon as possible after birth.  
34 During the 24 hours, the study can be terminated if requested by parents or the treating  
35 physicians. In case of withdrawal of a subject, an extra subject will be included.  
36  
37  
38  
39  
40  
41  
42  
43

### 44 **Safety**

45 Being a medical device study, this study was classified as a moderate risk.(21) A specified  
46 monitor plan for the study is made based on risk-classification.  
47  
48  
49  
50  
51

### 52 **STATISTICAL ANALYSIS**

53  
54 A detailed SAP can be found in the online supplement. Unless otherwise specified, all  
55 hypothesis tests are two-sided with a significance level of 0.05. All statistical analyses will be  
56  
57  
58  
59  
60



1  
2  
3 performed using R version 4.0 (the R Foundation for Statistical Computing; Vienna, Austria)  
4  
5 and SAS software version 9.4 (SAS Institute Inc., Cary, NC, USA).  
6  
7  
8  
9

10 The non-inferiority/equivalence margins based on expert opinions (survey send to  
11 neonatologists of different NICU's in the Netherlands) and literature (4, 22, 23) are described  
12  
13 in Table 1. In the different subparagraphs we refer to this table.  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 1** - The non-inferiority/equivalence margins for the primary and secondary outcomes

<i>Endpoints</i>	<i>acceptance-Prespecified margins</i> <sup>#</sup>
<i>LOA of second-to-second HR differences</i>	$\pm 8$ bpm
<i>LOA of RR trend differences</i>	$\pm 15$ brpm
<i>Sensitivity of brady-/tachycardia detection</i>	90% <sup>°</sup>
<i>PPV of brady-/tachycardia detection</i>	90% <sup>°</sup>
<i>Sensitivity of apnea/tachypnea detection</i>	70%
<i>PPV of apnea/tachypnea alarms</i>	0-100%*
<i>Data loss percentage</i>	5%
<i>Robust data percentage (HR)</i>	90%
<i>Robust data percentage (RR)</i>	70%

LOA: limits of agreement, HR: heart rate, RR: respiratory rate, PPV: positive predictive value.

Data loss is defined as the percentage of data without “Leads off” or “Bluetooth Loss Error” in the belt.

<sup>#</sup> The acceptance-prespecified margins are compared to confidence intervals with corresponding confidence levels (see SAP for more details).

<sup>°</sup>Note: all missed bradycardias are checked for clinical relevance by two independent experts.

\*Since the reference devices for apnea detection in the clinical practice are the peripheral oxygen saturation (SpO<sub>2</sub>) and electrocardiogram instead of the respiration signal and the performance for Chest Impedance to detect tachypnea is unsatisfactory due to the presence of cardiac interference, all values for PPV for apnea/tachypnea are acceptable. Interpretations will be made

## Summary and descriptive statistics

Categorical data will be summarized by numbers of counts and percentages. Continuous data will be summarized by mean, standard deviation if data is normal and median, interquartile range (IQR) if data is skewed. Minimum and maximum values will also be presented for continuous data when appropriate.

## Statistical analysis of the primary outcome

### Criteria 1: agreement in HR

To investigate the equivalence of HR measurement between the belt and ECG, we will fit a linear mixed model to the second-to-second HR difference between both. With this model, the 95% limits of agreement (Bland-Altman analysis) will be derived. The two-one-sided tests (TOST) with a multiplicity corrected alpha of 0.0167 and the prespecified margin ( $\pm 8$ bpm) will test equivalence between the two devices. In addition, based on a bivariate heteroscedastic model fitted to HR segments of a prespecified length, additional performance measures will be calculated as sensitivity analyses (details in SAP).

### Criteria 2: cardiac event detection

For HR monitoring, we also consider the detection of bradycardia and tachycardia. We will estimate the sensitivity and the positive predictive value (PPV) of the belt using the patient monitor data as the ground truth and perform a non-inferiority test with an alpha of 0.0167. The non-inferiority margin for the sensitivity and PPV are listed in Table 1. In case of missed bradycardias, one independent expert per center will qualify the ~~severity~~ safety and clinical consequences and acceptability of each missing event by answering the same questions per figure containing the ~~with the use of the~~ discrepancy in HR and the ECG-signals measured with CI and the belt and contextual information. ~~- These figures will be blinded and thus it will be unknown which signal corresponds CI or the belt.~~

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Confidential: For Review Only

### Criteria 3: signal quality

The quality of the investigational device will be quantified based on the percentage of time during the 24-hour period it produces any reading (percentage without data loss due to “Leads off” or “Bluetooth Loss Error”) and the percentage in time it produces a good-quality-reading (percentage of robust data) for the HR and RR, respectively. For the HR non-robust data can be caused by bad connection (suboptimal Bluetooth or skin-electrode connection). These criteria are built-in in the belt algorithm and therefore this data is automatically labeled. Hypothesis testing will be used to establish the non-inferiority of this “uptime” percentage (percentage without data loss and percentage of robust data) of the belt.

For the RR, the uptime percentage is also categorized as a) data readings without data loss and b) robust data readings, ~~i.e. good quality readings with acceptable signal-to-noise ratio's undisturbed by among others suboptimal connection and patient handling~~ i.e. readings without unrealistic (e.g. negative) values. Signal quality is only analyzed for the belt. However, these results are compared to prespecified margins. Again, the prespecified margins for the HR and RR are described in Table 1. As the HR monitored with CI is accurate and nearly continuous, while the RR is less relied upon and may be unreliable, the prespecified margin for the RR is lower than for the HR.

### Statistical analysis of secondary outcomes

Secondary analyses, based on the same statistical methods for the criteria of the primary outcome, include all secondary endpoints (apnea and tachypnea detection, RR trend analysis (see SAP)) and evaluation during different scenarios.

### ETHICS AND DISSEMINATION

The Medical Ethical Committee of the MMC (W21.042) and the Central Committee for Human Research in the Netherlands (CCMO, CCMO21/0167/PP) approved the study protocol (Version

1  
2  
3 2, 19<sup>th</sup> of May 2021). Local feasibility at the AmsterdamUMC was approved by the Medical  
4 Ethical Committee of the AMC (2021\_146). This study was registered in the Dutch Trial  
5 Register (<https://www.trialregister.nl>, NL9480). Regarding patient safety, no belt related events  
6 were observed in the pilot study and are therefore unexpected. Moreover, as every patient is  
7 monitored with ECG/CI and the belt, safety is guaranteed in case of missing belt data. The SAP  
8 will be used for the analyses. The results will be published in peer-reviewed journals and  
9 presented at future congresses.

10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20 The MMC started patient recruitment in July and the AmsterdamUMC in August 2021. The  
21 duration of this study will be approximately seven months.  
22  
23  
24  
25  
26  
27

## 28 **PATIENT AND PUBLIC INVOLVEMENT STATEMENT**

29  
30  
31 Patients were included in this study after obtaining parental informed consent. The patients  
32 could not be involved in the design, recruitment, conduction and dissemination of results of this  
33 study. Neither could we ask the burden of the study. The outcome measures were developed by  
34 combining clinical and statistical knowledge to ensure a SAP that enables confirmation of non-  
35 inferiority of the belt compared to ECG/CI.  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45

## 46 **REFERENCES**

- 47  
48 1. Joshi R, van Straaten H, van de Mortel H, et al. Does the architectural layout of a NICU  
49 affect alarm pressure? A comparative clinical audit of a single-family room and an open bay  
50 area NICU using a retrospective study design. *BMJ Open* 2018;8(6):e022813. doi:  
51 10.1136/bmjopen-2018-022813  
52  
53  
54  
55  
56  
57  
58 2. [Poets CF, Roberts RS, Schmidt B, Whyte RK, Asztalos E V, et al. Association Between](#)  
59 [Intermittent Hypoxemia or Bradycardia and Late Death or Disability in Extremely Preterm](#)  
60

1  
2  
3 [Infants. \*JAMA\*. \(2015\) 314\(6\):595–603. <https://doi.org/10.1001/jama.2015.8841>](https://doi.org/10.1001/jama.2015.8841)  
4 ~~Bouwstra S,~~  
5 ~~Chen W, Feijs LMG, et al. Designing the parents-to-infant bonding experience in the NICU.~~  
6 ~~*IARIA* 2012:7-13.~~  
7  
8  
9

10  
11 3.

12  
13  
14 3. Di Fiore JM. Neonatal cardiorespiratory monitoring techniques. *Semin Neonatol*  
15 2004;9(3):195-203. doi: 10.1016/j.siny.2003.11.009  
16  
17

18  
19 4. Kraaijenga JV, Hutten GJ, de Jongh FH, et al. Transcutaneous electromyography of the  
20 diaphragm: A cardio-respiratory monitor for preterm infants. *Pediatr Pulmonol*  
21 2015;50(9):889-95. doi: 10.1002/ppul.23116  
22  
23  
24

25  
26 5. Csoma ZR, Meszes A, Abraham R, et al. Iatrogenic Skin Disorders and Related Factors  
27 in Newborn Infants. *Pediatr Dermatol* 2016;33(5):543-8. doi: 10.1111/pde.12960  
28  
29

30  
31 6. Boundy EO, Dastjerdi R, Spiegelman D, et al. Kangaroo Mother Care and Neonatal  
32 Outcomes: A Meta-analysis. *Pediatrics* 2016;137(1). doi: 10.1542/peds.2015-2238  
33  
34  
35

36  
37 7. Conde-Agudelo A, Diaz-Rossello JL. Kangaroo mother care to reduce morbidity and  
38 mortality in low birthweight infants. *Cochrane Database Syst Rev* 2016(8). doi:  
39 <https://doi.org/10.1002/14651858.CD002771.pub4>  
40  
41  
42

43  
44 8. Chung HU, Rwei AY, Hourlier-Fargette A, et al. Skin-interfaced biosensors for  
45 advanced wireless physiological monitoring in neonatal and pediatric intensive-care units. *Nat*  
46 *Med* 2020;26(3):418-29. doi: <https://doi.org/10.1038/s41591-020-0792-9>  
47  
48  
49

50  
51 9. Henry C, Shipley L, Ward C, et al. Accurate neonatal heart rate monitoring using a new  
52 wireless, cap mounted device. *Acta Paediatr* 2021;110(1):72-8. doi:  
53 <https://doi.org/10.1111/apa.15303>  
54  
55  
56  
57  
58  
59  
60

10. Rimet Y, Brusquet Y, Ronayette D, et al. Surveillance of infants at risk of apparent life threatening events (ALTE) with the BBA bootee: a wearable multiparameter monitor. *Annu Int Conf IEEE Eng Med Biol Soc* 2007;2007:4997-5000. doi: 10.1109/IEMBS.2007.4353462
11. Chen H, Bao S, Lu C, et al. Design of an Integrated Wearable Multi-Sensor Platform Based on Flexible Materials for Neonatal Monitoring. *IEEE* 2020:23732-47. doi: 10.1109/ACCESS.2020.2970469
12. Chen W, Dols S, Bambang Oetomo S, et al. Monitoring body temperature of newborn infants at neonatal intensive care units using wearable sensors. *BodyNets '10* 2010. doi: <https://doi.org/10.1145/2221924.2221960>
13. Linti C, Horter H, Osterreicher P, et al. Sensory baby vest for the monitoring of infants. *IEEE* 2006. doi: 10.1109/BSN.2006.49
14. Pike H, Eilevstjonn J, Bjorland P, et al. Heart rate detection properties of dry-electrode ECG compared to conventional 3-lead gel-electrode ECG in newborns. *BMC Res Notes* 2021;14(1):166. doi: 10.1186/s13104-021-05576-x
15. van Leuteren RW, Hutten GJ, de Waal CG, et al. Processing transcutaneous electromyography measurements of respiratory muscles, a review of analysis techniques. *Journal of Electromyography and Kinesiology* 2019:176-86. doi: 10.1016/j.jelekin.2019.07.014
16. Lee H, Rusin CG, Lake DE, et al. A new algorithm for detecting central apnea in neonates. *Physiol Meas* 2012;33(1):1-17. doi: 10.1088/0967-3334/33/1/1
17. Li T, Matsushima M, Timpson W, et al. Epidemiology of patient monitoring alarms in the neonatal intensive care unit. *J Perinatol* 2018;38(8):1030-8. doi: 10.1038/s41372-018-0095-x



- 1  
2  
3 18. Reuter S, Moser C, Baack M. Respiratory distress in the newborn. *Pediatr Rev*  
4 2014;35(10):417-28. doi: 10.1542/pir.35-10-417  
5  
6  
7  
8 19. European Medicines Agency. ICH Topic E 9 – Statistical Principles for Clinical Trials.  
9 2006:21-3.  
10  
11  
12  
13 20. Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are  
14 promising: a practical guide with examples. *Stat Med* 2011;30(28):3267-84. doi:  
15 <https://doi.org/10.1002/sim.4102>  
16  
17  
18  
19  
20  
21 21. Central Committee on Research Involving Human Subjects.  
22  
23  
24 22. Anton O, Fernandez R, Rendon-Morales E, et al. Heart Rate Monitoring in Newborn  
25 Babies: A Systematic Review. *Neonatology* 2019;116(3):199-210. doi: 10.1159/000499675  
26  
27  
28  
29 23. Bretelet MJM, KleinJan EJ, Dohmen DAJ, et al. Vital Signs Monitoring with Wearable  
30 Sensors in High-risk Surgical Patients: A Clinical Validation Study. *Anesthesiology*  
31 2020;132(3):424-39. doi: 10.1097/ALN.0000000000003029  
32  
33  
34  
35  
36  
37  
38

### 39 AUTHOR CONTRIBUTIONS

40 AS, HN, MV, RL, FJ, AK, JH conceptualized the study. ZZ and EH made the statistical analysis  
41 plan, which was reviewed by all authors. AS wrote the first version of this manuscript. All  
42 authors contributed to the final draft of the manuscript.  
43  
44  
45  
46  
47  
48  
49

### 50 FUNDING STATEMENT

51 This work was supported by the Louise Vehmeijer foundation.  
52  
53  
54  
55  
56

### 57 COMPETING INTERESTS STATEMENT

58  
59  
60

1  
2  
3 The authors declare that the research was conducted in the absence of any commercial or  
4 ~~financial~~ relationships that could be construed as a potential conflict of interest. However, the  
5 MMC and AmsterdamUMC do receive a prespecified amount of money per performed  
6 measurement. The results of this study protocol will contribute to obtaining a CE-mark for the  
7 Bambi® belt.  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Confidential: For Review Only

**WHAT IS KNOWN ABOUT THE SUBJECT:**

- Disadvantages of the cardiorespiratory monitoring technique in neonates are indirect measurements of respiration, usage of adhesive electrodes and hindering wires.
- With transcutaneous electromyography of the diaphragm, respiratory activity is measured directly by recording the activity of the main respiratory muscle.
- The Bambi® belt is a novel wireless and non-adhesive belt that enables cardiorespiratory monitoring by measuring diaphragm activity with dry electrodes.

**WHAT THIS STUDY HOPES TO ADD:**

- Demonstration of the non-inferiority of the Bambi® belt compared to the electrocardiogram and chest impedance for cardiorespiratory monitoring in preterm and term infants.

**HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY**

- When non-inferiority of the Bambi® belt compared to the current cardiorespiratory monitor is confirmed, the belt could be used as a wireless and skin-friendly alternative.

## FIGURES

**FIGURE 1** - The measurement set-up. The adhesive electrodes used for standard cardiorespiratory monitoring are attached at the original location, visualised by the three blue dots. The diaphragm activity measured with the Bambi® belt is wirelessly transmitted with the Sensor Module to the Receiver Module where the data is processed to obtain an electrocardiogram and respiration waveform (and heart rate and respiratory rate). This data and the data measured with the patient monitor are transported to a personal bedside computer with Polybench software to synchronise and record these signals.

1  
2  
3 **MULTI-CENTER PAIRED NON-INFERIORITY STUDY OF THE**  
4  
5 **CARDIORESPIRATORY MONITORING PERFORMANCE OF THE WIRELESS**  
6  
7 **AND NON-ADHESIVE BAMBI® BELT MEASURING DIAPHRAGM ACTIVITY IN**  
8  
9 **NEONATES: STUDY PROTOCOL**  
10  
11

12  
13  
14 Anouk W.J. Scholten<sup>1,2\*†</sup>, Zhan Zhuozhao<sup>3†</sup>, Hendrik J. Niemarkt<sup>4</sup>, Marieke Vervoorn<sup>4</sup>, Ruud  
15  
16 W. van Leuteren<sup>1,2</sup>, Frans H. de Jongh<sup>1,5</sup>, Anton H. van Kaam<sup>1,6</sup>, Edwin R. van den Heuvel<sup>3</sup>, G.  
17  
18 Jeroen Hutten<sup>1,6</sup>  
19  
20

21 *<sup>1</sup>Department of neonatology, Emma Children's Hospital, AmsterdamUMC, University of*  
22  
23 *Amsterdam, Amsterdam, The Netherlands*  
24  
25

26 *<sup>2</sup>Amsterdam Reproduction & Development research institute, Amsterdam, The Netherlands*  
27

28 *<sup>3</sup>Department of Mathematics and Computer Science, Eindhoven University of Technology,*  
29  
30 *Eindhoven, The Netherlands*  
31

32 *<sup>4</sup>Department of neonatology, Maxis Medical Center, Veldhoven, The Netherlands*  
33  
34

35 *<sup>5</sup>Faculty of Science and Technology, University of Twente, Enschede, The Netherlands*  
36  
37

38 *<sup>6</sup>Department of neonatology, Emma Children's Hospital, AmsterdamUMC, Vrije Universiteit,*  
39  
40 *Amsterdam, the Netherlands*  
41

42 † These authors have contributed equally to this work and share first authorship  
43

44 **CORRESPONDING AUTHOR:**  
45

46 Anouk W.J. Scholten  
47

48 Department of Neonatology  
49

50 Emma Children's Hospital, AmsterdamUMC, location AMC  
51

52 Meibergdreef 9, 1105AZ Amsterdam, The Netherlands  
53

54 Phone: +31 20 5663675  
55

56 Email: a.w.j.scholten@amsterdamumc.nl  
57  
58  
59  
60

## ABSTRACT

**Introduction:** Cardiorespiratory monitoring is used in the Neonatal Intensive Care Unit (NICU) to assess the clinical status of newborn infants and detect critical deteriorations in cardiorespiratory function. Currently, heart rate is monitored by electrocardiography (ECG) and respiration by chest impedance (CI). Disadvantages of current monitoring techniques are usage of wired adhesive electrodes which may damage the skin and hinder care. The Bambi® belt is a wireless and non-adhesive alternative that enables cardiorespiratory monitoring by measuring electrical activity of the diaphragm via transcutaneous electromyography (dEMG). A previous study showed feasibility of the Bambi belt and this study compares the belt performance to ECG and CI.

**Methods and analysis:** This multi-center non-inferiority paired study will be performed in the NICU of the Máxima Medical Center (MMC) in Veldhoven and the Emma Children's Hospital AmsterdamUMC in Amsterdam, the Netherlands. 39 infants in different postmenstrual age groups (minimally 10 infants <30 weeks, between 30-32 weeks and >32 weeks) will be recruited. These infants will be monitored with the Bambi® belt in addition to standard ECG and CI for 24 h. The primary outcome is the heart rate (HR), studied with three criteria: 1) the limits of agreement of the HR measurements in terms of the second-to-second difference in the HR between the belt and standard ECG, 2) the detection of cardiac events consisting of bradycardia and tachycardia and 3) the quality of HR-monitoring. The secondary outcome is the respiratory rate (RR), studied with the criteria 1) agreement in RR trend monitoring, 2) apnea and tachypnea detection, and 3) reliable registrations.

**Ethics and dissemination:** This protocol was approved by the Medical Ethical Committee of the Máxima Medical Center and the Central Committee for Human Research (CCMO). The

1  
2  
3 MMC started patient recruitment in July and the AmsterdamUMC in August 2021. The results  
4  
5 will be presented at conferences and published in peer-reviewed journals.  
6  
7

8 **Trial registration number:** NL9480 (www.trialregister.nl)  
9  
10

## 11 INTRODUCTION

12  
13 In the Neonatal Intensive Care Unit (NICU), cardiorespiratory monitoring is crucial to assess  
14  
15 clinical condition and to timely detect and treat frequently occurring cardio-respiratory events  
16  
17 to prevent morbidity and mortality.(1, 2) To date, this is performed by measuring the  
18  
19 electrocardiogram (ECG) and chest impedance (CI) with three wired adhesive electrodes. CI  
20  
21 measures variation in electrical impedance across the chest during respiration caused by  
22  
23 changes in lung aeration and chest wall movement. These techniques provide continuous  
24  
25 monitoring of heart rate (HR), respiratory rate (RR), and breathing pattern. However, as CI  
26  
27 measures respiration indirectly, adequate detection of breathing cycles and apnea may not  
28  
29 always be optimal.(3)  
30  
31  
32  
33  
34  
35  
36  
37

38 With transcutaneous electromyography of the diaphragm (dEMG) breathing effort can be  
39  
40 recorded directly by measuring the electrical activity of this main respiratory muscle. To date,  
41  
42 this technique also uses three adhesive electrodes and provides information on respiration and  
43  
44 HR. Studies have shown its feasibility in the NICU-setting.(4)  
45  
46  
47  
48

49 The use of adhesive electrodes is restricted in infants with a postmenstrual age <26 weeks in  
50  
51 fear of skin damage.(5) Moreover, electrode removal may cause discomfort. Furthermore, the  
52  
53 wires attached to the electrodes restrict movements of the infant and may hinder parent-infant  
54  
55 interaction, nursing and kangaroo care. Restrictions in kangaroo care may impact patient  
56  
57 outcome as it has been associated with beneficial effects such as decreased mortality, decreased  
58  
59  
60

1  
2  
3 risk of severe infection/sepsis and hypothermia, and increased likelihood of exclusive breast  
4 feeding.(6, 7) All things considered, it is important to find alternatives for using wired adhesive  
5 electrodes.  
6  
7  
8  
9

10  
11  
12 In the past years, several wireless wearable sensors have been developed to measure various  
13 parameters in neonates such as ECG, HR, RR, peripheral oxygen saturation and (skin)  
14 temperature.(8-14) Recently, a novel wireless and non-adhesive sensor belt (Bambi® belt,  
15 Bambi B.V., Eindhoven, the Netherlands) was developed for neonatal use that measures ECG  
16 and respiration based on the dEMG technique. A recent pilot study showed that measuring HR  
17 and RR with this belt in preterm infants is feasible and that the measured HR and RR trend was  
18 similar to ECG and CI.(submitted for publication, NICU AmsterdamUMC, 2021) However,  
19 before replacing the current techniques using adhesive wired electrodes with the non-adhesive  
20 sensor belt, a larger study is required to demonstrate the non-inferiority of this belt as an  
21 alternative cardiorespiratory monitor. In this study, we compare the monitoring performance of  
22 the Bambi® belt to ECG and CI and hypothesize that the performance of the belt is non-inferior  
23 to the current monitoring techniques.  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41

## 42 **METHODS**

### 43 **Study design**

44  
45 This multi-center paired non-inferiority study will be performed in the NICU of Máxima  
46 Medical Center (MMC) in Veldhoven and the Emma Children's Hospital of the Amsterdam  
47 University Medical Centre (AmsterdamUMC), both located in the Netherlands. Each patient  
48 will be simultaneously measured with the belt and ECG/CI (paired design). To compare the  
49 devices, a non-inferiority/equivalence framework will be used. Here, equivalence is defined as  
50 the limit of agreement of the HR/RR between the belt and ECG/CI being within prespecified  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 margins (see Table 1 for the margins). Non-inferiority is defined as the performance of clinical  
4 event detection and quality criteria not being worse than prespecified margins.  
5  
6  
7  
8  
9  
10

### 11 **Study population**

12 Preterm and term infants being routinely monitored with the standard cardiorespiratory monitor  
13 (Intellivue MP90, Philips Healthcare, Eindhoven, The Netherlands) are included in the study.  
14  
15 To ensure a representative sample of the target population, infants in different age groups will  
16 be included. Infants with chest skin lesions, congenital anomalies, and other scenario's  
17 preventing belt placement, such as (effects of) surgery or wrap for therapeutic hypothermia,  
18 will be excluded.  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29

### 30 **Primary outcome**

31 As HR-monitoring is clinically most relied upon and both ECG and dEMG provide the HR by  
32 measuring cardiac electrical activity, while CI and dEMG measure respiration with a different  
33 technique, the HR is considered the primary outcome.(3, 15) This will be studied with three  
34 criteria, which will be compared to the prespecified margins in Table 1. 1) Reliable monitoring  
35 performance through second-to-second HR measurement agreement in terms of differences in  
36 measured HR between the belt and the ECG/CI monitoring. 2) The detection of a composite  
37 cardiac event consisting of bradycardia (HR < 100 beats per minute for at least five  
38 seconds)(16) and tachycardia (HR > 180 beats per minute for at least ten seconds)(17) between  
39 the belt and the ECG measured with adhesive electrodes. The minimal duration of a bradycardia  
40 or tachycardia will prevent the inclusion of technical errors (short drops or increases in the HR)  
41 in our analysis and is lower for bradycardia compared to tachycardia as bradycardias are shorter  
42 events.(1) The thresholds are empirically chosen to detect all low and high HR-values. 3) Non-  
43 inferior quality (percentage of time with HR recordings without data loss).  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Moreover, we will perform subgroup analyses to investigate whether the HR measurement performance is consistent under different clinical activities (e.g. kangaroo care, feeding) and in the different age groups.

### Secondary outcomes

The secondary outcome is the measured RR. This will be studied using the following three criteria, which will be compared to the prespecified margins in Table 1:

- 1) Comparing the trend in RR values provided by the belt and CI, based on the difference in the 10-minute moving averages. The RR-trend is studied as this is used in the clinical practice to detect for example increases in RR over time as a marker of clinical deterioration of a patient.(3) Since CI is widely used for neonatal respiratory monitoring, it is used as the reference technique.
- 2) Next to comparing the RR-trend, the ability to detect apnea and tachypnea is studied as the detection of these critical respiratory events based on RR is another purpose of the respiratory monitoring. Clinically relevant apneas are considered when indicated by a RR < 20 breaths per minute measured with CI for at least 10 seconds, associated with a desaturation (arterial oxygen saturation as measured by pulse oximetry (SpO<sub>2</sub>) <80% for at least 10 seconds) and/or bradycardia (HR <100 beats per minute for at least five seconds) (objective apnea measurement).(16) A RR<20 breaths per minute is chosen for the apnea definition as we solely use the numerical RR-values, because despite the two different measurement techniques this endpoint is equal, and to capture all periods of low breathing frequency.

Tachypnea is defined as a prolonged period of the averaged (moving average with a window size of 10 minutes) RR >60 breaths per minute and >100 breaths per minute

(approximately two times the average normal RR).(18) To cover short and long periods of tachypnea, 3 different durations are studied (30 seconds, 60 seconds, and 10 minutes).

- 3) Calculating the percentage of time with reliable respiratory monitoring (without data loss and with an acceptable signal-to-noise ratio).

### **Data collection**

The following basic characteristics and demographic information will be collected at the baseline of the study: gestational age, birth weight, gender, age and weight at day of measurement, relevant medical status (respiratory support, medication and underlying illness during measurement), chest circumference, nipple distance, skin type at study start by visual inspection (normal, dry, flaky, oily, moist, other).

### **Sample size calculation**

A power calculation is performed for the primary outcome using data collected in a previous study.(submitted for publication, NICU AmsterdamUMC, 2021) Among the three criteria, criteria 1 needs the largest sample size and is used for our study. This resulted in 39 required infants to achieve 80% power with an overall 5% type I error with a Bonferroni correction (details in the Statistical Analysis Plan (SAP) in the online supplement).

In addition, an interim analysis will be performed as the power calculation was based on the previous study and recruitment of infants without being able to answer research questions is unethical.(19) This will be performed after including 1/3th of the infants for sample size adaption using the method of Mehta and Pocock.(20) If the conditional power falls within the pre-defined “promising zone”, the sample size will be increased to an upper limit of 52 infants. Otherwise, the study will proceed with the original sample size. To ensure that a representative sample of the age distribution of infants at a NICU, infants in different postmenstrual age groups

1  
2  
3 will be recruited with the same proportions as in the target population (minimally 10 infants  
4  
5 <30 weeks, between 30-32 weeks and >32 weeks).  
6  
7  
8

### 9 10 **Study procedures**

11 The Bambi® belt system is a non CE-certified medical device, designed for wireless  
12 cardiorespiratory monitoring of (pre)term infants in a hospital environment. All included  
13 infants will be monitored with the belt in addition to standard ECG/CI for 24 hours to obtain  
14 representative clinical scenarios throughout the entire day. The measurement set-up is  
15 visualised in Figure 1 and consists of 1) dEMG measurement with the belt and 2) the extraction  
16 of patient monitor data.  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26

27 In the belt, three dry electrodes are incorporated. When placing the belt at the height of the  
28 diaphragm, the outer two electrodes are in the nipple line and the middle electrode is in line  
29 with the sternum. The three ECG/CI electrodes are attached at the original location without  
30 hindering belt placement. The measured electrical signal of the diaphragm with the belt is  
31 wirelessly transmitted to the Receiver Module (REM) by the Sensor Module (SEM). The REM  
32 processes the dEMG signal to obtain the ECG and respiration signal (averaged diaphragmatic  
33 activity). An inbuilt algorithm provides the HR and RR out of the ECG and respiration signal  
34 respectively. This data is transported to a bedside computer. The data from the patient monitor  
35 (ECG, HR, RR, and SpO<sub>2</sub>) is extracted from the bedside monitor using an isolated cable and is  
36 also transported to the bedside computer.  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52

53 The belt data from the REM and patient monitor are recorded and synchronised using a  
54 dedicated software package (Polybench, Applied Biosignals, Weener, Germany) on a personal  
55 bedside computer. Data is recorded at a sample rate of 1 to 500 Hz for rate and waveform data  
56 respectively. The bedside software also provides the possibility to make measurements  
57  
58  
59  
60

1  
2  
3 annotations by nurses and researchers during data recording, such as re-positioning of the  
4  
5 infant, nursing and kangaroo care.  
6  
7  
8  
9

10 During the study, daily routine care proceeds as usual. The location of the belt is regularly  
11  
12 checked and if necessary repositioned (similar to the clinical practice). Notifications are  
13  
14 visualised when contact between skin and the belt is lost (Leads off) or when there is no  
15  
16 connection between the SEM and REM (Bluetooth Loss Error). In case of the first notification,  
17  
18 the belt may be repositioned, while in case of Bluetooth loss the battery level of the SEM or  
19  
20 blocking of this sensor (e.g. by an arm) are checked.  
21  
22  
23  
24  
25

26 Preferably, the belt stays in place during the study. However, the belt can be removed during  
27  
28 diagnostic imaging, patient handling, or in case of skin irritation at the belt location. The reason  
29  
30 for removal will be annotated. If the belt is removed, the medical staff, parents and one of the  
31  
32 dedicated researchers will decide together if the belt can be re-applied.  
33  
34  
35  
36

### 37 **Recruitment**

38  
39 Parents of all eligible infants are approached for consent to obtain a sample as heterogeneous  
40  
41 and representative as possible. Preferably, infants are included as soon as possible after birth.  
42  
43 During the 24 hours, the study can be terminated if requested by parents or the treating  
44  
45 physicians. In case of withdrawal of a subject, an extra subject will be included.  
46  
47  
48  
49

### 50 **Safety**

51  
52 Being a medical device study, this study was classified as a moderate risk.(21) A specified  
53  
54 monitor plan for the study is made based on risk-classification.  
55  
56  
57  
58

### 59 **STATISTICAL ANALYSIS**

1  
2  
3 A detailed SAP can be found in the online supplement. Unless otherwise specified, all  
4 hypothesis tests are two-sided with a significance level of 0.05. All statistical analyses will be  
5 performed using R version 4.0 (the R Foundation for Statistical Computing; Vienna, Austria)  
6 and SAS software version 9.4 (SAS Institute Inc., Cary, NC, USA).  
7  
8  
9  
10  
11  
12  
13

14 The non-inferiority/equivalence margins based on expert opinions (survey send to  
15 neonatologists of different NICU's in the Netherlands) and literature (4, 22, 23) are described  
16 in Table 1. In the different subparagraphs we refer to this table.  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 1** - The non-inferiority/equivalence margins for the primary and secondary outcomes

<i>Endpoints</i>	<i>Prespecified margins<sup>#</sup></i>
<i>LOA of second-to-second HR differences</i>	$\pm 8$ bpm
<i>LOA of RR trend differences</i>	$\pm 15$ brpm
<i>Sensitivity of brady-/tachycardia detection</i>	90% <sup>°</sup>
<i>PPV of brady-/tachycardia detection</i>	90% <sup>°</sup>
<i>Sensitivity of apnea/tachypnea detection</i>	70%
<i>PPV of apnea/tachypnea alarms</i>	0-100%*
<i>Data loss percentage</i>	5%
<i>Robust data percentage (HR)</i>	90%
<i>Robust data percentage (RR)</i>	70%

LOA: limits of agreement, HR: heart rate, RR: respiratory rate, PPV: positive predictive value.

Data loss is defined as the percentage of data with “Leads off” or “Bluetooth Loss Error” in the belt.

<sup>#</sup> The prespecified margins are compared to confidence intervals with corresponding confidence levels (see SAP for more details).

<sup>°</sup>Note: all missed bradycardias are checked for clinical relevance by two independent experts.

\*Since the reference devices for apnea detection in the clinical practice are the peripheral oxygen saturation (SpO<sub>2</sub>) and electrocardiogram instead of the respiration signal and the performance for Chest Impedance to detect tachypnea is unsatisfactory due to the presence of cardiac interference, all values for PPV for apnea/tachypnea are acceptable. Interpretations will be made

### Summary and descriptive statistics

Categorical data will be summarized by numbers of counts and percentages. Continuous data will be summarized by mean, standard deviation if data is normal and median, interquartile range (IQR) if data is skewed. Minimum and maximum values will also be presented for continuous data when appropriate.

### Statistical analysis of the primary outcome

#### Criteria 1: agreement in HR

To investigate the equivalence of HR measurement between the belt and ECG, we will fit a linear mixed model to the second-to-second HR difference between both. With this model, the 95% limits of agreement (Bland-Altman analysis) will be derived. The two-one-sided tests (TOST) with a multiplicity corrected alpha of 0.0167 and the prespecified margin ( $\pm 8$ bpm) will test equivalence between the two devices. In addition, based on a bivariate heteroscedastic model fitted to HR segments of a prespecified length, additional performance measures will be calculated as sensitivity analyses (details in SAP).

#### Criteria 2: cardiac event detection

For HR monitoring, we also consider the detection of bradycardia and tachycardia. We will estimate the sensitivity and the positive predictive value (PPV) of the belt using the patient monitor data as the ground truth and perform a non-inferiority test with an alpha of 0.0167. The non-inferiority margin for the sensitivity and PPV are listed in Table 1. In case of missed bradycardias, one independent expert per center will qualify the safety and clinical consequences of each missing event by answering the same questions per figure containing the discrepancy in HR and the ECG-signals measured with CI and the belt. These figures will be blinded and thus it will be unknown which signal corresponds CI or the belt.

#### Criteria 3: signal quality



1  
2  
3 The quality of the investigational device will be quantified based on the percentage of time  
4 during the 24-hour period it produces any reading (percentage without data loss due to “Leads  
5 off” or “Bluetooth Loss Error”) and the percentage in time it produces a good-quality-reading  
6 (percentage of robust data) for the HR and RR, respectively. For the HR non-robust data can be  
7 caused by bad connection (suboptimal Bluetooth or skin-electrode connection). These criteria  
8 are built-in in the belt algorithm and therefore this data is automatically labeled. Hypothesis  
9 testing will be used to establish the non-inferiority of this “uptime” percentage (percentage  
10 without data loss and percentage of robust data) of the belt.  
11

12  
13 For the RR, the uptime percentage is also categorized as a) data readings without data loss and  
14 b) robust data readings, i.e. readings without unrealistic (e.g. negative) values. Signal quality is  
15 only analyzed for the belt. However, these results are compared to prespecified margins,  
16 described in Table 1. As the HR monitored with CI is accurate and nearly continuous, while the  
17 RR is less relied upon and may be unreliable, the prespecified margin for the RR is lower than  
18 for the HR.  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37

### 38 **Statistical analysis of secondary outcomes**

39 Secondary analyses, based on the same statistical methods for the criteria of the primary  
40 outcome, include all secondary endpoints (apnea and tachypnea detection, RR trend analysis  
41 (see SAP)) and evaluation during different scenarios.  
42  
43  
44  
45  
46  
47  
48

### 49 **ETHICS AND DISSEMINATION**

50 The Medical Ethical Committee of the MMC (W21.042) and the Central Committee for Human  
51 Research in the Netherlands (CCMO, CCMO21/0167/PP) approved the study protocol (Version  
52 2, 19<sup>th</sup> of May 2021). Local feasibility at the AmsterdamUMC was approved by the Medical  
53 Ethical Committee of the AMC (2021\_146). This study was registered in the Dutch Trial  
54 Register (<https://www.trialregister.nl>, NL9480). Regarding patient safety, no belt related events  
55  
56  
57  
58  
59  
60

1  
2  
3 were observed in the pilot study and are therefore unexpected. Moreover, as every patient is  
4 monitored with ECG/CI and the belt, safety is guaranteed in case of missing belt data. The SAP  
5 will be used for the analyses. The results will be published in peer-reviewed journals and  
6 presented at future congresses.  
7  
8  
9  
10  
11

12 The MMC started patient recruitment in July and the AmsterdamUMC in August 2021. The  
13 duration of this study will be approximately seven months.  
14  
15  
16  
17  
18  
19  
20

## 21 **PATIENT AND PUBLIC INVOLVEMENT STATEMENT**

22  
23 Patients were included in this study after obtaining parental informed consent. The patients  
24 could not be involved in the design, recruitment, conduction and dissemination of results of this  
25 study. Neither could we ask the burden of the study. The outcome measures were developed by  
26 combining clinical and statistical knowledge to ensure a SAP that enables confirmation of non-  
27 inferiority of the belt compared to ECG/CI.  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38

## 39 **REFERENCES**

- 40  
41 1. Joshi R, van Straaten H, van de Mortel H, et al. Does the architectural layout of a NICU  
42 affect alarm pressure? A comparative clinical audit of a single-family room and an open bay  
43 area NICU using a retrospective study design. *BMJ Open* 2018;8(6):e022813. doi:  
44 10.1136/bmjopen-2018-022813  
45  
46  
47  
48  
49  
50  
51 2. Poets CF, Roberts RS, Schmidt B, Whyte RK, Asztalos E V, et al. Association Between  
52 Intermittent Hypoxemia or Bradycardia and Late Death or Disability in Extremely Preterm  
53 Infants. *JAMA*. (2015) 314(6):595–603. <https://doi.org/10.1001/jama.2015.88413>.  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 3. Di Fiore JM. Neonatal cardiorespiratory monitoring techniques. *Semin Neonatol*  
4  
5 2004;9(3):195-203. doi: 10.1016/j.siny.2003.11.009  
6  
7
- 8  
9 4. Kraaijenga JV, Hutten GJ, de Jongh FH, et al. Transcutaneous electromyography of the  
10 diaphragm: A cardio-respiratory monitor for preterm infants. *Pediatr Pulmonol*  
11 2015;50(9):889-95. doi: 10.1002/ppul.23116  
12  
13
- 14  
15 5. Csoma ZR, Meszes A, Abraham R, et al. Iatrogenic Skin Disorders and Related Factors  
16 in Newborn Infants. *Pediatr Dermatol* 2016;33(5):543-8. doi: 10.1111/pde.12960  
17  
18
- 19  
20 6. Boundy EO, Dastjerdi R, Spiegelman D, et al. Kangaroo Mother Care and Neonatal  
21 Outcomes: A Meta-analysis. *Pediatrics* 2016;137(1). doi: 10.1542/peds.2015-2238  
22  
23
- 24  
25 7. Conde-Agudelo A, Diaz-Rossello JL. Kangaroo mother care to reduce morbidity and  
26 mortality in low birthweight infants. *Cochrane Database Syst Rev* 2016(8). doi:  
27 <https://doi.org/10.1002/14651858.CD002771.pub4>  
28  
29
- 30  
31 8. Chung HU, Rwei AY, Hourlier-Fargette A, et al. Skin-interfaced biosensors for  
32 advanced wireless physiological monitoring in neonatal and pediatric intensive-care units. *Nat*  
33 *Med* 2020;26(3):418-29. doi: <https://doi.org/10.1038/s41591-020-0792-9>  
34  
35
- 36  
37 9. Henry C, Shipley L, Ward C, et al. Accurate neonatal heart rate monitoring using a new  
38 wireless, cap mounted device. *Acta Paediatr* 2021;110(1):72-8. doi:  
39 <https://doi.org/10.1111/apa.15303>  
40  
41
- 42  
43 10. Rimet Y, Brusquet Y, Ronayette D, et al. Surveillance of infants at risk of apparent life  
44 threatening events (ALTE) with the BBA bootee: a wearable multiparameter monitor. *Annu Int*  
45 *Conf IEEE Eng Med Biol Soc* 2007;2007:4997-5000. doi: 10.1109/IEMBS.2007.4353462  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 11. Chen H, Bao S, Lu C, et al. Design of an Integrated Wearable Multi-Sensor Platform  
4 Based on Flexible Materials for Neonatal Monitoring. *IEEE* 2020:23732-47. doi:  
5 10.1109/ACCESS.2020.2970469  
6  
7  
8  
9
- 10 12. Chen W, Dols S, Bambang Oetomo S, et al. Monitoring body temperature of newborn  
11 infants at neonatal intensive care units using wearable sensors. *BodyNets '10* 2010. doi:  
12 <https://doi.org/10.1145/2221924.2221960>  
13  
14  
15  
16
- 17 13. Linti C, Horter H, Osterreicher P, et al. Sensory baby vest for the monitoring of infants.  
18 *IEEE* 2006. doi: 10.1109/BSN.2006.49  
19  
20  
21  
22
- 23 14. Pike H, Eilevstjonn J, Bjorland P, et al. Heart rate detection properties of dry-electrode  
24 ECG compared to conventional 3-lead gel-electrode ECG in newborns. *BMC Res Notes*  
25 2021;14(1):166. doi: 10.1186/s13104-021-05576-x  
26  
27  
28  
29  
30
- 31 15. van Leuteren RW, Hutten GJ, de Waal CG, et al. Processing transcutaneous  
32 electromyography measurements of respiratory muscles, a review of analysis techniques.  
33 *Journal of Electromyography and Kinesiology* 2019:176-86. doi:  
34 10.1016/j.jelekin.2019.07.014  
35  
36  
37  
38  
39  
40
- 41 16. Lee H, Rusin CG, Lake DE, et al. A new algorithm for detecting central apnea in  
42 neonates. *Physiol Meas* 2012;33(1):1-17. doi: 10.1088/0967-3334/33/1/1  
43  
44  
45
- 46 17. Li T, Matsushima M, Timpson W, et al. Epidemiology of patient monitoring alarms in  
47 the neonatal intensive care unit. *J Perinatol* 2018;38(8):1030-8. doi: 10.1038/s41372-018-  
48 0095-x  
49  
50  
51  
52
- 53 18. Reuter S, Moser C, Baack M. Respiratory distress in the newborn. *Pediatr Rev*  
54 2014;35(10):417-28. doi: 10.1542/pir.35-10-417  
55  
56  
57  
58  
59  
60

- 1  
2  
3 19. European Medicines Agency. ICH Topic E 9 – Statistical Principles for Clinical Trials.  
4  
5 2006:21-3.  
6  
7  
8 20. Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are  
9 promising: a practical guide with examples. *Stat Med* 2011;30(28):3267-84. doi:  
10 https://doi.org/10.1002/sim.4102  
11  
12  
13  
14  
15 21. Central Committee on Research Involving Human Subjects.  
16  
17  
18 22. Anton O, Fernandez R, Rendon-Morales E, et al. Heart Rate Monitoring in Newborn  
19 Babies: A Systematic Review. *Neonatology* 2019;116(3):199-210. doi: 10.1159/000499675  
20  
21  
22  
23 23. Breteler MJM, KleinJan EJ, Dohmen DAJ, et al. Vital Signs Monitoring with Wearable  
24 Sensors in High-risk Surgical Patients: A Clinical Validation Study. *Anesthesiology*  
25 2020;132(3):424-39. doi: 10.1097/ALN.0000000000003029  
26  
27  
28  
29  
30  
31  
32  
33

#### 34 **AUTHOR CONTRIBUTIONS**

35 AS, HN, MV, RL, FJ, AK, JH conceptualized the study. ZZ and EH made the statistical analysis  
36 plan, which was reviewed by all authors. AS wrote the first version of this manuscript. All  
37 authors contributed to the final draft of the manuscript.  
38  
39  
40  
41  
42  
43  
44

#### 45 **FUNDING STATEMENT**

46 This work was supported by the Louise Vehmeijer foundation.  
47  
48  
49  
50  
51

#### 52 **COMPETING INTERESTS STATEMENT**

53 The authors declare that the research was conducted in the absence of any commercial  
54 relationships that could be construed as a potential conflict of interest. However, the MMC  
55 and AmsterdamUMC do receive a prespecified amount of money per performed  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

measurement. The results of this study protocol will contribute to obtaining a CE-mark for the Bambi® belt.

Confidential: For Review Only

**WHAT IS KNOWN ABOUT THE SUBJECT:**

- Disadvantages of the cardiorespiratory monitoring technique in neonates are indirect measurements of respiration, usage of adhesive electrodes and hindering wires.
- With transcutaneous electromyography of the diaphragm, respiratory activity is measured directly by recording the activity of the main respiratory muscle.
- The Bambi® belt is a novel wireless and non-adhesive belt that enables cardiorespiratory monitoring by measuring diaphragm activity with dry electrodes.

**WHAT THIS STUDY HOPES TO ADD:**

- Demonstration of the non-inferiority of the Bambi® belt compared to the electrocardiogram and chest impedance for cardiorespiratory monitoring in preterm and term infants.

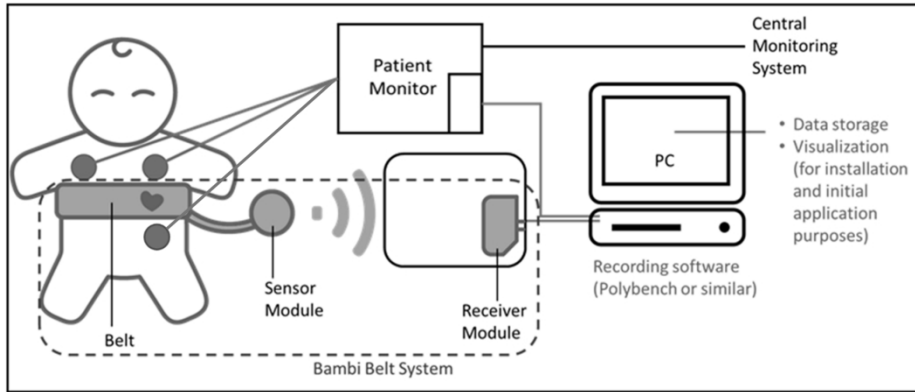
**HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY**

- When non-inferiority of the Bambi® belt compared to the current cardiorespiratory monitor is confirmed, the belt could be used as a wireless and skin-friendly alternative.

## FIGURES

**FIGURE 1** - The measurement set-up. The adhesive electrodes used for standard cardiorespiratory monitoring are attached at the original location, visualised by the three blue dots. The diaphragm activity measured with the Bambi® belt is wirelessly transmitted with the Sensor Module to the Receiver Module where the data is processed to obtain an electrocardiogram and respiration waveform (and heart rate and respiratory rate). This data and the data measured with the patient monitor are transported to a personal bedside computer with Polybench software to synchronise and record these signals.





170x80mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## ONLINE SUPPLEMENT – STATISTICAL ANALYSIS PLAN

### PURPOSE OF THE STATISTICAL ANALYSIS PLAN

The purpose of this statistical analysis plan (SAP) is to document technical and detailed specification for the analysis of data collected in the Bambi belt monitoring performance study.

The SAP has been written based on information contained in study protocol, dated 12th April 2021 before any data collection had taken place. It is prepared in compliance with the International Council on Harmonization (ICH) E9.

This SAP will be the guiding document for the analyses that will be conducted. Results of the analyses described in this SAP will be included in the Clinical Study Report (CSR). Any post hoc or unplanned analyses performed to provide results for inclusion in the CSR, but not identified in the prospective SAP will be identified in the given report. Additionally, the planned analyses of the primary aims will be included in future manuscripts. All the aims and research questions will be presented as an addendum as well.

### OVERVIEW AND DESCRIPTION OF THE STUDY

#### Study design

The study is a multi-center, paired design, clinical monitoring device measurement comparison study. The *investigational device* under consideration is the Bambi® belt monitoring system (using dry electrodes). The current standard device of cardiorespiratory monitoring through adhesive electrodes is considered as the clinical reference standard and thereafter referred to as the *reference device/method*. The Bambi® belt monitoring system will be used on infants by trained nurses in the neonatal intensive care units (NICU's) for continuous 24 hours monitoring in addition to the routine monitoring with the reference device on the same patients. Infants admitted to NICU's of the the Emma Children's Hospital of the Amsterdam University Medical

1  
2  
3 Centre (Amsterdam UMC) or Maxima Medical Center (MMC) will be measured at the earliest  
4  
5 suitable moment for clinical practice without interfering with infants' routine cycles.  
6  
7  
8  
9

### 10 **Randomization and blinding**

11  
12 No randomization is required for the paired design since both monitoring devices will be used  
13  
14 on the same patient at the same time. Blinding is also not possible since both the measurement  
15  
16 protocol and algorithmic characteristic differ substantially.  
17  
18  
19  
20

### 21 **Framework**

22  
23 The goal of this study is to establish the agreement between the investigational device and the  
24  
25 reference device. Unlike the traditional difference-based tests, non-inferiority and equivalence  
26  
27 techniques provide a better alternative for demonstrating the similarity between the two  
28  
29 measurement methods. Thus, we have adopted the non-inferiority/equivalence trial framework  
30  
31 for this primary objective of this study. This study considers three hypotheses ( $H_0$  denotes the  
32  
33 null hypothesis and  $H_A$  denotes the alternative hypothesis) for the first two primary outcomes:  
34  
35  
36  
37  
38  
39

40 1. Primary outcome, criterion 1: Heart rate measurement (second-by-second measurement)

41  
42  $H_0$ : The absolute difference between the investigational device and the reference device is  
43  
44 larger than the prespecified equivalence margin.

45  
46  $H_A$ : The absolute difference between the investigational device and the reference device is  
47  
48 within the prespecified equivalence margin.  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 2. Primary outcome, criterion 2: Brady-/tachy-cardia event detection  
4

5  $H_0$ : The composite cardiac event detection performances in terms of sensitivity and positive  
6 predictive value (PPV) based on the investigational device with respect to the reference device  
7  
8 is less than the prespecified non-inferiority margin.  
9

10  
11  
12  $H_A$ : The composite cardiac event detection performances in terms of sensitivity and PPV based  
13 on the investigational device with respect to the reference device is greater or equal to the  
14 prespecified non-inferiority margin.  
15  
16  
17  
18

19  
20  
21 3. Primary outcome, criterion 3: Reliable reading (percentage of the time)  
22

23  
24  $H_0$ : The percentage of the time the investigational device produces reliable readings is less than  
25 the prespecified non-inferiority margin.  
26  
27

28  
29  $H_A$ : The percentage of the time the investigational device produces reliable readings is greater  
30 or equal to the prespecified non-inferiority margin.  
31  
32  
33

34  
35 **Statistical interim analysis and stopping guidance**  
36

37 One interim analysis for sample size adaptation will be performed. That is, we will start with a  
38 certain sample size commitment which will be increased at the interim analysis in case the  
39 results obtained are reasonably promising. The interim analysis will be conducted after the  
40 prospectively recruited participant's number reaches one-third of the planned sample size.  
41  
42 Conditional power will be calculated for the analyses of the primary endpoints and compared  
43 to the boundary values of the conditional power for the promising zones (1, 2). In case the  
44 conditional power calculated at the interim analysis does fall inside the promising zone, the  
45 sample size will be increased to a predetermined limit. On the other hand, if the calculated  
46 conditional power is outside the promising zone, the study will proceed with the original sample  
47 size. Therefore, no early stopping rule is entailed in this study. Furthermore, a conventional  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 final analysis will be used without altering the level of type I error, since the promising zone is  
4  
5 defined as a set that ensures the type I error to be preserved conservatively for the final analysis.  
6  
7  
8  
9

## 10 **Study data**

11  
12 The following infant characteristics will be collected at baseline:

- 13 • Gestational age
- 14 • Postmenstrual age
- 15 • Gender
- 16 • Birth weight
- 17 • Weight at enrollment
- 18 • Ethnicity (derived from the electronic patient record or by asking the parents)
- 19 • Chest circumference
- 20 • Nipple distance
- 21 • Skin condition and abnormality

22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36 During the monitor study period, the following information will be measured:

- 37 • Clinical event
  - 38 • SpO<sub>2</sub>: Arterial oxygen saturation as measured by pulse oximetry
  - 39 • Medical status:
    - 40 ○ Ventilation support
    - 41 ○ Reports of medication and illness during the measurement
  - 42 • Lead status: Indicates whether at least one lead was off
  - 43 • Bluetooth link quality
  - 44 • Activities
    - 45 ○ Kangaroo care
    - 46 ○ Nurse care
- 47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- Feeding
- Medical Procedure
- Belt status
  - Moved: the belt is being moved
  - Open: the belt is removed from the patient
- Patient position
  - Unknown
  - Lying prone
  - Lying supine
  - Lying on the left side
  - Lying on the right side

## STATISTICAL ANALYSIS

Based on the collected information described above, the following total of variables will be derived:

- 10, 30, and 60 minutes moving average of the heart rate, and respiratory rate measured by both the investigational device and the reference device.
- Premature birth:
  - Premature (gestational age < 37 weeks)
  - Normal (gestational age  $\geq$  37 weeks)
- Desaturation: SpO<sub>2</sub> < 80% for at least 10 consecutive seconds
- Heart rate status (investigational and reference device):
  - Normal
  - Tachycardia (heart rate > 180 for at least 10 consecutive seconds)
  - Bradycardia (heart rate < 100 for at least 5 consecutive seconds)

- Respiration status (investigational and reference device):
  - Apnea (according to standard clinical definitions)
  - Tachypnea (respiratory rate >60 and >100 for 30 seconds, 1 minute, and 10 consecutive minutes in stationary signal)
- Measurement quality:
  - No anomalies
  - Poor data link: Bluetooth link is poor but data is still received
  - Unreliable data: One or more lead off, or no Bluetooth connection (Bluetooth Loss Error, BLE)

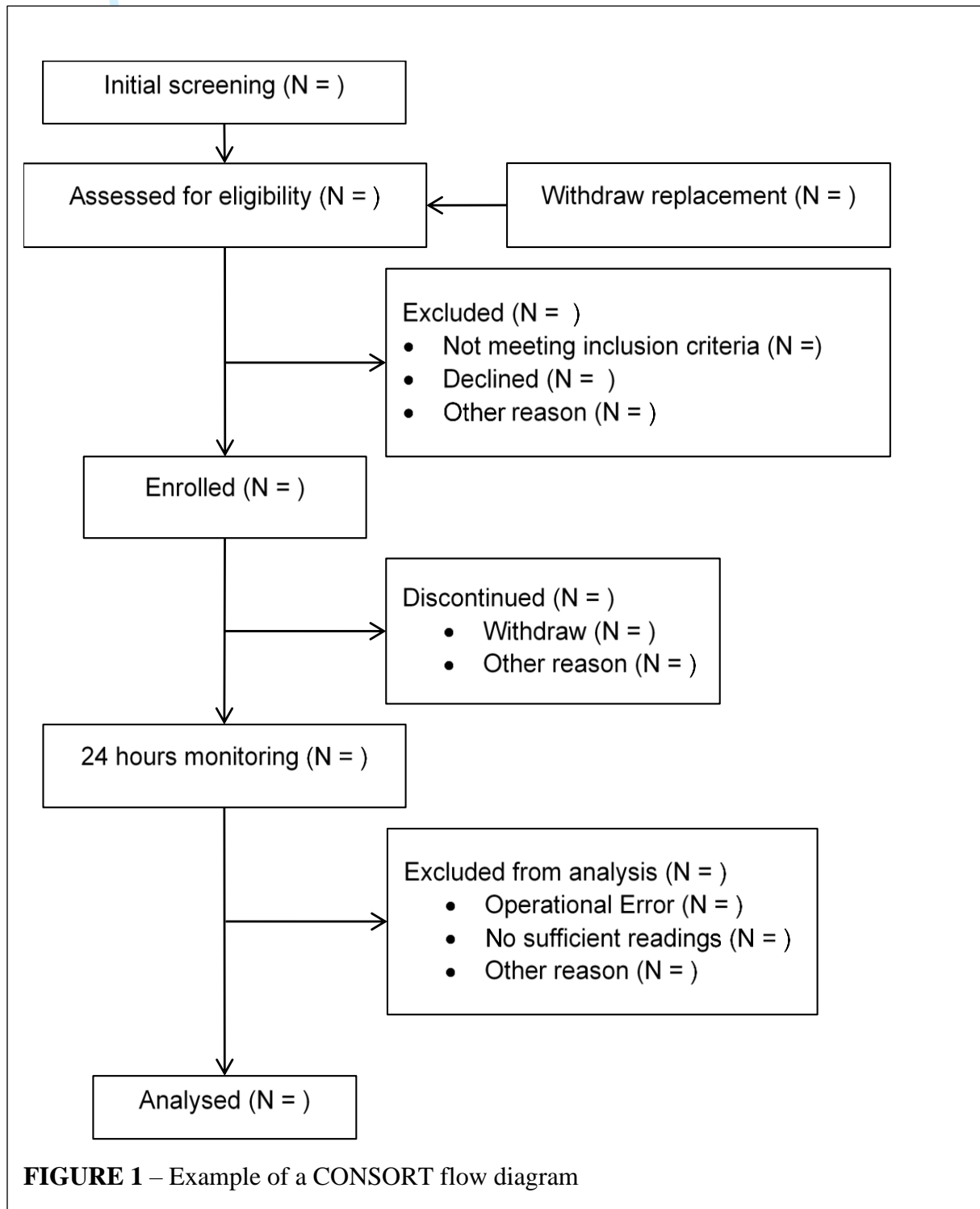
### Summary and descriptive statistics

Categorical data will be summarized by numbers and percentages. Continuous data will be summarized by mean, standard deviation if data are normal and median, interquartile range (IQR) if data are skewed. Minimum and maximum values will also be presented for continuous data. Tests of statistical significance will not be undertaken for baseline characteristics; rather the clinical importance of any imbalance will be noted.

A CONSORT flow diagram (example in Figure 1) will be used to summarize the number of infants who were:

- Assessed for eligibility at the screening
  - Eligible at screening
  - Ineligible at screening (with reasons)
- Eligible and enrolled
- Eligible but not enrolled
- Enrolled but did not receive any / sufficient measurements

- Discontinued
- Included in the analysis
- Excluded from the analysis





1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Confidential: For Review Only

## Analysis methods

Primary outcome, criterion 1: Heart rate measurement

To investigate and verify the equivalence of heart rate measurement between the investigational device and the reference device, we will fit a linear mixed model to the second-to-second heart rate difference between the two devices. Based on the estimates of the model, we will derive the 95% limits of agreement (3) as our main performance measure, known as the Bland-Altman analysis. The endpoints of the Bland-Altman 95% limits of agreement are the 2.5th percentile and 97.5th percentile for the distribution of the difference between paired measurements. We will calculate the  $(1 - \alpha/2)100\%$  confidence intervals of the percentiles according to Shieh (4), and conduct the two-one-sided t-tests (TOST) procedure with the prespecified equivalence margins (Table 1).

In addition, we will calculate the following performance measures to supplement the main analysis as sensitivity analyses to assess the agreement between the two devices from different aspects:

- The concordance correlation coefficient (5) and its variants
- Probability of Agreement (6) and Total Deviation Index (7)
- Coefficient of individual agreement (8)

These performance measures will be based on a bivariate heteroscedastic linear mixed-effects model fitted to each segment of the readings of a prespecified length from both devices. We will assume that measurements made with the two devices at the same time are correlated. Therefore, investigating the correlation between the two devices leads to the quantification of the degrees of agreement between them. Furthermore, we will consider the temporal correlations between measurements obtained with the same devices and the variabilities between different infants. Besides, we will start with a heteroscedastic model which does not assume equal variances for the two devices (namely, the measurement errors are not assumed

1  
2  
3 to be equal) and investigate the homogeneity of the measurement variabilities between the two  
4  
5 devices. Baseline characteristics of the infants and records of activities (listed in the study data  
6  
7 section) will be used as covariates in the model to partly explain the variabilities between the  
8  
9 infants. We will use the stepwise model selection procedure based on the Bayesian Information  
10  
11 Criteria (BIC) goodness-of-fit criteria.  
12  
13

#### 14 15 16 17 Primary outcome, criterion 2: Brady-/tachycardia event detection

18  
19 For brady-/tachycardia, the clinical event periods will be identified based on prespecified  
20  
21 margins. We will investigate the non-inferiority of sensitivity and positive predictive values  
22  
23 (PPV) of the event detected by the investigation device assuming that the reference device is  
24  
25 the predicate device and compare both values to the prespecified non-inferiority margins (Table  
26  
27 1). For the calculation of the sensitivity, when the event period identified based on the  
28  
29 investigational device overlaps with the event period identified by the reference device, it will  
30  
31 be counted as a true positive case. This is to prevent the repeated signaling of events from the  
32  
33 investigational device during a positive period identified by the reference device to inflate the  
34  
35 number of true positives. The same applies to the reference device when it comes to the  
36  
37 calculation of the PPV. That is, during an event period identified by the investigational device,  
38  
39 multiple event periods identified by the reference device will only be counted as one true-  
40  
41 positive case. Note that the true negative is ill-defined and will not be reported. Since true  
42  
43 negatives are used in the calculation of specificity, specificity will not be reported either.  
44  
45  
46  
47  
48  
49  
50

#### 51 Primary outcome, criterion 3: Safety and Quality

52  
53 *Safety:* The investigation of safety and tolerability is a multidimensional problem. Although we  
54  
55 don't anticipate any specific adverse effects for the investigational device, new and  
56  
57 unforeseeable effects are always possible. This background underlies the statistical difficulties  
58  
59  
60

1  
2  
3 associated with the analytical evaluation of the safety and tolerability of the device. We will  
4 address the safety and tolerability implications by applying descriptive statistical methods to  
5 the data, supplemented by calculation of confidence intervals whenever this aids interpretation  
6 and make use of graphical presentations in which patterns of adverse events are displayed.  
7  
8  
9  
10  
11

12  
13  
14 *Quality:* The quality of the investigational device will be quantified in terms of the point  
15 estimate and 95% confidence intervals based on the estimated percentages in time during the  
16 24-hour period it produces reliable readings for heart rate and respiratory rate, respectively.  
17 Reliable readings are defined in the study protocol. The uptime percentages are the percentage  
18 of data loss and the percentage of robust data readings. For each outcome, hypothesis testing  
19 will be used to establish the non-inferiority of the uptime percentages of the investigational  
20 device considering a non-inferiority margin specified in Table 1. The uptime percentages will  
21 be estimated based on a Generalized Estimation Equations (GEE) model.  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34

### 35 Missing data

36  
37 To get an idea about the complexity of the missing data problem in the data and information  
38 about the location of the missing values, the missing data pattern will be evaluated and reported.  
39  
40 We expect missing data in the primary outcomes measured by the investigational device to be  
41 the results of external causes such as the movement of the belt, signal losses, poor Bluetooth  
42 link qualities and so on. Therefore, it will be reasonable to assume that data are missing  
43 completely at random (MCAR). Formally, we will investigate the validity of such an  
44 assumption using Little's MCAR test. Furthermore, the availability of the data from the  
45 reference device (since it depends on a separate measurement system) provides us the  
46 opportunity to investigate whether the missingness is related to the underlying measurand. That  
47 is, whether the missing data mechanism is missing not at random (MNAR). This is rarely  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 possible in other types of studies. Nevertheless, considering the pair of bivariate measurements  
4  
5 from the investigational and the reference device, we will investigate the assumption using the  
6  
7 covariate-dependent missing (CDM) test proposed in Li (9). Note that CDM is usually  
8  
9 considered as missing at random (MAR), we here simply exploit the advantage of the data from  
10  
11 the reference device to test the dependencies between the missingness and the underlying  
12  
13 measurand. Furthermore, we will use the CDM test on other covariates (excluding the reference  
14  
15 device data) as well to test if the missingness is MAR.  
16  
17

18  
19 In the case of MAR (i.e., CDM without measurements from reference device), list wise deletion  
20  
21 can still be unbiased and will be used if the percentage of missingness is less than 5%.  
22  
23 Otherwise, multiple imputations (MI) will be considered. We will not use the measurements  
24  
25 from the reference device for the MI to avoid biasing the results towards the equivalence of the  
26  
27 two devices. On the other hand, if the missingness is related to the measurand after taking into  
28  
29 account all covariates, this indicates a potential problem of the measurement device, and a  
30  
31 separate analysis will be carried out to investigate the associations between the missingness and  
32  
33 the measurand.  
34  
35

36  
37 For multiple imputations, we will use the fully conditional specification method. Unrealistic  
38  
39 values (e.g., negative values for strictly positive variable) will be checked and corrected (e.g.,  
40  
41 using truncations). The imputation will be repeated at least 5 times and Rubin's rule will be  
42  
43 used to combine estimates and standard errors from the imputed data.  
44  
45  
46  
47  
48  
49

## 50 Secondary analyses

51  
52 If the sample size permits, we will perform subset analyses to explore the performances of the  
53  
54 investigational device under different scenarios.  
55  
56  
57  
58  
59  
60

### Subset analyses: primary endpoints

For each of the primary endpoints, we will consider additional exploratory analyses on the following subsets:

- During periods of a clinical event (e.g., apnea, bradycardia)
- During activities (e.g., Kangaroo care, feeding)
- During periods where the reference device's readings are stable
- Gestational age (e.g., preterm birth)
- Respiratory support (e.g., mechanical ventilation)

For these subsets, we will use the same model as the primary outcome to investigate the performances of the investigational device under various scenarios/activities of the infants. In case the subset does not contain enough data to fit the same model as the primary one, we will resort to a simpler model for case-by-case analyses.

### Respiratory rate analysis

It is known the reference device does not provide point-by-point accurate measurement resulting in large variabilities (measurement errors) in the measured respiratory rates. The intended clinical use of the readings in the NICU thus consists of two different aspects:

1. The trend of the respiratory rates over time;
2. Signaling of potentially respiratory related clinical events (i.e. apnea related desaturation and/or bradycardia, and potentially disease related tachypnea);

For the first usage, we will apply the same analysis method as the one used for heart rate on the moving average of the respiratory rate. We will primarily focus on the 10 minutes moving average for the respiratory rate. Analysis of the 1 minute and 5 minutes moving averages will

1  
2  
3 be used as a sensitivity analysis to establish the robustness of the conclusions made for the 10  
4  
5 minutes moving average.  
6

7  
8 For apnea and tachypnea, respectively, the clinical event periods will be identified based on  
9  
10 clinical definitions and the same methods as the brady-/tachycardia event detection will be used  
11  
12 to compare the sensitivity and PPV to the prespecified non-inferiority limits (Table 1).  
13  
14 However, it should be noted that since the reference device is known to have an unsatisfactory  
15  
16 performance of apnea/tachypnea detection, cautions are needed to interpret the sensitivity and  
17  
18 PPV as if the reference device is the truth.  
19  
20  
21  
22

#### 23 24 Statistical software

25  
26 All statistical analyses will be performed using R version 4.0 (the R Foundation for Statistical  
27  
28 Computing; Vienna, Austria) and SAS software version 9.4 (SAS Institute Inc., Cary, NC,  
29  
30 USA).  
31  
32  
33  
34

#### 35 36 Non-inferiority/equivalence criteria

37  
38 In Table 1 the non-inferiority/equivalence criteria for the primary and secondary outcomes are  
39  
40 visualized.  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 1** - The non-inferiority/equivalence margins for the primary and secondary outcomes

<i>Endpoints</i>	<i>Prespecified margins<sup>#</sup></i>
<i>LOA of second-to-second HR differences</i>	$\pm 8$ bpm
<i>LOA of RR trend differences</i>	$\pm 15$ brpm
<i>Sensitivity of brady-/tachycardia detection</i>	90% <sup>°</sup>
<i>PPV of brady-/tachycardia detection</i>	90% <sup>°</sup>
<i>Sensitivity of apnea/tachypnea detection</i>	70%
<i>PPV of apnea/tachypnea alarms</i>	0-100%*
<i>Data loss percentage</i>	5%
<i>Robust data percentage (HR)</i>	90%
<i>Robust data percentage (RR)</i>	70%

LOA: limits of agreement, HR: heart rate, RR: respiratory rate, PPV: positive predictive value.

Data loss is defined as the percentage of data with “Leads off” or “Bluetooth Loss Error” in the belt.

<sup>#</sup> The prespecified margins are compared to confidence intervals with corresponding confidence levels (see SAP for more details).

<sup>°</sup>Note: all missed bradycardias are checked for clinical relevance by two independent experts.

\*Since the reference devices for apnea detection in the clinical practice are the peripheral oxygen saturation (SpO<sub>2</sub>) and electrocardiogram instead of the respiration signal and the performance for Chest Impedance to detect tachypnea is unsatisfactory due to the presence of cardiac

## Sample size

Based on preliminary analysis of data collected in a feasibility study on a total of 13 infants with measurements from both the investigational device and the reference device, we were able to obtain preliminary information with regards to the characteristics of the primary endpoints



1  
2  
3 upon which we have formulated our sample size calculation.(submitted for publication, NICU  
4  
5 AmsterdamUMC, 2021)  
6  
7  
8  
9

10 A detailed specification of the sample size calculation can be found in the sections below. In  
11  
12 summary, for the monitor performance, 39 infants are needed to achieve 80% power with a 5%  
13  
14 overall type I error with a Bonferroni correction for multiplicity. It is worth noting that no  
15  
16 dropout was assumed during the sample size calculation. This is because we plan to include an  
17  
18 extra infant in case of withdrawal of an infant to fulfil the required sample size. Infants who  
19  
20 withdraw from the study will be followed up by one of the investigators and responsible medical  
21  
22 staff to obtain detailed reasons behind the withdraw. Dropout rate for the monitor performance  
23  
24 study is expected to be low, between 0-5%.  
25  
26  
27  
28  
29

30 While the preplanned sample size is 39 infants, we will include an adaptive sample size re-  
31  
32 estimation procedure as per the “promising zone” methodology of Mehta and Pocock (2) using  
33  
34 the data from the first 1/3 infants. This procedure involves the evaluation of conditional power  
35  
36 in the interim analysis, and if it were to fall in the prespecified “promising zone”, the sample  
37  
38 size will be increased, subject to a predetermined upper limit (52 infants) to increase the  
39  
40 conditional power to 80%. The boundary of the conditional power for the “promising zone” is  
41  
42 0.36 and 0.8.  
43  
44  
45  
46  
47  
48

49 Monitoring study: Primary endpoints: Heart rate

50  
51 For the sample size calculation, we will assume the measured heart rate difference  $D_{ij}$  between  
52  
53 the investigational device and the reference device at time point  $j$  ( $j = 1, \dots, m$ ) on infant  $i$  ( $i =$   
54  
55  $1, \dots, n$ ) can be modelled as:  
56  
57

$$D_{ij} = d + a_i + e_{ij}$$

where  $d$  is the overall difference,  $a_i$  is a random effect with  $a_i \sim N(0, \sigma_a^2)$ , and  $e_{ij} \sim N(0, \sigma_e^2)$  is the random error independent of  $a_i$ . Though, we considered a bivariate mixed-effects model for our analysis, the variance component model for the difference can be derived from the bivariate mixed-effects model, therefore we will use this variance component model for the sample size calculation. The variance of the difference will be estimated from the aforementioned model via  $\hat{\sigma}_d^2 = \hat{\sigma}_a^2 + \hat{\sigma}_e^2$ . Here  $\hat{\sigma}_a^2$  and  $\hat{\sigma}_e^2$  is the estimator of the between-subject variability  $\sigma_a^2$  and residual variability  $\sigma_e^2$ , respectively. The 95% limit of agreement (LOA) can be estimated as  $LOA = \hat{d} \pm 1.96 \hat{\sigma}_d$  with  $\hat{d}$  and  $\hat{\sigma}_d$  denotes the estimator of  $d$  and  $\sigma_d$ , respectively. The variance of the LOA estimator is  $\text{var}(\hat{d} \pm 1.96 \hat{\sigma}_d) = \text{var}(\hat{d}) + 1.96^2 \text{var}(\hat{\sigma}_d)$  ( $\hat{d}$  and  $\hat{\sigma}_d$  is asymptotically independent). Since for  $\hat{\sigma}_d^2 = \hat{\sigma}_a^2 + \hat{\sigma}_e^2$ , we have  $\text{var}(\hat{\sigma}_d^2) = \text{var}(\hat{\sigma}_a^2) + \text{var}(\hat{\sigma}_e^2) + \text{cov}(\hat{\sigma}_a^2, \hat{\sigma}_e^2)$ . Furthermore, each term on the right-hand side (assuming  $m$  is large) is given by:

$$\text{var}(\hat{\sigma}_a^2) = \frac{2}{m^2} \left[ \frac{(m\sigma_a^2 + \sigma_e^2)^2}{n-1} + \frac{\sigma_e^4}{n(m-1)} \right] \approx \frac{2\sigma_a^4}{n-1}, \text{var}(\hat{\sigma}_e^2) = \frac{2\sigma_e^4}{n(m-1)+2} \approx 0,$$

$$\text{cov}(\hat{\sigma}_a^2, \hat{\sigma}_e^2) = -\frac{2\sigma_e^4}{nm(m-1)} \approx 0;$$

This leads to  $\text{var}(\hat{\sigma}_d^2) \approx 2\sigma_a^4/(n-1)$ . Therefore, by the delta method, we have  $\text{var}(\hat{\sigma}_d) = \frac{1}{4\sigma_d^2} \text{var}(\hat{\sigma}_d^2) = \frac{\sigma_a^4}{2(n-1)\sigma_d^2}$ . According to Lu et al. (10), the power for the TOST is given by:

$$1 - \beta = 1 - T_{n-1} \left( t_{1-\frac{\alpha}{2}}, \frac{\delta - d - 1.96\sigma_d}{se_{LOA}} \right) - T_{n-1} \left( t_{1-\frac{\alpha}{2}}, \frac{\delta + d - 1.96\sigma_d}{se_{LOA}} \right)$$

where  $\alpha$ ,  $\beta$  denotes type I and type II error respectively,  $\delta$  is the predefined limit,  $se_{LOA} \approx$

$\sqrt{\frac{\sigma_d^2}{n} + \frac{1.96^2 \sigma_a^4}{2(n-1)\sigma_d^2}}$  is the standard error of the LOA estimate calculated according to the variance

component model, and  $T_{n-1}(\cdot, \tau)$  denotes the cumulative distribution function of a non-central Student's t-distribution with  $n-1$  degrees of freedom, and non-centrality parameter  $\tau$ .

For a 5% overall type I error rate, with a multiplicity correction factor of 3, and 80% power, the minimum sample size required is calculated at  $n = 39$ , for  $d = -0.5$ ,  $\sigma_a = 0.3$ , and  $\sigma_d = 3$ .

Primary endpoints: Brady-/tachycardia event detection

Suppose the total number of true events is  $M$  and are 100% detected by the reference device.

Assuming the true sensitivity is 95% for the investigational device, then a non-inferiority test using Z-test with normal approximation to the binomial distribution leads to a required  $M$  of 271 for a power of 80% and  $\alpha = 0.05/3 \approx 0.01667$  assuming the detection between each event (conditioning on the event itself) is independent. Considering the incidence of bradycardia to be 1 event per hour per infant according to the preliminary analysis of data from the feasibility study, at least 12 infants are needed to satisfy the required  $M$  (assuming each infant is measured for 24 hours long). The calculation is the same for PPV if we assume the investigational device is the truth. Assuming an incidence rate of tachycardia of 1.5 per hour per infant according to the preliminary analysis of data from the feasibility study, the required sample size is 8. Note that in the aforementioned calculation, we assume that the event-detection performance of the investigational device is homogeneous (or independent) among infants. A

1  
2  
3 sensitivity/robustness investigation regarding the sample size for infant-specific heterogeneous  
4 performances was performed, with results from which we can see that with  $n = 39$ , we have  
5 more than 90% power to detect a heterogeneous performance scenario where 15% of the  
6 population would have sensitivities between 80% - 90% and less than 5% of the population  
7 have sensitivities less than 80%.  
8  
9  
10  
11  
12  
13  
14  
15  
16

#### 17 Primary endpoints: Safety and quality

18  
19 Based on preliminary analysis of data collected in the feasibility study, we will assume that the  
20 overall probability of producing an erroneous reading at any time  $p_e$  is 2% and is constant  
21 across all participants. We will consider a non-inferiority test using normal approximation and  
22 a Z-test with the null hypothesis of  $H_0: p_e > 0.05$  and the alternative hypothesis of  $H_A: p_e \leq$   
23 0.05. The required number of observations for a given type I error of 1.667% ( $\approx 5\%/3$ ) to  
24 achieve 80% power is 376. Here the sample size 376 refers to 376 independent observations.  
25  
26 Considering the large numbers of repeated measurements (more than 376) within each  
27 participant, we will have sufficient power for this non-inferiority test even with 1 participant.  
28  
29 However, the assumption of independence can be too strong in the setting of our study.  
30  
31 Therefore, if we would assume an AR(1)-type dependency with correlation parameter  $\rho = 0.8$   
32 between two measurements within a participant, the variance inflation factor (VIF) for the  
33 asymptotic variance of the GEE estimator  $\hat{p}_e$  according to Pan is approximately (with the  
34 number of repeats  $m = 376$ ):  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

$$50 \quad 1 + \frac{2\rho}{1 - \rho} = 9$$

51  
52  
53 in the case of identity working correlation matrix when the true correlation has an AR(1)  
54 structure. To achieve the same power as the independent case calculated before, we need  
55  
56  
57  
58  
59  
60

$$\text{var}(\hat{p}_e) := \text{VIF} \frac{p_e(1-p_e)}{nm} = \frac{p_e(1-p_e)}{m}$$

Thus, we can conclude that at least  $n = \text{VIF} = 9$  participants will be needed to provide enough power for the non-inferiority test based on the GEE estimator using the identity working correlation matrix using the inverse proportionality between the required sample size and the variance of the estimator used in the Z-test. The same calculation can be carried out for the robust data percentages. It can be seen that only the number of repeats  $m$  will differ when the probabilities and the non-inferiority margins change while the VIF remains the same for the same value of the correlation parameter  $\rho$ . Among all settings, the largest  $m$  needed will be 718 when we assume the probability of producing robust data for respiratory rate is 75% with the corresponding non-inferiority margin equals to 70%. This number of repeats is still fully covered by the high-frequency measurements found in the study.

### Protocol deviations and analysis sets

#### Definition of protocol deviations

Protocol deviations (PD) occurring during the study will be determined for all enrolled infants, mainly from the clinical database by either clinical and/or medical review processes.

The mapping of the protocol deviations from the clinical database to analysis will be performed as per Table 2:

Table 2 – The influence of protocol deviations on the statistical analysis plan (SAP)

<i>Database label</i>	<i>SAP</i>
<i>Minor</i>	Not required
<i>Major</i>	Important
<i>Critical</i>	Important
<i>Clinical (a subset of Critical)</i>	Important

Important protocol deviations are protocol deviations that might significantly affect the completeness, accuracy, and/or reliability of the study data or that might significantly affect a subject's rights, safety, or well-being.

Important protocol deviations may also be recorded as "Major" protocol deviations in the database, but will be presented only as important in the analysis output.

Important protocol deviations include:

- Infants that are included in the study despite not satisfying the eligibility criteria;
- Infants that develop exclusion criteria while on the study but not withdrawn;
- Infants being measured with operational human errors;
- Deviation from Good Clinical Practice (ICE E6)

Clinically Important protocol deviations are the protocol deviations marked as important in Table 2, which lead to the exclusion of a subject from the analysis set.

The following deviations will be identified and confirmed before the partial database lock for the final analysis.

- Important protocol deviations including
  - Deviations from the inclusion and exclusion criteria
  - Deviations post inclusion

1  
2  
3 Protocol deviations may be identified by the data managers, clinical and medical staff either by  
4  
5 programmed validation checks or data listings/reports or manual verification of data sources.  
6  
7 Some important/major protocol deviation criteria may be identified in the clinical database via  
8  
9 biostatistical programs. Every important protocol deviation will be documented in the database  
10  
11 whether identified through sites monitoring, medical review or programming.  
12  
13  
14  
15  
16

## 17 REFERENCES

- 18  
19  
20 1. European Medicines Agency. ICH Topic E 9 – Statistical Principles for Clinical Trials.  
21  
22 2006:21-3.  
23  
24
- 25 2. Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are  
26  
27 promising: a practical guide with examples. *Stat Med* 2011;30(28):3267-84. doi:  
28  
29 <https://doi.org/10.1002/sim.4102>  
30  
31
- 32 3. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat*  
33  
34 *Methods Med Res* 1999;8(2):135-60. doi: 10.1177/096228029900800204  
35  
36  
37
- 38 4. Shieh G. The appropriateness of Bland-Altman's approximate confidence intervals for  
39  
40 limits of agreement. *BMC Med Res Methodol* 2018;18(1):45. doi:  
41  
42 <https://doi.org/10.1186/s12874-018-0505-y>  
43  
44
- 45 5. Lin L. I.-K. A Concordance Correlation Coefficient to Evaluate Reproducibility.  
46  
47 *Biometrics* 1989:255-68.  
48  
49
- 50 6. Stevens NT, Steiner SH, MacKay RJ. Assessing agreement between two measurement  
51  
52 systems: An alternative to the limits of agreement approach. *Stat Methods Med Res*  
53  
54 2017;26(6):2487-504. doi: 10.1177/0962280215601133  
55  
56  
57  
58  
59  
60

- 1  
2  
3 7. Lin L. I.-K. Total deviation index for measuring individual agreement with applications  
4 in laboratory performance and bioequivalence. *Statistics in Medicine* 2000;255-70.  
5  
6  
7
- 8 8. Haber M, Barnhart HX. A general approach to evaluating agreement between two  
9 observers or methods of measurement from quantitative data with replicated measurements.  
10  
11 *Stat Methods Med Res* 2008;17(2):151-69.  
12  
13
- 14 9. Li C. Little's Test of Missing Completely at Random. *The Stata Journal* 2013. p. 795-  
15 809. doi: 10.1177/1536867X1301300407  
16  
17  
18  
19
- 20 10. Lu MJ, Zhong WH, Liu YX, et al. Sample Size for Assessing Agreement between Two  
21 Methods of Measurement by Bland-Altman Method. *Int J Biostat* 2016;12(2). doi: 10.1515/ijb-  
22 2015-0039  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



# BMJ Paediatrics Open

## Multi-center paired non-inferiority study of the cardiorespiratory monitoring performance of the wireless and non-adhesive Bambi® belt measuring diaphragm activity in neonates: Study Protocol

Journal:	<i>BMJ Paediatrics Open</i>
Manuscript ID	bmjpo-2022-001430.R2
Article Type:	Protocol
Date Submitted by the Author:	25-Apr-2022
Complete List of Authors:	Scholten, Anouk; Amsterdam UMC Locatie AMC, department of Neonatology; Amsterdam UMC Locatie AMC, Amsterdam Reproduction & Development research institute Zhuozhao, Zhan; Eindhoven University of Technology, department of Mathematics and Computer Science Niemarkt, Hendrik; Máxima Medisch Centrum, department of Neonatology Vervoorn, Marieke; Máxima Medisch Centrum, department of Neonatology van Leutenen, Ruud; Amsterdam UMC Locatie AMC, department of Neonatology; Amsterdam UMC Locatie AMC, Amsterdam Reproduction & Development research institute de Jongh, Frans; Amsterdam UMC Locatie AMC, department of Neonatology; University of Twente, Faculty of Science and Technology van Kaam, Anton; Amsterdam UMC Locatie AMC, department of Neonatology; Amsterdam UMC Locatie AMC, Amsterdam Reproduction & Development research institute van den Heuvel, Edwin; Eindhoven University of Technology, department of Mathematics and Computer Science Hutten, G. Jeroen; Amsterdam UMC Locatie AMC, department of Neonatology; Amsterdam UMC Locatie AMC, Amsterdam Reproduction & Development research institute
Keywords:	Neonatology, Technology

SCHOLARONE™  
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1  
2  
3 **MULTI-CENTER PAIRED NON-INFERIORITY STUDY OF THE**  
4  
5 **CARDIORESPIRATORY MONITORING PERFORMANCE OF THE WIRELESS**  
6  
7 **AND NON-ADHESIVE BAMBI® BELT MEASURING DIAPHRAGM ACTIVITY IN**  
8  
9 **NEONATES: STUDY PROTOCOL**  
10

11  
12  
13  
14 Anouk W.J. Scholten<sup>1,2\*†</sup>, Zhan Zhuozhao<sup>3†</sup>, Hendrik J. Niemarkt<sup>4</sup>, Marieke Vervoorn<sup>4</sup>, Ruud  
15  
16 W. van Leuteren<sup>1,2</sup>, Frans H. de Jongh<sup>1,5</sup>, Anton H. van Kaam<sup>1,2</sup>, Edwin R. van den Heuvel<sup>3</sup>, G.  
17  
18 Jeroen Hutten<sup>1,2</sup>  
19

20  
21 *<sup>1</sup>Amsterdam UMC location University of Amsterdam, department of Neonatology,*  
22  
23 *Meibergdreef 9, Amsterdam, The Netherlands*  
24

25  
26 *<sup>2</sup>Amsterdam Reproduction & Development research institute, Amsterdam, The Netherlands*  
27

28  
29 *<sup>3</sup>Eindhoven University of Technology, department of Mathematics and Computer Science,*  
30  
31 *Eindhoven, The Netherlands*  
32

33  
34 *<sup>4</sup>Màxima Medical Center, department of neonatology, Veldhoven, The Netherlands*  
35

36  
37 *<sup>5</sup>University of Twente, Faculty of Science and Technology, Enschede, The Netherlands*  
38

39  
40 † These authors have contributed equally to this work and share first authorship  
41

42 **CORRESPONDING AUTHOR:**  
43

44 Anouk W.J. Scholten  
45

46 Department of Neonatology  
47

48 Emma Children's Hospital, AmsterdamUMC, location AMC  
49

50 Meibergdreef 9, 1105AZ Amsterdam, The Netherlands  
51

52 Phone: +31 20 5663675  
53

54 Email: a.w.j.scholten@amsterdamumc.nl  
55

56 **WORD COUNT: 2777**  
57  
58  
59  
60

**WHAT IS KNOWN ABOUT THE SUBJECT:**

- Disadvantages of the cardiorespiratory monitoring technique in neonates are indirect measurements of respiration, usage of adhesive electrodes and hindering wires.
- With transcutaneous electromyography of the diaphragm, respiratory activity is measured directly by recording the activity of the main respiratory muscle.
- The Bambi® belt is a novel wireless and non-adhesive belt that enables cardiorespiratory monitoring by measuring diaphragm activity with dry electrodes.

**WHAT THIS STUDY HOPES TO ADD:**

- Demonstration of the non-inferiority of the Bambi® belt compared to the electrocardiogram and chest impedance for cardiorespiratory monitoring in preterm and term infants.

## ABSTRACT

**Introduction:** Cardiorespiratory monitoring is used in the Neonatal Intensive Care Unit (NICU) to assess the clinical status of newborn infants and detect critical deteriorations in cardiorespiratory function. Currently, heart rate is monitored by electrocardiography (ECG) and respiration by chest impedance (CI). Disadvantages of current monitoring techniques are usage of wired adhesive electrodes which may damage the skin and hinder care. The Bambi® belt is a wireless and non-adhesive alternative that enables cardiorespiratory monitoring by measuring electrical activity of the diaphragm via transcutaneous electromyography (dEMG). A previous study showed feasibility of the Bambi belt and this study compares the belt performance to ECG and CI.

**Methods and analysis:** This multi-center non-inferiority paired study will be performed in the NICU of the Máxima Medical Center (MMC) in Veldhoven and the Emma Children's Hospital AmsterdamUMC in Amsterdam, the Netherlands. 39 infants in different postmenstrual age groups (minimally 10 infants <30 weeks, between 30-32 weeks and >32 weeks) will be recruited. These infants will be monitored with the Bambi® belt in addition to standard ECG and CI for 24 h. The primary outcome is the heart rate (HR), studied with three criteria: 1) the limits of agreement of the HR measurements in terms of the second-to-second difference in the HR between the belt and standard ECG, 2) the detection of cardiac events consisting of bradycardia and tachycardia and 3) the quality of HR-monitoring. The secondary outcome is the respiratory rate (RR), studied with the criteria 1) agreement in RR trend monitoring, 2) apnea and tachypnea detection, and 3) reliable registrations.

**Ethics and dissemination:** This protocol was approved by the Medical Ethical Committee of the Máxima Medical Center and the Central Committee for Human Research (CCMO). The

1  
2  
3 MMC started patient recruitment in July and the AmsterdamUMC in August 2021. The results  
4  
5 will be presented at conferences and published in peer-reviewed journals.  
6  
7

8 **Trial registration number:** NL9480 ([www.trialregister.nl](http://www.trialregister.nl))  
9  
10

## 11 INTRODUCTION

12  
13 In the Neonatal Intensive Care Unit (NICU), cardiorespiratory monitoring is crucial to assess  
14  
15 clinical condition and to timely detect and treat frequently occurring cardio-respiratory events  
16  
17 to prevent morbidity and mortality.(1, 2) To date, this is performed by measuring the  
18  
19 electrocardiogram (ECG) and chest impedance (CI) with three wired adhesive electrodes. CI  
20  
21 measures variation in electrical impedance across the chest during respiration caused by  
22  
23 changes in lung aeration and chest wall movement. These techniques provide continuous  
24  
25 monitoring of heart rate (HR), respiratory rate (RR), and breathing pattern. However, as CI  
26  
27 measures respiration indirectly, adequate detection of breathing cycles and apnea may not  
28  
29 always be optimal.(3)  
30  
31  
32  
33  
34  
35  
36  
37

38 With transcutaneous electromyography of the diaphragm (dEMG) breathing effort can be  
39  
40 recorded directly by measuring the electrical activity of this main respiratory muscle. To date,  
41  
42 this technique also uses three adhesive electrodes and provides information on respiration and  
43  
44 HR. Studies have shown its feasibility in the NICU-setting.(4)  
45  
46  
47  
48

49 The use of adhesive electrodes is restricted in infants with a postmenstrual age <26 weeks in  
50  
51 fear of skin damage.(5) Moreover, electrode removal may cause discomfort. Furthermore, the  
52  
53 wires attached to the electrodes restrict movements of the infant and may hinder parent-infant  
54  
55 interaction, nursing and kangaroo care. Restrictions in kangaroo care may impact patient  
56  
57 outcome as it has been associated with beneficial effects such as decreased mortality, decreased  
58  
59  
60

1  
2  
3 risk of severe infection/sepsis and hypothermia, and increased likelihood of exclusive breast  
4 feeding.(6, 7) All things considered, it is important to find alternatives for using wired adhesive  
5 electrodes.  
6  
7  
8  
9

10  
11  
12 In the past years, several wireless wearable sensors have been developed to measure various  
13 parameters in neonates such as ECG, HR, RR, peripheral oxygen saturation and (skin)  
14 temperature.(8-14) Recently, a novel wireless and non-adhesive sensor belt (Bambi® belt,  
15 Bambi B.V., Eindhoven, the Netherlands) was developed for neonatal use that measures ECG  
16 and respiration based on the dEMG technique. A recent pilot study showed that measuring HR  
17 and RR with this belt in preterm infants is feasible and that the measured HR and RR trend was  
18 similar to ECG and CI.(15) However, before replacing the current techniques using adhesive  
19 wired electrodes with the non-adhesive sensor belt, a larger study is required to demonstrate the  
20 non-inferiority of this belt as an alternative cardiorespiratory monitor. In this study, we compare  
21 the monitoring performance of the Bambi® belt to ECG and CI and hypothesize that the  
22 performance of the belt is non-inferior to the current monitoring techniques.  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

## 40 **METHODS**

### 41 **Study design**

42 This multi-center paired non-inferiority study will be performed in the NICU of Máxima  
43 Medical Center (MMC) in Veldhoven and the Emma Children's Hospital of the Amsterdam  
44 University Medical Centre (AmsterdamUMC), both located in the Netherlands. Each patient  
45 will be simultaneously measured with the belt and ECG/CI (paired design). To compare the  
46 devices, a non-inferiority/equivalence framework will be used. Here, equivalence is defined as  
47 the limit of agreement of the HR/RR between the belt and ECG/CI being within prespecified  
48 margins (see Table 1 for the margins). Non-inferiority is defined as the performance of clinical  
49 event detection and quality criteria not being worse than prespecified margins.  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

### Study population

Preterm and term infants being routinely monitored with the standard cardiorespiratory monitor (Intellivue MP90, Philips Healthcare, Eindhoven, The Netherlands) are included in the study. To ensure a representative sample of the target population, infants in different age groups will be included. Infants with chest skin lesions, congenital anomalies, and other scenarios preventing belt placement, such as (effects of) surgery or wrap for therapeutic hypothermia, will be excluded.

### Primary outcome

As HR-monitoring is clinically most relied upon and both ECG and dEMG provide the HR by measuring cardiac electrical activity, while CI and dEMG measure respiration with a different technique, the HR is considered the primary outcome.<sup>(3, 16)</sup> This will be studied with three criteria, which will be compared to the prespecified margins in Table 1. 1) Reliable monitoring performance through second-to-second HR measurement agreement in terms of differences in measured HR between the belt and the ECG/CI monitoring. 2) The detection of a composite cardiac event consisting of bradycardia (HR < 100 beats per minute for at least five seconds)<sup>(17)</sup> and tachycardia (HR > 180 beats per minute for at least ten seconds)<sup>(18)</sup> between the belt and the ECG measured with adhesive electrodes. The minimal duration of a bradycardia or tachycardia will prevent the inclusion of technical errors (short drops or increases in the HR) in our analysis and is lower for bradycardia compared to tachycardia as bradycardias are shorter events.<sup>(1)</sup> The thresholds are empirically chosen to detect all low and high HR-values. 3) Non-inferior quality (percentage of time with HR recordings without data loss).



Moreover, we will perform subgroup analyses to investigate whether the HR measurement performance is consistent under different clinical activities (e.g. kangaroo care, feeding) and in the different age groups.

### Secondary outcomes

The secondary outcome is the measured RR. This will be studied using the following three criteria, which will be compared to the prespecified margins in Table 1:

- 1) Comparing the trend in RR values provided by the belt and CI, based on the difference in the 10-minute moving averages. The RR-trend is studied as this is used in the clinical practice to detect for example increases in RR over time as a marker of clinical deterioration of a patient.(3) Since CI is widely used for neonatal respiratory monitoring, it is used as the reference technique.
- 2) Next to comparing the RR-trend, the ability to detect apnea and tachypnea is studied as the detection of these critical respiratory events based on RR is another purpose of the respiratory monitoring. Clinically relevant apneas are considered when indicated by a RR < 20 breaths per minute measured with CI for at least 10 seconds, associated with a desaturation (arterial oxygen saturation as measured by pulse oximetry (SpO<sub>2</sub>) <80% for at least 10 seconds) and/or bradycardia (HR <100 beats per minute for at least five seconds) (objective apnea measurement).(17) A RR<20 breaths per minute is chosen for the apnea definition as we solely use the numerical RR-values, because despite the two different measurement techniques this endpoint is equal, and to capture all periods of low breathing frequency.

Tachypnea is defined as a prolonged period of the averaged (moving average with a window size of 10 minutes) RR >60 breaths per minute and >100 breaths per minute (approximately two times the average normal RR).(19) To cover short and long periods of tachypnea, 3 different durations are studied (30 seconds, 60 seconds, and 10 minutes).

- 1  
2  
3 3) Calculating the percentage of time with reliable respiratory monitoring (without data  
4  
5 loss and with an acceptable signal-to-noise ratio).  
6  
7  
8  
9

### 10 **Data collection**

11 The following basic characteristics and demographic information will be collected at the  
12 baseline of the study: gestational age, birth weight, gender, age and weight at day of  
13 measurement, relevant medical status (respiratory support, medication and underlying illness  
14 during measurement), chest circumference, nipple distance, skin type at study start by visual  
15 inspection (normal, dry, flaky, oily, moist, other).  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25

### 26 **Sample size calculation**

27 A power calculation is performed for the primary outcome using data collected in a previous  
28 study.(15) Among the three criteria, criteria 1 needs the largest sample size and is used for our  
29 study. This resulted in 39 required infants to achieve 80% power with an overall 5% type I error  
30 with a Bonferroni correction (details in the Statistical Analysis Plan (SAP) in the online  
31 supplement).  
32  
33  
34  
35  
36  
37  
38

39 In addition, an interim analysis will be performed as the power calculation was based on the  
40 previous study and recruitment of infants without being able to answer research questions is  
41 unethical.(20) This will be performed after including 1/3th of the infants for sample size  
42 adaption using the method of Mehta and Pocock.(21) If the conditional power falls within the  
43 pre-defined “promising zone”, the sample size will be increased to an upper limit of 52 infants.  
44 Otherwise, the study will proceed with the original sample size. To ensure that a representative  
45 sample of the age distribution of infants at a NICU, infants in different postmenstrual age groups  
46 will be recruited with the same proportions as in the target population (minimally 10 infants  
47 <30 weeks, between 30-32 weeks and >32 weeks).  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Study procedures

The Bambi® belt system is a non CE-certified medical device, designed for wireless cardiorespiratory monitoring of (pre)term infants in a hospital environment. All included infants will be monitored with the belt in addition to standard ECG/CI for 24 hours to obtain representative clinical scenarios throughout the entire day. The measurement set-up is visualised in Figure 1 and consists of 1) dEMG measurement with the belt and 2) the extraction of patient monitor data.

In the belt, three dry electrodes are incorporated (Figure 2). When placing the belt at the height of the diaphragm, the outer two electrodes are in the nipple line and the middle electrode is in line with the sternum. The three ECG/CI electrodes are attached at the original location without hindering belt placement. The measured electrical signal of the diaphragm with the belt is wirelessly transmitted to the Receiver Module (REM) by the Sensor Module (SEM). The REM processes the dEMG signal to obtain the ECG and respiration signal (averaged diaphragmatic activity). An inbuilt algorithm provides the HR and RR out of the ECG and respiration signal respectively. This data is transported to a bedside computer. The data from the patient monitor (ECG, HR, RR, and SpO<sub>2</sub>) is extracted from the bedside monitor using an isolated cable and is also transported to the bedside computer.

The belt data from the REM and patient monitor are recorded and synchronised using a dedicated software package (Polybench, Applied Biosignals, Weener, Germany) on a personal bedside computer. Data is recorded at a sample rate of 1 to 500 Hz for rate and waveform data respectively. The bedside software also provides the possibility to make measurements annotations by nurses and researchers during data recording, such as re-positioning of the infant, nursing and kangaroo care.

1  
2  
3 During the study, daily routine care proceeds as usual. The location of the belt is regularly  
4 checked and if necessary repositioned (similar to the clinical practice). Notifications are  
5 visualised when contact between skin and the belt is lost (Leads off) or when there is no  
6 connection between the SEM and REM (Bluetooth Loss Error). In case of the first notification,  
7 the belt may be repositioned, while in case of Bluetooth loss the battery level of the SEM or  
8 blocking of this sensor (e.g. by an arm) are checked.  
9

10  
11  
12  
13  
14  
15  
16  
17  
18  
19 Preferably, the belt stays in place during the study. However, the belt can be removed during  
20 diagnostic imaging, patient handling, or in case of skin irritation at the belt location. The reason  
21 for removal will be annotated. If the belt is removed, the medical staff, parents and one of the  
22 dedicated researchers will decide together if the belt can be re-applied.  
23  
24  
25  
26  
27  
28  
29

### 30 **Recruitment**

31  
32 Parents of all eligible infants are approached for consent to obtain a sample as heterogeneous  
33 and representative as possible. Preferably, infants are included as soon as possible after birth.  
34 During the 24 hours, the study can be terminated if requested by parents or the treating  
35 physicians. In case of withdrawal of a subject, an extra subject will be included.  
36  
37  
38  
39  
40  
41  
42  
43

### 44 **Safety**

45 Being a medical device study, this study was classified as a moderate risk.(22) A specified  
46 monitor plan for the study is made based on risk-classification.  
47  
48  
49  
50  
51

### 52 **STATISTICAL ANALYSIS**

53  
54 A detailed SAP can be found in the online supplement. Unless otherwise specified, all  
55 hypothesis tests are two-sided with a significance level of 0.05. All statistical analyses will be  
56  
57  
58  
59  
60

1  
2  
3 performed using R version 4.0 (the R Foundation for Statistical Computing; Vienna, Austria)  
4  
5 and SAS software version 9.4 (SAS Institute Inc., Cary, NC, USA).  
6  
7  
8  
9

10 The non-inferiority/equivalence margins based on expert opinions (survey send to  
11 neonatologists of different NICU's in the Netherlands) and literature (4, 23, 24) are described  
12  
13 in Table 1. In the different subparagraphs we refer to this table.  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 1** - The non-inferiority/equivalence margins for the primary and secondary outcomes

<i>Endpoints</i>	<i>Prespecified margins<sup>#</sup></i>
<i>LOA of second-to-second HR differences</i>	$\pm 8$ bpm
<i>LOA of RR trend differences</i>	$\pm 15$ brpm
<i>Sensitivity of brady-/tachycardia detection</i>	90% <sup>°</sup>
<i>PPV of brady-/tachycardia detection</i>	90% <sup>°</sup>
<i>Sensitivity of apnea/tachypnea detection</i>	70%
<i>PPV of apnea/tachypnea alarms</i>	0-100%*
<i>Data loss percentage</i>	5%
<i>Robust data percentage (HR)</i>	90%
<i>Robust data percentage (RR)</i>	70%

LOA: limits of agreement, HR: heart rate, RR: respiratory rate, PPV: positive predictive value.

Data loss is defined as the percentage of data with “Leads off” or “Bluetooth Loss Error” in the belt.

<sup>#</sup> The prespecified margins are compared to confidence intervals with corresponding confidence levels (see SAP for more details).

<sup>°</sup>Note: all missed bradycardias are checked for clinical relevance by two independent experts.

\*Since the reference devices for apnea detection in the clinical practice are the peripheral oxygen saturation (SpO<sub>2</sub>) and electrocardiogram instead of the respiration signal and the performance for Chest Impedance to detect tachypnea is unsatisfactory due to the presence of cardiac interference, all values for PPV for apnea/tachypnea are acceptable. Interpretations will be made

## Summary and descriptive statistics

Categorical data will be summarized by numbers of counts and percentages. Continuous data will be summarized by mean, standard deviation if data is normal and median, interquartile range (IQR) if data is skewed. Minimum and maximum values will also be presented for continuous data when appropriate.

## Statistical analysis of the primary outcome

### Criteria 1: agreement in HR

To investigate the equivalence of HR measurement between the belt and ECG, we will fit a linear mixed model to the second-to-second HR difference between both. With this model, the 95% limits of agreement (Bland-Altman analysis) will be derived. The two-one-sided tests (TOST) with a multiplicity corrected alpha of 0.0167 and the prespecified margin ( $\pm 8$ bpm) will test equivalence between the two devices. In addition, based on a bivariate heteroscedastic model fitted to HR segments of a prespecified length, additional performance measures will be calculated as sensitivity analyses (details in SAP).

### Criteria 2: cardiac event detection

For HR monitoring, we also consider the detection of bradycardia and tachycardia. We will estimate the sensitivity and the positive predictive value (PPV) of the belt using the patient monitor data as the ground truth and perform a non-inferiority test with an alpha of 0.0167. The non-inferiority margin for the sensitivity and PPV are listed in Table 1. In case of missed bradycardias, one independent expert per center will qualify the safety and clinical consequences of each missing event by answering the same questions per figure containing the discrepancy in HR and the ECG-signals measured with CI and the belt. These figures will be blinded and thus it will be unknown which signal corresponds CI or the belt.

### Criteria 3: signal quality

1  
2  
3 The quality of the investigational device will be quantified based on the percentage of time  
4 during the 24-hour period it produces any reading (percentage without data loss due to “Leads  
5 off” or “Bluetooth Loss Error”) and the percentage in time it produces a good-quality-reading  
6 (percentage of robust data) for the HR and RR, respectively. For the HR non-robust data can be  
7 caused by bad connection (suboptimal Bluetooth or skin-electrode connection). These criteria  
8 are built-in in the belt algorithm and therefore this data is automatically labeled. Hypothesis  
9 testing will be used to establish the non-inferiority of this “uptime” percentage (percentage  
10 without data loss and percentage of robust data) of the belt.  
11

12  
13 For the RR, the uptime percentage is also categorized as a) data readings without data loss and  
14 b) robust data readings, i.e. readings without unrealistic (e.g. negative) values. Signal quality is  
15 only analyzed for the belt. However, these results are compared to prespecified margins,  
16 described in Table 1. As the HR monitored with CI is accurate and nearly continuous, while the  
17 RR is less relied upon and may be unreliable, the prespecified margin for the RR is lower than  
18 for the HR.  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37

### 38 **Statistical analysis of secondary outcomes**

39 Secondary analyses, based on the same statistical methods for the criteria of the primary  
40 outcome, include all secondary endpoints (apnea and tachypnea detection, RR trend analysis  
41 (see SAP)) and evaluation during different scenarios.  
42  
43  
44  
45  
46  
47  
48

### 49 **ETHICS AND DISSEMINATION**

50 The Medical Ethical Committee of the MMC (W21.042) and the Central Committee for Human  
51 Research in the Netherlands (CCMO, CCMO21/0167/PP) approved the study protocol (Version  
52 2, 19<sup>th</sup> of May 2021). Local feasibility at the AmsterdamUMC was approved by the Medical  
53 Ethical Committee of the AMC (2021\_146). This study was registered in the Dutch Trial  
54 Register (<https://www.trialregister.nl>, NL9480). Regarding patient safety, no belt related events  
55  
56  
57  
58  
59  
60



1  
2  
3 were observed in the pilot study and are therefore unexpected. Moreover, as every patient is  
4 monitored with ECG/CI and the belt, safety is guaranteed in case of missing belt data. The SAP  
5 will be used for the analyses. The results will be published in peer-reviewed journals and  
6 presented at future congresses.  
7  
8  
9  
10  
11

12 The MMC started patient recruitment in July and the AmsterdamUMC in August 2021. The  
13 duration of this study will be approximately seven months.  
14  
15  
16  
17  
18  
19  
20

## 21 **PATIENT AND PUBLIC INVOLVEMENT STATEMENT**

22 Patients were included in this study after obtaining parental informed consent. The patients  
23 could not be involved in the design, recruitment, conduction and dissemination of results of this  
24 study. Neither could we ask the burden of the study. The outcome measures were developed by  
25 combining clinical and statistical knowledge to ensure a SAP that enables confirmation of non-  
26 inferiority of the belt compared to ECG/CI.  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38

## 39 **REFERENCES**

- 40  
41 1. Joshi R, van Straaten H, van de Mortel H, et al. Does the architectural layout of a NICU  
42 affect alarm pressure? A comparative clinical audit of a single-family room and an open bay  
43 area NICU using a retrospective study design. *BMJ Open* 2018;8(6):e022813. doi:  
44 10.1136/bmjopen-2018-022813  
45  
46  
47  
48  
49
- 50  
51 2. Poets CF, Roberts RS, Schmidt B, Whyte RK, Asztalos E V, et al. Association Between  
52 Intermittent Hypoxemia or Bradycardia and Late Death or Disability in Extremely Preterm  
53 Infants. *JAMA*. (2015) 314(6):595–603. <https://doi.org/10.1001/jama.2015.88413>.  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 3. Di Fiore JM. Neonatal cardiorespiratory monitoring techniques. *Semin Neonatol*  
4  
5 2004;9(3):195-203. doi: 10.1016/j.siny.2003.11.009  
6  
7
- 8  
9 4. Kraaijenga JV, Hutten GJ, de Jongh FH, et al. Transcutaneous electromyography of the  
10 diaphragm: A cardio-respiratory monitor for preterm infants. *Pediatr Pulmonol*  
11  
12 2015;50(9):889-95. doi: 10.1002/ppul.23116  
13  
14
- 15  
16 5. Csoma ZR, Meszes A, Abraham R, et al. Iatrogenic Skin Disorders and Related Factors  
17 in Newborn Infants. *Pediatr Dermatol* 2016;33(5):543-8. doi: 10.1111/pde.12960  
18  
19
- 20  
21 6. Boundy EO, Dastjerdi R, Spiegelman D, et al. Kangaroo Mother Care and Neonatal  
22 Outcomes: A Meta-analysis. *Pediatrics* 2016;137(1). doi: 10.1542/peds.2015-2238  
23  
24
- 25  
26 7. Conde-Agudelo A, Diaz-Rossello JL. Kangaroo mother care to reduce morbidity and  
27 mortality in low birthweight infants. *Cochrane Database Syst Rev* 2016(8). doi:  
28  
29 <https://doi.org/10.1002/14651858.CD002771.pub4>  
30  
31
- 32  
33 8. Chung HU, Rwei AY, Hourlier-Fargette A, et al. Skin-interfaced biosensors for  
34 advanced wireless physiological monitoring in neonatal and pediatric intensive-care units. *Nat*  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
*Med* 2020;26(3):418-29. doi: <https://doi.org/10.1038/s41591-020-0792-9>
9. Henry C, Shipley L, Ward C, et al. Accurate neonatal heart rate monitoring using a new  
wireless, cap mounted device. *Acta Paediatr* 2021;110(1):72-8. doi:  
<https://doi.org/10.1111/apa.15303>
10. Rimet Y, Brusquet Y, Ronayette D, et al. Surveillance of infants at risk of apparent life  
threatening events (ALTE) with the BBA bootee: a wearable multiparameter monitor. *Annu Int  
Conf IEEE Eng Med Biol Soc* 2007;2007:4997-5000. doi: 10.1109/IEMBS.2007.4353462

- 1  
2  
3 11. Chen H, Bao S, Lu C, et al. Design of an Integrated Wearable Multi-Sensor Platform  
4 Based on Flexible Materials for Neonatal Monitoring. *IEEE* 2020:23732-47. doi:  
5  
6 10.1109/ACCESS.2020.2970469  
7  
8
- 9  
10 12. Chen W, Dols S, Bambang Oetomo S, et al. Monitoring body temperature of newborn  
11 infants at neonatal intensive care units using wearable sensors. *BodyNets '10* 2010. doi:  
12  
13 <https://doi.org/10.1145/2221924.2221960>  
14  
15
- 16 13. Linti C, Horter H, Osterreicher P, et al. Sensory baby vest for the monitoring of infants.  
17  
18 *IEEE* 2006. doi: 10.1109/BSN.2006.49  
19  
20
- 21 14. Pike H, Eilevstjonn J, Bjorland P, et al. Heart rate detection properties of dry-electrode  
22  
23 ECG compared to conventional 3-lead gel-electrode ECG in newborns. *BMC Res Notes*  
24  
25 2021;14(1):166. doi: 10.1186/s13104-021-05576-x  
26  
27
- 28 15. Scholten AWJ, van Leuteren RW, de Waal CG, de Jongh FH, van Kaam AH, Hutten  
29  
30 GJ. Feasibility of wireless cardiorespiratory monitoring with dry electrodes incorporated in a  
31  
32 belt in preterm infants. *Physiological Measurement* 2022. doi: [10.1088/1361-6579/ac69a9](https://doi.org/10.1088/1361-6579/ac69a9)  
33  
34
- 35 16. van Leuteren RW, Hutten GJ, de Waal CG, et al. Processing transcutaneous  
36  
37 electromyography measurements of respiratory muscles, a review of analysis techniques.  
38  
39 *Journal of Electromyography and Kinesiology* 2019:176-86. doi:  
40  
41 10.1016/j.jelekin.2019.07.014  
42  
43
- 44 17. Lee H, Rusin CG, Lake DE, et al. A new algorithm for detecting central apnea in  
45  
46 neonates. *Physiol Meas* 2012;33(1):1-17. doi: 10.1088/0967-3334/33/1/1  
47  
48
- 49 18. Li T, Matsushima M, Timpson W, et al. Epidemiology of patient monitoring alarms in  
50  
51 the neonatal intensive care unit. *J Perinatol* 2018;38(8):1030-8. doi: 10.1038/s41372-018-  
52  
53 0095-x  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 19. Reuter S, Moser C, Baack M. Respiratory distress in the newborn. *Pediatr Rev*  
4 2014;35(10):417-28. doi: 10.1542/pir.35-10-417  
5  
6  
7  
8 20. European Medicines Agency. ICH Topic E 9 – Statistical Principles for Clinical Trials.  
9 2006:21-3.  
10  
11  
12  
13 21. Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are  
14 promising: a practical guide with examples. *Stat Med* 2011;30(28):3267-84. doi:  
15 https://doi.org/10.1002/sim.4102  
16  
17  
18  
19  
20  
21 22. Central Committee on Research Involving Human Subjects.  
22  
23  
24 23. Anton O, Fernandez R, Rendon-Morales E, et al. Heart Rate Monitoring in Newborn  
25 Babies: A Systematic Review. *Neonatology* 2019;116(3):199-210. doi: 10.1159/000499675  
26  
27  
28  
29 24. Bretelet MJM, KleinJan EJ, Dohmen DAJ, et al. Vital Signs Monitoring with Wearable  
30 Sensors in High-risk Surgical Patients: A Clinical Validation Study. *Anesthesiology*  
31 2020;132(3):424-39. doi: 10.1097/ALN.0000000000003029  
32  
33  
34  
35  
36  
37  
38

### 39 AUTHOR CONTRIBUTIONS

40 AS, HN, MV, RL, FJ, AK, JH conceptualized the study. ZZ and EH made the statistical analysis  
41 plan, which was reviewed by all authors. AS wrote the first version of this manuscript. All  
42 authors contributed to the final draft of the manuscript.  
43  
44  
45  
46  
47  
48  
49

### 50 FUNDING STATEMENT

51 This work was supported by the Louise Vehmeijer foundation.  
52  
53  
54  
55  
56

### 57 COMPETING INTERESTS STATEMENT

58  
59  
60

1  
2  
3 Bambi Medical B.V. supports the study by a financial grant and use of equipment free of  
4  
5 charge. Data collection, analysis, interpretation, and reporting will be done independent of  
6  
7 Bambi Medical B.V..  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Confidential: For Review Only

1  
2  
3  
4  
5 **HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY**  
6  
7

- 8 • When non-inferiority of the Bambi® belt compared to the current cardiorespiratory  
9 monitor is confirmed, the belt could be used as a wireless and skin-friendly alternative.  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

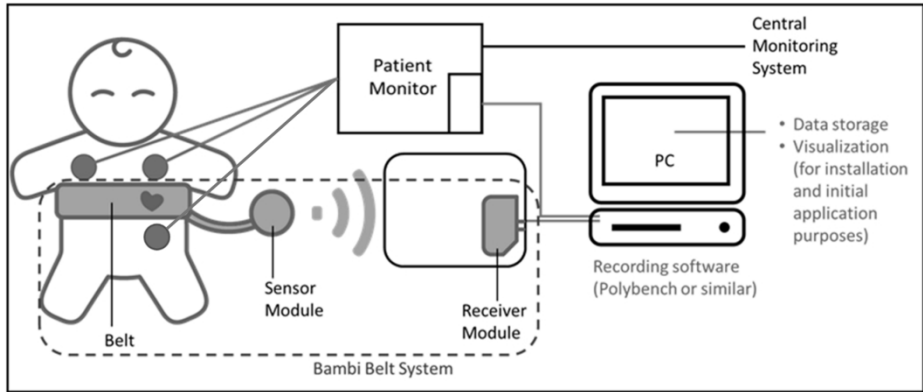
Confidential: For Review Only

## FIGURES

**FIGURE 1** - The measurement set-up. The adhesive electrodes used for standard cardiorespiratory monitoring are attached at the original location, visualised by the three grey dots. The diaphragm activity measured with the Bambi® belt is wirelessly transmitted with the Sensor Module to the Receiver Module where the data is processed to obtain an electrocardiogram and respiration waveform (and heart rate and respiratory rate). This data and the data measured with the patient monitor are transported to a personal bedside computer with Polybench software to synchronise and record these signals.

**FIGURE 2** – The Bambi® belt is a wireless non-adhesive belt designed for cardiorespiratory monitoring of (pre)term infants. The three dry electrodes (2) measure electrical activity of the diaphragm via transcutaneous electromyography. This data is wirelessly transmitted with the sensor module (1) to a receiver module that processes the diaphragm activity to obtain the ECG, respiration signal, heart rate and respiratory rate.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



170x80mm (300 x 300 DPI)





170x39mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## ONLINE SUPPLEMENT – STATISTICAL ANALYSIS PLAN

### PURPOSE OF THE STATISTICAL ANALYSIS PLAN

The purpose of this statistical analysis plan (SAP) is to document technical and detailed specification for the analysis of data collected in the Bambi belt monitoring performance study.

The SAP has been written based on information contained in study protocol, dated 12th April 2021 before any data collection had taken place. It is prepared in compliance with the International Council on Harmonization (ICH) E9.

This SAP will be the guiding document for the analyses that will be conducted. Results of the analyses described in this SAP will be included in the Clinical Study Report (CSR). Any post hoc or unplanned analyses performed to provide results for inclusion in the CSR, but not identified in the prospective SAP will be identified in the given report. Additionally, the planned analyses of the primary aims will be included in future manuscripts. All the aims and research questions will be presented as an addendum as well.

### OVERVIEW AND DESCRIPTION OF THE STUDY

#### Study design

The study is a multi-center, paired design, clinical monitoring device measurement comparison study. The *investigational device* under consideration is the Bambi® belt monitoring system (using dry electrodes). The current standard device of cardiorespiratory monitoring through adhesive electrodes is considered as the clinical reference standard and thereafter referred to as the *reference device/method*. The Bambi® belt monitoring system will be used on infants by trained nurses in the neonatal intensive care units (NICU's) for continuous 24 hours monitoring in addition to the routine monitoring with the reference device on the same patients. Infants admitted to NICU's of the the Emma Children's Hospital of the Amsterdam University Medical

Centre (Amsterdam UMC) or Maxima Medical Center (MMC) will be measured at the earliest suitable moment for clinical practice without interfering with infants' routine cycles.

### Randomization and blinding

No randomization is required for the paired design since both monitoring devices will be used on the same patient at the same time. Blinding is also not possible since both the measurement protocol and algorithmic characteristic differ substantially.

### Framework

The goal of this study is to establish the agreement between the investigational device and the reference device. Unlike the traditional difference-based tests, non-inferiority and equivalence techniques provide a better alternative for demonstrating the similarity between the two measurement methods. Thus, we have adopted the non-inferiority/equivalence trial framework for this primary objective of this study. This study considers three hypotheses ( $H_0$  denotes the null hypothesis and  $H_A$  denotes the alternative hypothesis) for the first two primary outcomes:

1. Primary outcome, criterion 1: Heart rate measurement (second-by-second measurement)

$H_0$ : The absolute difference between the investigational device and the reference device is larger than the prespecified equivalence margin.

$H_A$ : The absolute difference between the investigational device and the reference device is within the prespecified equivalence margin.

1  
2  
3 2. Primary outcome, criterion 2: Brady-/tachy-cardia event detection  
4

5  $H_0$ : The composite cardiac event detection performances in terms of sensitivity and positive  
6 predictive value (PPV) based on the investigational device with respect to the reference device  
7  
8 is less than the prespecified non-inferiority margin.  
9

10  
11  
12  $H_A$ : The composite cardiac event detection performances in terms of sensitivity and PPV based  
13 on the investigational device with respect to the reference device is greater or equal to the  
14 prespecified non-inferiority margin.  
15  
16  
17  
18

19  
20  
21 3. Primary outcome, criterion 3: Reliable reading (percentage of the time)  
22

23  
24  $H_0$ : The percentage of the time the investigational device produces reliable readings is less than  
25 the prespecified non-inferiority margin.  
26

27  
28  $H_A$ : The percentage of the time the investigational device produces reliable readings is greater  
29 or equal to the prespecified non-inferiority margin.  
30  
31  
32

33  
34  
35 **Statistical interim analysis and stopping guidance**  
36

37 One interim analysis for sample size adaptation will be performed. That is, we will start with a  
38 certain sample size commitment which will be increased at the interim analysis in case the  
39 results obtained are reasonably promising. The interim analysis will be conducted after the  
40 prospectively recruited participant's number reaches one-third of the planned sample size.  
41  
42 Conditional power will be calculated for the analyses of the primary endpoints and compared  
43 to the boundary values of the conditional power for the promising zones (1, 2). In case the  
44 conditional power calculated at the interim analysis does fall inside the promising zone, the  
45 sample size will be increased to a predetermined limit. On the other hand, if the calculated  
46 conditional power is outside the promising zone, the study will proceed with the original sample  
47 size. Therefore, no early stopping rule is entailed in this study. Furthermore, a conventional  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 final analysis will be used without altering the level of type I error, since the promising zone is  
4  
5 defined as a set that ensures the type I error to be preserved conservatively for the final analysis.  
6  
7  
8  
9

## 10 **Study data**

11  
12 The following infant characteristics will be collected at baseline:

- 13
- 14
- 15 • Gestational age
- 16
- 17 • Postmenstrual age
- 18
- 19 • Gender
- 20
- 21
- 22 • Birth weight
- 23
- 24 • Weight at enrollment
- 25
- 26 • Ethnicity (derived from the electronic patient record or by asking the parents)
- 27
- 28 • Chest circumference
- 29
- 30 • Nipple distance
- 31
- 32 • Skin condition and abnormality
- 33
- 34
- 35

36 During the monitor study period, the following information will be measured:

- 37
- 38 • Clinical event
- 39
- 40 • SpO<sub>2</sub>: Arterial oxygen saturation as measured by pulse oximetry
- 41
- 42 • Medical status:
  - 43 ○ Ventilation support
  - 44
  - 45 ○ Reports of medication and illness during the measurement
  - 46
  - 47
  - 48
  - 49
- 50 • Lead status: Indicates whether at least one lead was off
- 51
- 52 • Bluetooth link quality
- 53
- 54 • Activities
  - 55 ○ Kangaroo care
  - 56
  - 57 ○ Nurse care
  - 58
  - 59
  - 60

- Feeding
- Medical Procedure
- Belt status
  - Moved: the belt is being moved
  - Open: the belt is removed from the patient
- Patient position
  - Unknown
  - Lying prone
  - Lying supine
  - Lying on the left side
  - Lying on the right side

## STATISTICAL ANALYSIS

Based on the collected information described above, the following total of variables will be derived:

- 10, 30, and 60 minutes moving average of the heart rate, and respiratory rate measured by both the investigational device and the reference device.
- Premature birth:
  - Premature (gestational age < 37 weeks)
  - Normal (gestational age  $\geq$  37 weeks)
- Desaturation: SpO<sub>2</sub> < 80% for at least 10 consecutive seconds
- Heart rate status (investigational and reference device):
  - Normal
  - Tachycardia (heart rate > 180 for at least 10 consecutive seconds)
  - Bradycardia (heart rate < 100 for at least 5 consecutive seconds)

- Respiration status (investigational and reference device):
  - Apnea (according to standard clinical definitions)
  - Tachypnea (respiratory rate >60 and >100 for 30 seconds, 1 minute, and 10 consecutive minutes in stationary signal)
- Measurement quality:
  - No anomalies
  - Poor data link: Bluetooth link is poor but data is still received
  - Unreliable data: One or more lead off, or no Bluetooth connection (Bluetooth Loss Error, BLE)

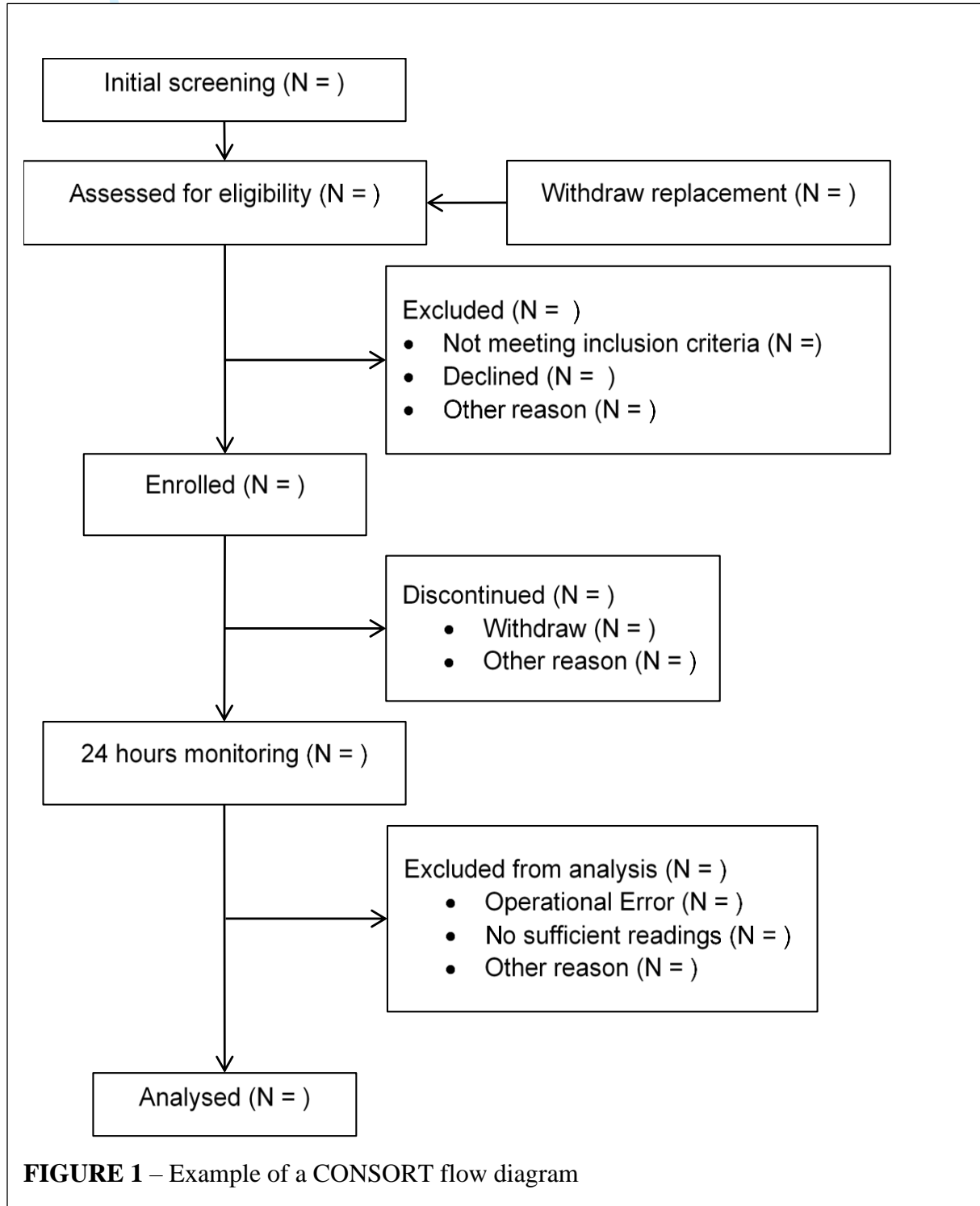
### Summary and descriptive statistics

Categorical data will be summarized by numbers and percentages. Continuous data will be summarized by mean, standard deviation if data are normal and median, interquartile range (IQR) if data are skewed. Minimum and maximum values will also be presented for continuous data. Tests of statistical significance will not be undertaken for baseline characteristics; rather the clinical importance of any imbalance will be noted.

A CONSORT flow diagram (example in Figure 1) will be used to summarize the number of infants who were:

- Assessed for eligibility at the screening
  - Eligible at screening
  - Ineligible at screening (with reasons)
- Eligible and enrolled
- Eligible but not enrolled
- Enrolled but did not receive any / sufficient measurements

- Discontinued
- Included in the analysis
- Excluded from the analysis





1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Confidential: For Review Only

## Analysis methods

### Primary outcome, criterion 1: Heart rate measurement

To investigate and verify the equivalence of heart rate measurement between the investigational device and the reference device, we will fit a linear mixed model to the second-to-second heart rate difference between the two devices. Based on the estimates of the model, we will derive the 95% limits of agreement (3) as our main performance measure, known as the Bland-Altman analysis. The endpoints of the Bland-Altman 95% limits of agreement are the 2.5th percentile and 97.5th percentile for the distribution of the difference between paired measurements. We will calculate the  $(1 - \alpha/2)100\%$  confidence intervals of the percentiles according to Shieh (4), and conduct the two-one-sided t-tests (TOST) procedure with the prespecified equivalence margins (Table 1).

In addition, we will calculate the following performance measures to supplement the main analysis as sensitivity analyses to assess the agreement between the two devices from different aspects:

- The concordance correlation coefficient (5) and its variants
- Probability of Agreement (6) and Total Deviation Index (7)
- Coefficient of individual agreement (8)

These performance measures will be based on a bivariate heteroscedastic linear mixed-effects model fitted to each segment of the readings of a prespecified length from both devices. We will assume that measurements made with the two devices at the same time are correlated. Therefore, investigating the correlation between the two devices leads to the quantification of the degrees of agreement between them. Furthermore, we will consider the temporal correlations between measurements obtained with the same devices and the variabilities between different infants. Besides, we will start with a heteroscedastic model which does not assume equal variances for the two devices (namely, the measurement errors are not assumed

1  
2  
3 to be equal) and investigate the homogeneity of the measurement variabilities between the two  
4  
5 devices. Baseline characteristics of the infants and records of activities (listed in the study data  
6  
7 section) will be used as covariates in the model to partly explain the variabilities between the  
8  
9 infants. We will use the stepwise model selection procedure based on the Bayesian Information  
10  
11 Criteria (BIC) goodness-of-fit criteria.  
12  
13

#### 14 15 16 17 Primary outcome, criterion 2: Brady-/tachycardia event detection

18  
19 For brady-/tachycardia, the clinical event periods will be identified based on prespecified  
20  
21 margins. We will investigate the non-inferiority of sensitivity and positive predictive values  
22  
23 (PPV) of the event detected by the investigation device assuming that the reference device is  
24  
25 the predicate device and compare both values to the prespecified non-inferiority margins (Table  
26  
27 1). For the calculation of the sensitivity, when the event period identified based on the  
28  
29 investigational device overlaps with the event period identified by the reference device, it will  
30  
31 be counted as a true positive case. This is to prevent the repeated signaling of events from the  
32  
33 investigational device during a positive period identified by the reference device to inflate the  
34  
35 number of true positives. The same applies to the reference device when it comes to the  
36  
37 calculation of the PPV. That is, during an event period identified by the investigational device,  
38  
39 multiple event periods identified by the reference device will only be counted as one true-  
40  
41 positive case. Note that the true negative is ill-defined and will not be reported. Since true  
42  
43 negatives are used in the calculation of specificity, specificity will not be reported either.  
44  
45  
46  
47  
48  
49

#### 50 51 Primary outcome, criterion 3: Safety and Quality

52  
53 *Safety:* The investigation of safety and tolerability is a multidimensional problem. Although we  
54  
55 don't anticipate any specific adverse effects for the investigational device, new and  
56  
57 unforeseeable effects are always possible. This background underlies the statistical difficulties  
58  
59  
60

1  
2  
3 associated with the analytical evaluation of the safety and tolerability of the device. We will  
4 address the safety and tolerability implications by applying descriptive statistical methods to  
5 the data, supplemented by calculation of confidence intervals whenever this aids interpretation  
6 and make use of graphical presentations in which patterns of adverse events are displayed.  
7  
8  
9  
10  
11  
12  
13

14 *Quality:* The quality of the investigational device will be quantified in terms of the point  
15 estimate and 95% confidence intervals based on the estimated percentages in time during the  
16 24-hour period it produces reliable readings for heart rate and respiratory rate, respectively.  
17 Reliable readings are defined in the study protocol. The uptime percentages are the percentage  
18 of data loss and the percentage of robust data readings. For each outcome, hypothesis testing  
19 will be used to establish the non-inferiority of the uptime percentages of the investigational  
20 device considering a non-inferiority margin specified in Table 1. The uptime percentages will  
21 be estimated based on a Generalized Estimation Equations (GEE) model.  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34

### 35 Missing data

36  
37 To get an idea about the complexity of the missing data problem in the data and information  
38 about the location of the missing values, the missing data pattern will be evaluated and reported.  
39  
40 We expect missing data in the primary outcomes measured by the investigational device to be  
41 the results of external causes such as the movement of the belt, signal losses, poor Bluetooth  
42 link qualities and so on. Therefore, it will be reasonable to assume that data are missing  
43 completely at random (MCAR). Formally, we will investigate the validity of such an  
44 assumption using Little's MCAR test. Furthermore, the availability of the data from the  
45 reference device (since it depends on a separate measurement system) provides us the  
46 opportunity to investigate whether the missingness is related to the underlying measurand. That  
47 is, whether the missing data mechanism is missing not at random (MNAR). This is rarely  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 possible in other types of studies. Nevertheless, considering the pair of bivariate measurements  
4  
5 from the investigational and the reference device, we will investigate the assumption using the  
6  
7 covariate-dependent missing (CDM) test proposed in Li (9). Note that CDM is usually  
8  
9 considered as missing at random (MAR), we here simply exploit the advantage of the data from  
10  
11 the reference device to test the dependencies between the missingness and the underlying  
12  
13 measurand. Furthermore, we will use the CDM test on other covariates (excluding the reference  
14  
15 device data) as well to test if the missingness is MAR.  
16  
17

18  
19 In the case of MAR (i.e., CDM without measurements from reference device), list wise deletion  
20  
21 can still be unbiased and will be used if the percentage of missingness is less than 5%.  
22  
23 Otherwise, multiple imputations (MI) will be considered. We will not use the measurements  
24  
25 from the reference device for the MI to avoid biasing the results towards the equivalence of the  
26  
27 two devices. On the other hand, if the missingness is related to the measurand after taking into  
28  
29 account all covariates, this indicates a potential problem of the measurement device, and a  
30  
31 separate analysis will be carried out to investigate the associations between the missingness and  
32  
33 the measurand.  
34  
35

36  
37 For multiple imputations, we will use the fully conditional specification method. Unrealistic  
38  
39 values (e.g., negative values for strictly positive variable) will be checked and corrected (e.g.,  
40  
41 using truncations). The imputation will be repeated at least 5 times and Rubin's rule will be  
42  
43 used to combine estimates and standard errors from the imputed data.  
44  
45  
46  
47  
48  
49

## 50 Secondary analyses

51  
52  
53 If the sample size permits, we will perform subset analyses to explore the performances of the  
54  
55 investigational device under different scenarios.  
56  
57  
58  
59  
60

### Subset analyses: primary endpoints

For each of the primary endpoints, we will consider additional exploratory analyses on the following subsets:

- During periods of a clinical event (e.g., apnea, bradycardia)
- During activities (e.g., Kangaroo care, feeding)
- During periods where the reference device's readings are stable
- Gestational age (e.g., preterm birth)
- Respiratory support (e.g., mechanical ventilation)

For these subsets, we will use the same model as the primary outcome to investigate the performances of the investigational device under various scenarios/activities of the infants. In case the subset does not contain enough data to fit the same model as the primary one, we will resort to a simpler model for case-by-case analyses.

### Respiratory rate analysis

It is known the reference device does not provide point-by-point accurate measurement resulting in large variabilities (measurement errors) in the measured respiratory rates. The intended clinical use of the readings in the NICU thus consists of two different aspects:

1. The trend of the respiratory rates over time;
2. Signaling of potentially respiratory related clinical events (i.e. apnea related desaturation and/or bradycardia, and potentially disease related tachypnea);

For the first usage, we will apply the same analysis method as the one used for heart rate on the moving average of the respiratory rate. We will primarily focus on the 10 minutes moving average for the respiratory rate. Analysis of the 1 minute and 5 minutes moving averages will

1  
2  
3 be used as a sensitivity analysis to establish the robustness of the conclusions made for the 10  
4  
5 minutes moving average.  
6

7  
8 For apnea and tachypnea, respectively, the clinical event periods will be identified based on  
9  
10 clinical definitions and the same methods as the brady-/tachycardia event detection will be used  
11  
12 to compare the sensitivity and PPV to the prespecified non-inferiority limits (Table 1).  
13  
14 However, it should be noted that since the reference device is known to have an unsatisfactory  
15  
16 performance of apnea/tachypnea detection, cautions are needed to interpret the sensitivity and  
17  
18 PPV as if the reference device is the truth.  
19  
20  
21  
22

#### 23 24 Statistical software

25  
26 All statistical analyses will be performed using R version 4.0 (the R Foundation for Statistical  
27  
28 Computing; Vienna, Austria) and SAS software version 9.4 (SAS Institute Inc., Cary, NC,  
29  
30 USA).  
31  
32  
33  
34

#### 35 36 Non-inferiority/equivalence criteria

37  
38 In Table 1 the non-inferiority/equivalence criteria for the primary and secondary outcomes are  
39  
40 visualized.  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 1** - The non-inferiority/equivalence margins for the primary and secondary outcomes

<i>Endpoints</i>	<i>Prespecified margins<sup>#</sup></i>
<i>LOA of second-to-second HR differences</i>	$\pm 8$ bpm
<i>LOA of RR trend differences</i>	$\pm 15$ brpm
<i>Sensitivity of brady-/tachycardia detection</i>	90% <sup>°</sup>
<i>PPV of brady-/tachycardia detection</i>	90% <sup>°</sup>
<i>Sensitivity of apnea/tachypnea detection</i>	70%
<i>PPV of apnea/tachypnea alarms</i>	0-100%*
<i>Data loss percentage</i>	5%
<i>Robust data percentage (HR)</i>	90%
<i>Robust data percentage (RR)</i>	70%

LOA: limits of agreement, HR: heart rate, RR: respiratory rate, PPV: positive predictive value.

Data loss is defined as the percentage of data with “Leads off” or “Bluetooth Loss Error” in the belt.

<sup>#</sup> The prespecified margins are compared to confidence intervals with corresponding confidence levels (see SAP for more details).

<sup>°</sup>Note: all missed bradycardias are checked for clinical relevance by two independent experts.

\*Since the reference devices for apnea detection in the clinical practice are the peripheral oxygen saturation (SpO<sub>2</sub>) and electrocardiogram instead of the respiration signal and the performance for Chest Impedance to detect tachypnea is unsatisfactory due to the presence of cardiac

## Sample size

Based on preliminary analysis of data collected in a feasibility study on a total of 13 infants with measurements from both the investigational device and the reference device, we were able to obtain preliminary information with regards to the characteristics of the primary endpoints



1  
2  
3 upon which we have formulated our sample size calculation.(submitted for publication, NICU  
4  
5 AmsterdamUMC, 2021)  
6  
7  
8  
9

10 A detailed specification of the sample size calculation can be found in the sections below. In  
11 summary, for the monitor performance, 39 infants are needed to achieve 80% power with a 5%  
12 overall type I error with a Bonferroni correction for multiplicity. It is worth noting that no  
13 dropout was assumed during the sample size calculation. This is because we plan to include an  
14 extra infant in case of withdrawal of an infant to fulfil the required sample size. Infants who  
15 withdraw from the study will be followed up by one of the investigators and responsible medical  
16 staff to obtain detailed reasons behind the withdraw. Dropout rate for the monitor performance  
17 study is expected to be low, between 0-5%.  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29

30 While the preplanned sample size is 39 infants, we will include an adaptive sample size re-  
31 estimation procedure as per the “promising zone” methodology of Mehta and Pocock (2) using  
32 the data from the first 1/3 infants. This procedure involves the evaluation of conditional power  
33 in the interim analysis, and if it were to fall in the prespecified “promising zone”, the sample  
34 size will be increased, subject to a predetermined upper limit (52 infants) to increase the  
35 conditional power to 80%. The boundary of the conditional power for the “promising zone” is  
36 0.36 and 0.8.  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48

49 Monitoring study: Primary endpoints: Heart rate

50  
51 For the sample size calculation, we will assume the measured heart rate difference  $D_{ij}$  between  
52 the investigational device and the reference device at time point  $j$  ( $j = 1, \dots, m$ ) on infant  $i$  ( $i =$   
53  
54  
55  
56  
57  
58  
59  
60  
1, ...,  $n$ ) can be modelled as:

$$D_{ij} = d + a_i + e_{ij}$$

where  $d$  is the overall difference,  $a_i$  is a random effect with  $a_i \sim N(0, \sigma_a^2)$ , and  $e_{ij} \sim N(0, \sigma_e^2)$  is the random error independent of  $a_i$ . Though, we considered a bivariate mixed-effects model for our analysis, the variance component model for the difference can be derived from the bivariate mixed-effects model, therefore we will use this variance component model for the sample size calculation. The variance of the difference will be estimated from the aforementioned model via  $\hat{\sigma}_d^2 = \hat{\sigma}_a^2 + \hat{\sigma}_e^2$ . Here  $\hat{\sigma}_a^2$  and  $\hat{\sigma}_e^2$  is the estimator of the between-subject variability  $\sigma_a^2$  and residual variability  $\sigma_e^2$ , respectively. The 95% limit of agreement (LOA) can be estimated as  $\text{LOA} = \hat{d} \pm 1.96 \hat{\sigma}_d$  with  $\hat{d}$  and  $\hat{\sigma}_d$  denotes the estimator of  $d$  and  $\sigma_d$ , respectively. The variance of the LOA estimator is  $\text{var}(\hat{d} \pm 1.96 \hat{\sigma}_d) = \text{var}(\hat{d}) + 1.96^2 \text{var}(\hat{\sigma}_d)$  ( $\hat{d}$  and  $\hat{\sigma}_d$  is asymptotically independent). Since for  $\hat{\sigma}_d^2 = \hat{\sigma}_a^2 + \hat{\sigma}_e^2$ , we have  $\text{var}(\hat{\sigma}_d^2) = \text{var}(\hat{\sigma}_a^2) + \text{var}(\hat{\sigma}_e^2) + \text{cov}(\hat{\sigma}_a^2, \hat{\sigma}_e^2)$ . Furthermore, each term on the right-hand side (assuming  $m$  is large) is given by:

$$\text{var}(\hat{\sigma}_a^2) = \frac{2}{m^2} \left[ \frac{(m\sigma_a^2 + \sigma_e^2)^2}{n-1} + \frac{\sigma_e^4}{n(m-1)} \right] \approx \frac{2\sigma_a^4}{n-1}, \text{var}(\hat{\sigma}_e^2) = \frac{2\sigma_e^4}{n(m-1)+2} \approx 0,$$

$$\text{cov}(\hat{\sigma}_a^2, \hat{\sigma}_e^2) = -\frac{2\sigma_e^4}{nm(m-1)} \approx 0;$$

This leads to  $\text{var}(\hat{\sigma}_d^2) \approx 2\sigma_a^4/(n-1)$ . Therefore, by the delta method, we have  $\text{var}(\hat{\sigma}_d) = \frac{1}{4\sigma_d^2} \text{var}(\hat{\sigma}_d^2) = \frac{\sigma_a^4}{2(n-1)\sigma_d^2}$ . According to Lu et al. (10), the power for the TOST is given by:

$$1 - \beta = 1 - T_{n-1} \left( t_{1-\frac{\alpha}{2}}, \frac{\delta - d - 1.96\sigma_d}{se_{LOA}} \right) - T_{n-1} \left( t_{1-\frac{\alpha}{2}}, \frac{\delta + d - 1.96\sigma_d}{se_{LOA}} \right)$$

where  $\alpha$ ,  $\beta$  denotes type I and type II error respectively,  $\delta$  is the predefined limit,  $se_{LOA} \approx$

$\sqrt{\frac{\sigma_d^2}{n} + \frac{1.96^2 \sigma_a^4}{2(n-1)\sigma_d^2}}$  is the standard error of the LOA estimate calculated according to the variance

component model, and  $T_{n-1}(\cdot, \tau)$  denotes the cumulative distribution function of a non-central Student's t-distribution with  $n-1$  degrees of freedom, and non-centrality parameter  $\tau$ .

For a 5% overall type I error rate, with a multiplicity correction factor of 3, and 80% power, the minimum sample size required is calculated at  $n = 39$ , for  $d = -0.5$ ,  $\sigma_a = 0.3$ , and  $\sigma_d = 3$ .

Primary endpoints: Brady-/tachycardia event detection

Suppose the total number of true events is  $M$  and are 100% detected by the reference device.

Assuming the true sensitivity is 95% for the investigational device, then a non-inferiority test using Z-test with normal approximation to the binomial distribution leads to a required  $M$  of 271 for a power of 80% and  $\alpha = 0.05/3 \approx 0.01667$  assuming the detection between each event (conditioning on the event itself) is independent. Considering the incidence of bradycardia to be 1 event per hour per infant according to the preliminary analysis of data from the feasibility study, at least 12 infants are needed to satisfy the required  $M$  (assuming each infant is measured for 24 hours long). The calculation is the same for PPV if we assume the investigational device is the truth. Assuming an incidence rate of tachycardia of 1.5 per hour per infant according to the preliminary analysis of data from the feasibility study, the required sample size is 8. Note that in the aforementioned calculation, we assume that the event-detection performance of the investigational device is homogeneous (or independent) among infants. A

1  
2  
3 sensitivity/robustness investigation regarding the sample size for infant-specific heterogeneous  
4 performances was performed, with results from which we can see that with  $n = 39$ , we have  
5 more than 90% power to detect a heterogeneous performance scenario where 15% of the  
6 population would have sensitivities between 80% - 90% and less than 5% of the population  
7 have sensitivities less than 80%.  
8  
9  
10  
11  
12  
13  
14  
15  
16

#### 17 Primary endpoints: Safety and quality

18  
19 Based on preliminary analysis of data collected in the feasibility study, we will assume that the  
20 overall probability of producing an erroneous reading at any time  $p_e$  is 2% and is constant  
21 across all participants. We will consider a non-inferiority test using normal approximation and  
22 a Z-test with the null hypothesis of  $H_0: p_e > 0.05$  and the alternative hypothesis of  $H_A: p_e \leq$   
23 0.05. The required number of observations for a given type I error of 1.667% ( $\approx 5\%/3$ ) to  
24 achieve 80% power is 376. Here the sample size 376 refers to 376 independent observations.  
25  
26 Considering the large numbers of repeated measurements (more than 376) within each  
27 participant, we will have sufficient power for this non-inferiority test even with 1 participant.  
28  
29 However, the assumption of independence can be too strong in the setting of our study.  
30  
31 Therefore, if we would assume an AR(1)-type dependency with correlation parameter  $\rho = 0.8$   
32 between two measurements within a participant, the variance inflation factor (VIF) for the  
33 asymptotic variance of the GEE estimator  $\hat{p}_e$  according to Pan is approximately (with the  
34 number of repeats  $m = 376$ ):  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52

$$1 + \frac{2\rho}{1 - \rho} = 9$$

53 in the case of identity working correlation matrix when the true correlation has an AR(1)  
54 structure. To achieve the same power as the independent case calculated before, we need  
55  
56  
57  
58  
59  
60

$$\text{var}(\hat{p}_e) := \text{VIF} \frac{p_e(1-p_e)}{nm} = \frac{p_e(1-p_e)}{m}$$

Thus, we can conclude that at least  $n = \text{VIF} = 9$  participants will be needed to provide enough power for the non-inferiority test based on the GEE estimator using the identity working correlation matrix using the inverse proportionality between the required sample size and the variance of the estimator used in the Z-test. The same calculation can be carried out for the robust data percentages. It can be seen that only the number of repeats  $m$  will differ when the probabilities and the non-inferiority margins change while the VIF remains the same for the same value of the correlation parameter  $\rho$ . Among all settings, the largest  $m$  needed will be 718 when we assume the probability of producing robust data for respiratory rate is 75% with the corresponding non-inferiority margin equals to 70%. This number of repeats is still fully covered by the high-frequency measurements found in the study.

### Protocol deviations and analysis sets

#### Definition of protocol deviations

Protocol deviations (PD) occurring during the study will be determined for all enrolled infants, mainly from the clinical database by either clinical and/or medical review processes.

The mapping of the protocol deviations from the clinical database to analysis will be performed as per Table 2:

Table 2 – The influence of protocol deviations on the statistical analysis plan (SAP)

<i>Database label</i>	<i>SAP</i>
<i>Minor</i>	Not required
<i>Major</i>	Important
<i>Critical</i>	Important
<i>Clinical (a subset of Critical)</i>	Important

Important protocol deviations are protocol deviations that might significantly affect the completeness, accuracy, and/or reliability of the study data or that might significantly affect a subject's rights, safety, or well-being.

Important protocol deviations may also be recorded as "Major" protocol deviations in the database, but will be presented only as important in the analysis output.

Important protocol deviations include:

- Infants that are included in the study despite not satisfying the eligibility criteria;
- Infants that develop exclusion criteria while on the study but not withdrawn;
- Infants being measured with operational human errors;
- Deviation from Good Clinical Practice (ICE E6)

Clinically Important protocol deviations are the protocol deviations marked as important in Table 2, which lead to the exclusion of a subject from the analysis set.

The following deviations will be identified and confirmed before the partial database lock for the final analysis.

- Important protocol deviations including
  - Deviations from the inclusion and exclusion criteria
  - Deviations post inclusion

1  
2  
3 Protocol deviations may be identified by the data managers, clinical and medical staff either by  
4  
5 programmed validation checks or data listings/reports or manual verification of data sources.  
6  
7 Some important/major protocol deviation criteria may be identified in the clinical database via  
8  
9 biostatistical programs. Every important protocol deviation will be documented in the database  
10  
11 whether identified through sites monitoring, medical review or programming.  
12  
13  
14  
15  
16

## 17 REFERENCES

- 18  
19  
20 1. European Medicines Agency. ICH Topic E 9 – Statistical Principles for Clinical Trials.  
21  
22 2006:21-3.  
23  
24
- 25 2. Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are  
26  
27 promising: a practical guide with examples. *Stat Med* 2011;30(28):3267-84. doi:  
28  
29 <https://doi.org/10.1002/sim.4102>  
30  
31
- 32 3. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat*  
33  
34 *Methods Med Res* 1999;8(2):135-60. doi: 10.1177/096228029900800204  
35  
36  
37
- 38 4. Shieh G. The appropriateness of Bland-Altman's approximate confidence intervals for  
39  
40 limits of agreement. *BMC Med Res Methodol* 2018;18(1):45. doi:  
41  
42 <https://doi.org/10.1186/s12874-018-0505-y>  
43  
44
- 45 5. Lin L. I.-K. A Concordance Correlation Coefficient to Evaluate Reproducibility.  
46  
47 *Biometrics* 1989:255-68.  
48  
49
- 50 6. Stevens NT, Steiner SH, MacKay RJ. Assessing agreement between two measurement  
51  
52 systems: An alternative to the limits of agreement approach. *Stat Methods Med Res*  
53  
54 2017;26(6):2487-504. doi: 10.1177/0962280215601133  
55  
56  
57  
58  
59  
60

- 1  
2  
3 7. Lin L. I.-K. Total deviation index for measuring individual agreement with applications  
4 in laboratory performance and bioequivalence. *Statistics in Medicine* 2000;255-70.  
5  
6  
7  
8 8. Haber M, Barnhart HX. A general approach to evaluating agreement between two  
9 observers or methods of measurement from quantitative data with replicated measurements.  
10 *Stat Methods Med Res* 2008;17(2):151-69.  
11  
12  
13 9. Li C. Little's Test of Missing Completely at Random. *The Stata Journal* 2013. p. 795-  
14 809. doi: 10.1177/1536867X1301300407  
15  
16  
17  
18 10. Lu MJ, Zhong WH, Liu YX, et al. Sample Size for Assessing Agreement between Two  
19 Methods of Measurement by Bland-Altman Method. *Int J Biostat* 2016;12(2). doi: 10.1515/ijb-  
20 2015-0039  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60