BMJ Open Respiratory Research

New method to measure interbreath intervals in infants for the assessment of apnoea and respiration

Tricia Adjei,[1] Ryan Purdy,[1] João Jorge,[2] Eleri Adams,[3] Miranda Buckle,[1] Ria Evans Fry,[1] Gabrielle Green,[1] Chetan Patel,[4] Richard Rogers,[5] Rebeccah Slater,[1] Lionel Tarassenko,[2] Mauricio Villarroel,[2] Caroline Hartley [1]

To cite: Adjei T, Purdy R, Jorge J, *et al*. New method to measure interbreath intervals in infants for the assessment of apnoea and respiration. *BMJ Open Resp Res* 2021;**8**:e001042. doi:10.1136/bmjresp-2021-001042

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjresp-2021-001042).

RP and JJ contributed equally.

Received 30 June 2021
Accepted 18 November 2021

[1]Department of Paediatrics, University of Oxford, Oxford, UK
[2]Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK
[3]Newborn Care Unit, Oxford University Hospitals NHS Foundation Trust, Oxford, UK
[4]Department of Ophthalmology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK
[5]Department of Anaesthetics, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

Correspondence to
Caroline Hartley;
caroline.hartley@paediatrics.ox.ac.uk

ABSTRACT

Background Respiratory disorders, including apnoea, are common in preterm infants due to their immature respiratory control compared with term-born infants. However, our inability to accurately measure respiratory rate in hospitalised infants results in unreported episodes of apnoea and an incomplete picture of respiratory activity.

Methods We develop, validate and use a novel algorithm to identify interbreath intervals (IBIs) and apnoeas in preterm infants. In 42 preterm infants (1600 hours of recordings), we assess IBIs from the chest electrical impedance pneumograph using an adaptive amplitude threshold for the detection of breaths. The algorithm is refined by comparing its accuracy with clinically observed breaths and pauses in breathing. We develop an automated classifier to differentiate periods of true apnoea from artefactually low amplitude signal. We assess the performance of this algorithm in the detection of morphine-induced respiratory depression. Finally, we use the algorithm to investigate whether retinopathy of prematurity (ROP) screening alters the IBI distribution.

Results Individual breaths were detected with a false-positive rate of 13% and a false-negative rate of 12%. The classifier identified true apnoeas with an accuracy of 93%. As expected, morphine caused a significant shift in the IBI distribution towards longer IBIs. Following ROP screening, there was a significant increase in pauses in breathing that lasted more than 10 s (t-statistic=1.82, p=0.023). This was not reflected by changes in the monitor-derived respiratory rate and no episodes of apnoea were recorded in the medical records.

Conclusions We show that our algorithm offers an improved method for the identification of IBIs and apnoeas in preterm infants. Following ROP screening, increased respiratory instability can occur even in the absence of clinically significant apnoeas. Accurate assessment of infant respiratory activity is essential to inform clinical practice.

Key messages

- Can we improve the detection of apnoeas and respiratory activity in infants?
- We develop, validate and use a novel algorithm to identify interbreath intervals (IBIs) and apnoeas in infants, demonstrating improved sensitivity compared with the monitor-derived respiratory rate and clinically documented apnoeas.
- Respiratory disorders are common in preterm infants but without better measurement of respiratory activity, we will not fully understand the pathology and improve the treatment of these disorders.

INTRODUCTION

Immature respiratory control in premature infants results in irregular patterns of breathing, with frequent pauses in breathing of variable duration.[1] Apnoea (often defined as a pause in breathing lasting more than 20 s, or shorter if associated with a bradycardia or oxygen desaturation[2 3]) is a common pathology of prematurity, affecting more than 50% of preterm infants.[3] These events can be life-threatening, result in reduced tissue oxygenation[4] and may have long-term effects including reduced cognitive ability in childhood.[5 6] Respiratory disorders are a common reason for admission to a neonatal unit.[7] An infant's respiratory activity may also be affected by pathologies including sepsis,[8] pharmacological interventions including caffeine[9 10] (administered as a treatment for apnoea of prematurity) and opioids[11] (respiratory depressants) and painful clinically indicated procedures such as retinopathy of prematurity (ROP) screening.[12] Despite the high prevalence of problems with respiratory control, clinical measurement of infant respiration is inadequate.[13 14] While clinicians can rely on other physiological measurements to initiate the treatment of apnoeic episodes (eg, reductions in oxygen saturation and heart rate occur during prolonged pauses in breathing), self-resolving apnoeas may be missed[14] and more subtle changes in respiratory activity will not be observed. Accurate assessment of respiration is essential to inform clinical practice and to understand respiratory development in health and disease.

Infants' physiological data are continuously monitored in neonatal intensive care. Respiration is often computed by measuring changes in the electrical impedance of a patient's thorax using the same electrodes that monitor the electrocardiograph (ECG). The use of impedance pneumography (IP) to assess respiratory function has known limitations, in particular susceptibility to noise, and was found by Lim *et al* to be less accurate than capsule pneumography;[15] however, IP remains popular. Commercially available physiological monitors use built-in algorithms to process the IP signal and calculate the respiratory rate, often through the identification of peaks in the signal classified as breaths as a result of a specified amplitude threshold being exceeded.[16–18] However, this approach is limited due to cardiac interference and artefacts caused by non-respiratory-related movements.[13 16 17] Moreover, the manufacturers of many physiological monitors warn that their methods have yet to be validated for apnoea detection in infants.[16 17] Research investigations have demonstrated the limitations of these monitors with high false-alarm rates and missed apnoeas.[13 14] Lee and colleagues previously developed an algorithm to remove cardiac-frequency noise from the IP signal and demonstrated improved performance compared with built-in physiological monitor algorithms in the detection of neonatal apnoeas with a false positive rate of 5% and a false negative rate of 2.5%.[13] However, they note that low amplitude signal related to factors such as poor electrode positioning or shallow breathing can be falsely identified as apnoeas; in a sample of 114 built-in apnoea monitor alarms, Lee reported that almost two-thirds were found to be false by clinicians (and similar rates have been found in other studies[19]). While their algorithm reduced this false alarm rate substantially to 37%,[13] artefactually low amplitude signal remains a problem in apnoea detection. Additionally, accurate assessment of interbreath intervals (IBIs), and not just the identification of apnoeas as in the work of Lee *et al* is needed to gain a better understanding of the effects of pathology and interventions on respiration. For example, the assessment of more subtle changes in IBIs will improve classification of underlying pathology and may allow for the early detection and prediction of apnoeas.[20]

Here we develop a new method for identifying IBIs and apnoeas (defined here as pauses in breathing of at least 20 s) from an infant's IP signal. We then use the algorithm to check its sensitivity to detect respiratory depression following morphine administration. Finally, we investigate changes in IBIs following ROP screening.

METHODS

Study design

We designed, validated and tested our algorithm using three separate data sets. Data set 1 was used to determine the optimal threshold parameters for breath detection in the IP signals, by comparing the breaths identified with the algorithm to those manually annotated by clinical staff. Data set 2 was first used to verify that the parameters identified using data set 1 could detect pauses in respiration of at least 5 s. It was then used to develop and validate a classifier to detect true central apnoea as opposed to artefactually low amplitude signal. We then tested the algorithm, exploring its ability to identify morphine-induced respiratory depression, using a subset of data set 2. Finally, data sets 2 and 3 were used to evaluate changes in IBIs following ROP screening.

Study participants

A total of 42 infants were included in this study. Data set 1 was collected as a subset of the MONITOR study.[21] It comprises 181 sequences of approximately 40 breaths each (in total 7632 breaths), recorded from five preterm infants (postmenstrual age (PMA) at study range 30.6–34.3 weeks). Each breath was manually annotated by clinical staff in real time by visual observation of the infant. Data set 2 comprised physiological data collected during the Poppi trial, a single-centre, masked, randomised, placebo-controlled trial which investigated whether oral morphine was an effective and safe analgesic for procedural pain in premature-born infants.[11] Physiological data were collected for 24 hours before and after the clinical procedure—a heel lance followed by ROP screening—in 30 infants (15 received morphine, 15 received placebo, PMA at study 34–39 weeks). Data set 3 is a previously unpublished data set of seven infants (PMA at study 30–37 weeks) whose physiological data were recorded before and after ROP screening. Further details for all studies are given in the online supplemental methods.

All data sets were collected at the Newborn Care Unit, John Radcliffe Hospital (Oxford University Hospitals NHS Foundation Trust, Oxford, UK). Written informed parental consent for all three data sets was gained. Approval was obtained from South Central Research Ethics Committee (REC) (13/SC/0597) for the MONITOR study, the Medicines and Healthcare products Regulatory Agency (MHRA) and Northampton REC (15/EM/0310) for the Poppi trial and from South Central REC (12/SC/0447) for data set 3. All studies conformed to the standards set by the Declaration of Helsinki.

Physiological recordings

All infants were monitored using a Philips IntelliVue MX800 monitor, and physiological data were continuously downloaded from the monitor using ixTrend software (ixitos GmbH, Germany). Further details are given in the online supplemental methods.

Breath detection from the IP signal

The algorithm presented here to identify IBIs from the IP signal consists of three main steps (figure 1):

1. Removal of artefacts.
2. Application of an adaptive threshold to identify breaths.
3. Identification of true apnoeas using support vector machine (SVM) classification.

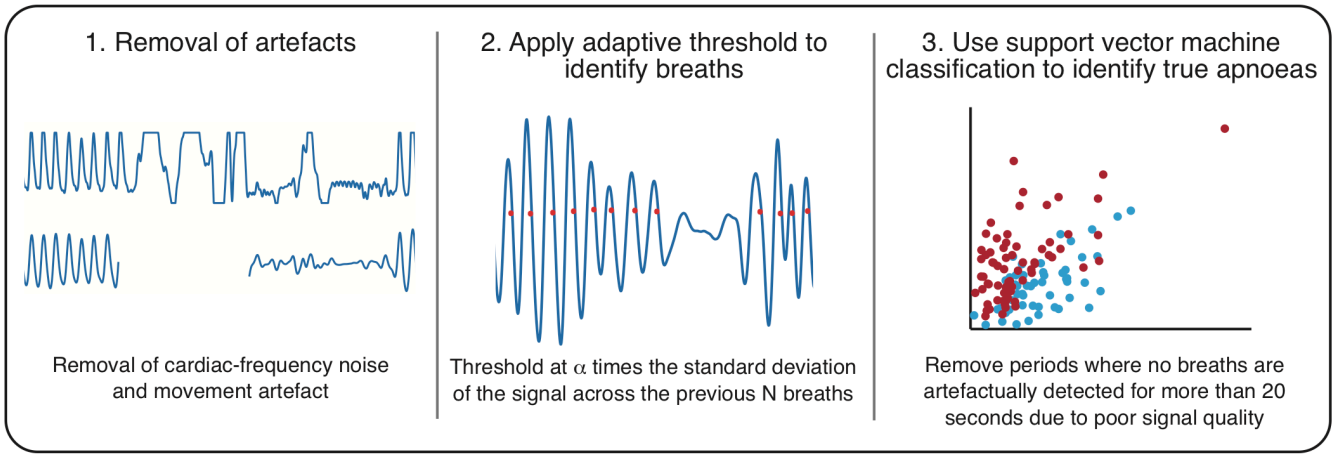


Figure 1 Schematic of the proposed algorithm for detection of interbreath intervals (IBIs) from the impedance pneumograph (IP) in infants.

The code for this algorithm is available from https://gitlab.com/paediatric_neuroimaging/identify_ibi_from_ip.git. For further details of all parts of the algorithm, see the online supplemental methods, figures 1,2 and tables 1,2. Briefly, first, the IP signals were filtered to remove artefacts not related to respiration, for example, large-amplitude changes caused by movements of the infant, and cardiac-frequency noise; the filtering process also zeroed the IP signals. Second, individual breaths were identified from the IP signal as the point at which an adaptive threshold is crossed (an adaptive threshold, ie, one that is updated across the recording,[22–24] was used to account for changes in the amplitude of the signal for a variety of physiological and non-physiological reasons, such as shallow breathing and changes in the electrode and infant positioning). We identified the optimal threshold parameters for breath detection by comparing the breaths detected by the algorithm for different parameters with recordings where individual breaths were annotated in real time by a clinical member of staff visually observing the infant's breathing (data set 1). The optimal parameters were chosen to be values which achieved the best compromise between the percentages of false positives and false negatives. We then verified that these parameters were also suitable for detection of pauses in breathing with a duration greater than 5 s by comparison of pauses in breathing detected by the algorithm with those that were retrospectively identified by two investigators (data set 2, first hour of recording, in 15 infants).

Finally, a linear SVM classifier was used to identify true central apnoeas (defined here as IBIs≥20 s) as opposed to artefactually low amplitude signal. The model input features are the magnitude (root-mean-square) of the filtered IP signal during the apnoea, in the 10 s prior to the apnoea, and in the 10 s after the end of apnoea, and the change in heart rate and oxygen saturation in the 60 s from the onset of the apnoea. The model was trained and tested using labels (true apnoea/false alarm) provided by two investigators for all potential apnoeas identified in data set 2 (training set, 15 infants who received morphine, test set, 15 infants who received placebo); 24% of potential apnoeas were classed differently by the two investigators and so were not included in the analysis.

Performance of apnoea identification

To compare the accuracy of our approach with the accuracy of the current standard, all periods where the monitor-derived respiratory rate reached 0 were viewed by two investigators (see online supplemental material) and rated according to whether the investigator thought this period was a true central apnoea or a false alarm (90% inter-rater agreement occurred here). To compare with Lee *et al*, all episodes of apnoea accompanied by bradycardia (<100 bpm) and oxygen desaturation (<80%) detected by the algorithm were compared with investigator ratings to calculate the false-positive rate. To calculate the false-negative rate, all episodes of bradycardia (<100 bpm for at least 15 s) were identified in the recordings. Those that were not accompanied by a pause in breathing of at least 5 s (IBI >5) identified by the algorithm were rated according to whether the investigator thought a pause (true positive) in breathing occurred during this episode.

Comparison with medical records

The time of apnoeas identified by our algorithm was compared with apnoeas documented in each infant's medical records and nursing observation charts, specifically on the apnoea/bradycardia/desaturation chart (the term medical records is used to describe both in the rest of the paper). Apnoeas are documented if the clinical/nursing staff observe the infant having an episode of apnoea along with a description of how they were resolved, that is, self-resolving, requirement for increased oxygen, requirement for stimulation or requirement for resuscitation.

Use of the algorithm to evaluate respiratory depression following morphine administration

We tested the algorithm by examining the changes in the IBI distribution following morphine administration in the 15 infants in data set 2 who received morphine. In the Poppi trial, we previously demonstrated a significant decrease in the respiratory rate (recorded on the monitor) in the morphine-treated infants compared with the placebo-treated infants, with a peak decrease approximately 2.5 hours following drug administration.[11] We examined the IBI distribution in the 1 hour period prior to drug administration and the 1 hour period after the clinical procedure (from the end of the ROP screening, on average 1.3–2.3 hours after drug administration), and calculated the mean, median and SD of the IBI distributions, the proportion of IBIs longer than 5 s and the proportion of IBIs longer than 10 s (time periods commonly used to assess pauses in breathing[2]). We compared this with the mean monitor-derived respiratory rate calculated for the same periods.

Use of the algorithm to evaluate changes in IBIs following ROP screening

We used the algorithm to investigate changes in the IBI distribution following ROP screening in a total of 22 infants—the 15 infants who received placebo in data set 2 and the seven infants in data set 3. In the placebo-treated infants, we compared the 1-hour period prior to placebo administration with the 1-hour period after the clinical procedure. In data set 3, we similarly compared the 1-hour after ROP screening with the 1-hour period 2.3–1.3 hours prior to ROP screening. We also compared the 12-hour period before and after ROP screening in the subset of 19 infants with at least 12 hours of recording before and after ROP screening.

Statistical analysis

All data analysis was undertaken with MATLAB 2019b (MathWorks, USA). Model performance of the SVM classification was assessed with accuracy, false-positive rate, false-negative rate and Matthew's correlation coefficient (MCC) using leave-one-subject-out cross-validation in the training set and independently in the test set using the model constructed from all infants in the training set. Differences in the IBI distribution and mean respiratory rate before and after morphine administration and ROP screening were compared using paired non-parametric t-tests with statistical significance assessed using permutation testing (10 000 random permutations) performed using FSLs PALM software.[25] P values were adjusted for multiple comparisons using Hochberg's method in R (The R Project for Statistical Computing).

Patient and public involvement

A parent focus group, organised in collaboration with the charity SSNAP (Supporting the Sick Newborn and their Parents, a local charity based on the Newborn Care Unit at the John Radcliffe Hospital, Oxford), was held to discuss the Poppi Trial (data set 2) prior to the trial starting.

RESULTS

Optimising the adaptive threshold

A threshold of 0.4 times the SD of the filtered IP signal for the 15 previous breaths provided a good compromise between the false-positive and false-negative rates of breath detection (figure 2A, online supplemental results, figure 3). At this threshold, a mean (across all recordings in data set 1) of 12% of the manually annotated breaths were missed by the algorithm (false negatives), and 13% of breaths detected by the algorithm were false positives.

We examined whether these threshold parameters could also accurately identify pauses in breathing of at least 5 s. Using the same parameters, 13 pauses out of the 162 identified by both investigators were missed by our algorithm (false-negative rate: 8%) and 44 pauses out of the 229 identified by the algorithm were not identified by either investigator (false-positive rate: 19%). Varying the parameters confirmed that those selected achieved a good balance between false positives and false negatives (figure 2B).

Optimising apnoea detection using machine learning

Applying the adaptive threshold to all recordings from data set 2 identified a total of 164 potential apnoeas. Of these episodes, 68 (41%) were classified by both investigators as true apnoeas and 57 (35%) were classified by both investigators as false alarms (no agreement for 39 (24%) episodes). This already represents a major improvement in detection rate from the monitor-derived respiratory rate—of the 71 occasions for which the monitor-derived respiratory rate reached a value of 0 breaths per minute, two episodes were classified by both investigators as true apnoeas (3%) and 62 (87%) were classified by both investigators as false alarms.

An SVM classifier was trained to distinguish between episodes detected by the adaptive threshold and classify them as either true apnoeas or false alarms (examples shown in figure 3A,B). In the training set (15 infants), using features derived from the IP signal alone (figure 3C, online supplemental methods), the classifier had an accuracy of 75% in the detection of true apnoeas (MCC=0.49, 62% of 69 episodes in the training set were true apnoeas). Including additional features related to the change in oxygen saturation and heart rate as inputs to the classifier and retraining it on the same training set increased the accuracy to 87% (MCC=0.74, false-positive rate=5%, false-negative rate=16%, figure 3D). Applying the best classification model to the test set gave an accuracy of 93% (MCC=0.87, false-positive rate=14%, false-negative rate=0%, 25 of 56 episodes in the test set were true apnoeas), validating the model in this independent data set.

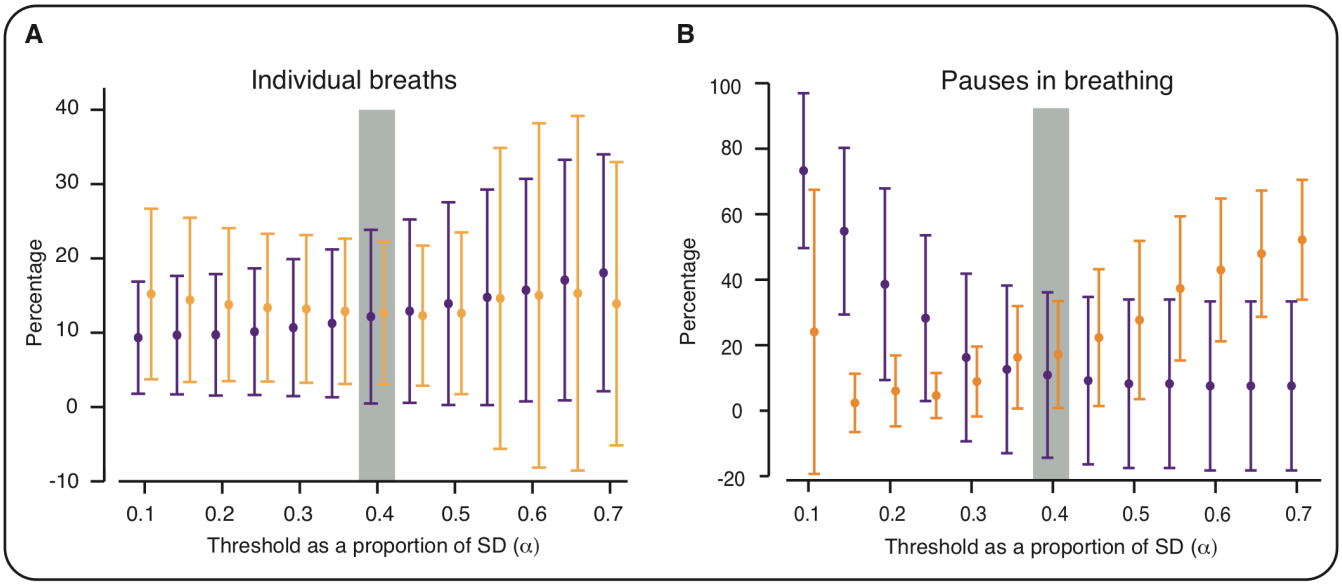


Figure 2 Optimising the threshold for breath detection. To optimise the threshold parameters, we investigated the performance of different threshold values (defined as a multiple (α) of the SD of the IP signal across the previous N breaths) to identify individual breaths and pauses in breathing. Figures show the percentage of false positives (orange) and false negatives (purple) for different values of α (with N=15). (A) Values calculated by comparing algorithm-identified breaths with breaths manually annotated at the time of the recording by visual observation (data set 1). (B) Values calculated by comparing algorithm-identified pauses in breathing with pauses (of at least 5 s) manually annotated by two investigators (first hour of recording in 15 infants from data set 2). Error bars indicate mean and SD (across the recordings). Values are jittered on the X-axis so that false positive and false negative bars do not overlap. Grey shading indicates selected threshold parameters; with these parameters (α=0.4, N=15), there was the optimal balance between the percentages of false positives and false negatives in the identification of individual breaths (A). These parameters also achieved a good balance between false positives and negatives in the identification of pauses in breathing (B).

For comparison with Lee *et al*,[13] we performed three analyses. First, we assessed the performance of our algorithm for the detection of apnoeas with co-occurring bradycardia and oxygen desaturation. Of 26 such episodes (in both training and test set) detected by our algorithm, all were classified as true apnoeas by both investigators (0% false-positive rate). Second, we assessed the performance of our algorithm for the detection of pauses in breathing associated with bradycardias. A total of 109 episodes of bradycardia occurred in our data. Of the 13 episodes of bradycardia where a pause in breathing was not detected by our algorithm, only three were thought to be associated with pauses in breathing (3% false-negative rate) by the investigators. Finally, of the 62 false alarms where the monitor-derived respiratory rate reached a value of 0 breaths per minute, 3 (5%) were detected as apnoeas using the adaptive threshold alone. After applying the SVM classifier, none of these were detected as apnoeas by our algorithm (0% false alarm rate).

Comparison with medical records

Of the 60 true apnoeas identified by our method, 88% were not recorded in the medical records. During the recording period, a total of 24 apnoeas were recorded in the medical records, the majority of which were associated with an IBI of at least 10 s; however, four events were not associated with a prolonged pause in breathing detected by the algorithm but instead with a prolonged loss of signal due to artefacts. We hypothesise that such artefacts were caused by clinical intervention in response to the apnoea.

Use of the algorithm to evaluate respiratory depression following morphine administration

As expected, there was a significant decrease in the monitor-derived respiratory rate following morphine administration (p=0.0004, non-parametric permutation t-test corrected for multiple comparisons, n=15, table 1, figure 4A, n=15). This was reflected in the IBI distribution, which showed a clear shift in the distribution towards longer IBIs following morphine administration (figure 4B), and significant differences in all IBI metrics assessed (figure 4C,D, table 1).

Use of the algorithm to evaluate changes in IBIs following ROP screening

There was a shift in the IBI distribution in the 1 hour following ROP screening towards longer IBIs (figure 4F), with a significant increase in the proportion of IBIs longer than 10 s (p=0.023, figure 4H, table 1, n=22). This was not reflected by a change in the monitor-derived respiratory rate (p=0.89, figure 4E, table 1). Moreover, there was

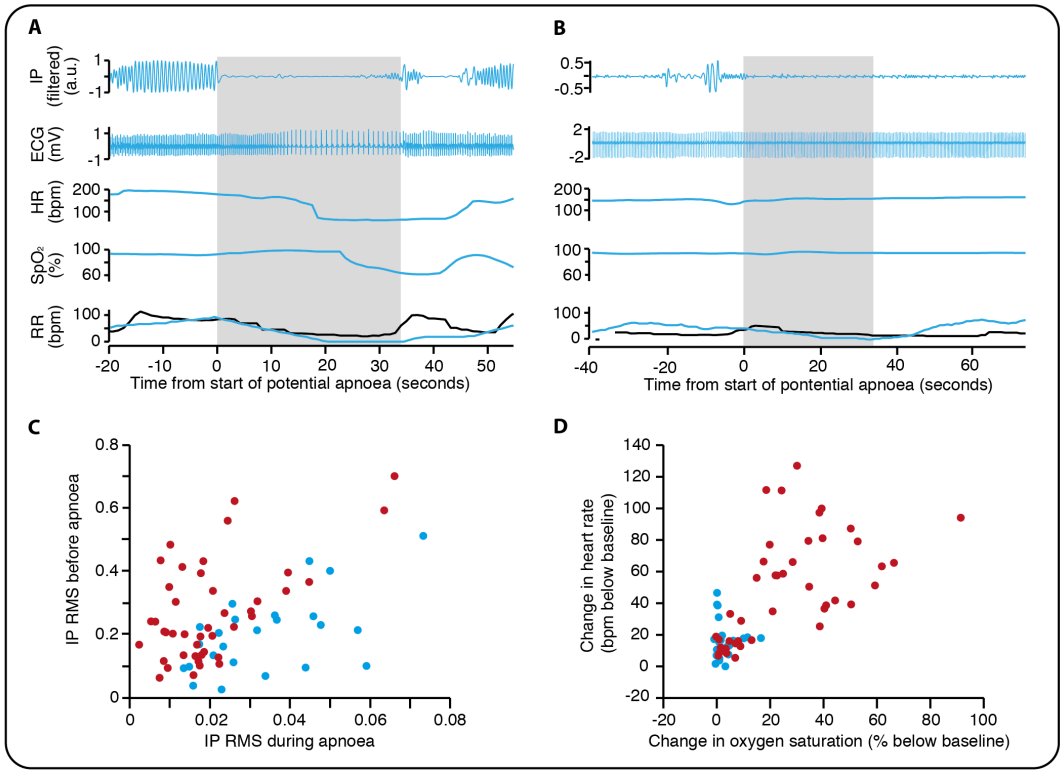


Figure 3 Using support vector machine classification to identify true apnoeas. (A) An example of a pause in breathing lasting longer than 20 s identified as a true apnoea. IP, the electrical impedance pneumograph after filtering to remove cardiac-frequency noise and movement artefact. HR, heart rate in beats per minute. SpO_2, oxygen saturation. RR, respiratory rate in breaths per minute, recorded by the infant's patient monitor (black) and calculated using our algorithm (blue). Note that the RR does not reach zero on the infant's patient monitor and so this episode does not lead to a monitor apnoea alarm. Grey shading indicates the period during which no breaths were detected by our algorithm. (B) A potential apnoea initially detected by the algorithm but classified by investigators as a false alarm. (C) The root mean square (RMS) of the IP signal before and during the apnoea (see Methods for further details). Red circles indicate episodes classified by both investigators as true apnoeas, and blue circles are those episodes classified by both investigators as false alarms. (D) Change in oxygen saturation and heart rate for true apnoeas (red) compared with false alarms (blue).

a significant increase in the proportion of IBIs longer than 10 s in the 12 hours after ROP screening compared with the 12 hours before (p=0.037, t-statistic=1.77, n=19). No apnoeas were recorded in the medical records in the 12 hours before or after ROP screening for any of these infants. Infant demographics are shown in online supplemental table 3.

DISCUSSION

We developed a new algorithm to detect IBIs from the IP signal in infants. Following the removal of cardiac artefact from the IP signals using a method introduced in Lee *et al*[13] we used an adaptive amplitude threshold to identify individual breaths, validating the threshold by comparison with visually identified breaths and pauses in breathing. Previous studies have reported that signals with low amplitude due to poor electrode placement or shallow breathing can be erroneously detected as episodes of apnoea. To overcome this problem, we used machine learning to identify true apnoeas from periods of artifactually low amplitude. We tested our algorithm by investigating changes in IBIs following morphine administration, observing a clear shift in the IBI distribution consistent with the reduction in respiratory rate seen on the infants' patient monitors. Finally, we used our algorithm to investigate changes in IBIs following ROP screening. We observed a significant shift in the IBI distribution following ROP screening which was not reflected by a change in the monitor-derived respiratory rate. This demonstrates the increased sensitivity of our method in detecting changes in respiratory activity, compared with the monitors and highlights the increase in physiological instability in infants following ROP screening.

Premature infants are born with immature cerebral and respiratory function compared with term-born infants, and consequently have a higher incidence of respiratory disorders. Current inadequacies in the measurement of respiration in infants leads to missed opportunities to better understand respiratory development and could potentially lead to suboptimal clinical treatment. For example, caffeine therapy, given for apnoea of prematurity, is stopped in infants between 33 and 35 weeks PMA if the infant appears clinically stable.[26] However, in 10% of infants, caffeine treatment is restarted,[27] which may

Table 1 Changes in interbreath intervals following morphine administration and ROP screening

	Mean before	Mean after	t-statistic	Uncorrected p value	Corrected p value
Morphine (n=15 infants)					
Mean respiratory rate (bpm)	52.01	44.66	–3.54	0.0001	0.0004***
Mean IBI (s)	1.07	1.34	5.18	0.0001	0.0004***
Median IBI (s)	0.93	1.03	3.96	0.0012	0.0012**
SD IBI (seconds)	0.61	1.26	5.86	0.0001	0.0004***
IBI >5 s (%)	0.56	2.54	4.17	0.0004	0.0008***
IBI >10 s (%)	0.02	0.49	3.39	0.0002	0.0006***
ROP screening (n=22 infants)					
Mean respiratory rate (bpm)	51.07	50.92	–0.14	0.89	0.89
Mean IBI (s)	1.09	1.12	1.32	0.20	0.81
Median IBI (s)	0.97	0.98	0.25	0.84	0.89
SD IBI (s)	0.56	0.66	2.45	0.021	0.10
IBI >5 s (%)	0.49	0.63	1.14	0.28	0.83
IBI >10 s (%)	0.02	0.06	1.82	0.0039	0.023*

Comparison of the respiratory rate (recorded by the patient monitor) and interbreath interval (IBI) distribution 1 hour before and after morphine administration and 1 hour before and after ROP screening. The table indicates the mean across all infants in each group, and the t-statistic and p-values for each comparison (permutation test). P-values were corrected for multiple comparisons using Hochberg's method (*p<0.05, **p<0.01, ***p<0.001).
ROP, retinopathy of prematurity.

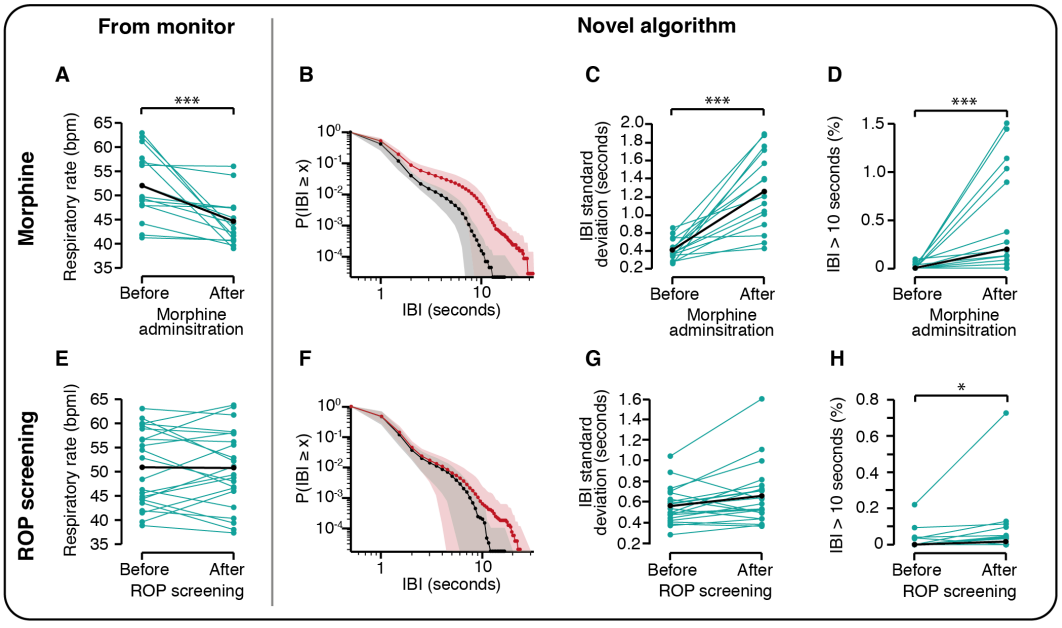


Figure 4 Interbreath intervals are altered by morphine administration and following ROP screening. (A–D) Respiratory rate and interbreath intervals (IBIs) in the 1-hour period prior to morphine administration compared with a 1-hour period after morphine administration (the 1-hour period immediately following ROP screening, approximately 1.3–2.3 hours after morphine administration) in the 15 infants who received morphine in the Poppi clinical trial. (E–H) Respiratory rate and IBIs 1 hour before and after ROP screening in 22 infants. (A, E) Mean respiratory rate from the infants' patient monitor. (B–D, F–H) Metrics calculated using the novel algorithm proposed in this paper to identify the IBIs. Black lines and points indicate the group mean (A, C, E, F) or median (D, H). (B) IBI distribution in the 1-hour period prior to (black) compared with 1.3–2.3 hours after morphine administration (red). (F) IBI distribution in the 1-hour period before (black) and after (red) ROP screening. Y-axis indicates the probability of an IBI of duration greater than or equal to the X-axis value. Dotted line indicates the mean and shaded area the SD. (*p<0.05, **p<0.01, ***p<0.001, p-values corrected for multiple comparisons). ROP, retinopathy of prematurity.

suggest that caffeine was withdrawn too early, exposing the infants to the adverse consequences of lack of treatment. We found that 88% of apnoeas identified using our algorithm were not recorded in the medical records, consistent with previous results highlighting the inaccuracies in clinical documentation of apnoea.[14 28 29] While there are multiple reasons why apnoeas may not be recorded accurately in clinical observations, including the under-recognition of apnoeas that come within periods of periodic breathing, this substantial increase in the number of apnoeas identified using our algorithm demonstrates the potential for improving apnoea detection. Improved measurement of respiration is essential to optimise clinical treatment of apnoea and could enhance treatment for other clinical conditions or procedures which alter respiration.

Many drugs will alter infants' physiology. Our results confirm the applicability of the algorithm to analyse morphine-related respiration depression. Using this approach to investigate respiratory changes in relation to other drugs commonly prescribed in neonatal care may enhance our understanding of pharmacodynamics. Additionally, analysis of physiological recordings may be useful to develop predictive models to tailor individualised care.[30 31] We recently showed in a post-hoc analysis of the morphine-treated infants in the Poppi trial that we could predict the risk of adverse cardiorespiratory effects in individual infants from their baseline physiological stability.[32] To date, measures of respiration are often not included in the development of predictive tools, which is likely due to the relatively poor quality of the currently available measurement tools.[31] Here we provide a more accurate measure of IBIs, which will allow for more complex metrics, such as respiratory rate variability, to be computed.

ROP screening, an eye examination that is thought to be painful and distressing for infants,[33] has previously been shown to increase the rate of apnoea in the 24–48 hours following the screen from clinical chart review.[12] In an exploratory analysis, we demonstrated a significant increase in the proportion of IBIs longer than 10 s in the 1-hour and 12-hour periods after ROP screening, which was not reflected by a change in the monitor-derived respiratory rate. This demonstrates the improved sensitivity of our method for identifying changes in respiratory activity and suggests that even those infants without clinically significant apnoeas may still experience changes in respiratory activity with a shift towards longer IBIs. Further research in a larger cohort across a wider age range is needed to explore the relationship between an infant's respiratory activity following ROP screening and changes with age. Identifying older infants that are at risk of physiological instability after ROP screening would be particularly important for those ex-premature infants who have ROP screening in outpatient clinics and may benefit from observation before leaving hospital.[34]

To remove cardiac-frequency noise from the IP signals, we used the approach of Lee and colleagues,[13] which we modified (online supplemental table S2) predominantly due to the poor performance of the ECG R peak detection used by Lee *et al* in our data. Our algorithm had similar rates of false positives and negatives in the identification of apnoeas to those reported by Lee. Importantly, we also trained a classifier to identify true apnoeas compared with artifactually low amplitude signal; the classifier reduced the false alarm rate compared with using the adaptive threshold alone and to that reported by Lee and colleagues. Additionally, our algorithm used an adaptive threshold to identify individual breaths (with thresholds optimised with and without the prior removal of cardiac-frequency noise). Thus, unlike the algorithm of Lee, our algorithm can be used both in the identification of apnoea and also to examine changes in the pattern of IBIs of an infant.

By using an adaptive threshold which we validated for infants, our algorithm performed substantially better than the monitor derived respiratory rate. However, limitations of this study are the relatively small sample size and narrow age range of the infants included (from 30 to 39 weeks PMA). Further validation should be carried out in younger infants. Moreover, this method identifies central apnoea; it cannot detect obstructive apnoea—alternative measures, such as nasal air flow, are needed to detect these events. Additionally, apnoea that necessitates intervention by clinical staff may not be detected or the reported duration may be shorter than the true duration of the episode as interventions are likely to lead to large artefacts in the IP signal. While this is not a problem for clinical management, as the infant is receiving the appropriate clinical intervention to support their breathing, this should be taken into account in research studies so that apnoeas are not missed in the analysis.

In summary, despite the common occurrence of respiratory pathology in preterm infants, current methods used to measure respiration are inadequate. We developed a new method to measure respiration in infants, demonstrating the improved sensitivity of the method compared with current standards; the increased sensitivity provided by our algorithm could aid clinical teams in the care of infants. Furthermore, we identified a significant increase in respiratory instability in infants following ROP screening. A better understanding of respiratory activity in infants is critical to improve neonatal care.

Acknowledgements We would like to thank the MONITOR Trial team for collecting data set 1, conducting analysis and interpreting results as part of the original MONITOR study. We would like to thank the Poppi Trial team for collecting data set 2, conducting analysis and interpreting results as part of the original Poppi trial. We would also like to thank Gabriela Schmidt Mellado for assistance with data collection for data set 3, Fahiza Begum for assistance with analysis in an early form of this algorithm, and Professor John Delos for sharing the code related to Lee *et al. A new algorithm for detecting central apnea in neonates. Physiol Meas* 2012; 33: 1–17.

Contributors CH conceived the idea for the study and acts as guarantor. TA, RP, JJ, REF, RR, MV and CH conducted the analysis. TA, RP, JJ, RR, RS, MV and CH interpreted the data. TA and CH wrote the first draft of the manuscript. Data set 1 was originally collected for the MONITOR study by JJ, GG, MV, LT and members of the MONITOR study team. Data set 2 was originally collected for the Poppi Trial by CH, GG, MB, RR, CP, EA, RS and members of the Poppi trial team. Data set 3 was

collected by MB and CP. All authors critically reviewed the data and revised the paper.

Funding This work was funded by the Wellcome Trust and Royal Society through a Sir Henry Dale Fellowship (grant reference number: 213486/Z/18/Z)

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The code for the algorithm developed in this paper is available from https://gitlab.com/paediatric_neuroimaging/identify_ibi_from_ip.git.

ORCID iD

Caroline Hartley http://orcid.org/0000-0002-7981-0836

REFERENCES

1 Martin RJ, Abu-Shaweesh JM. Control of breathing and neonatal apnea. *Biol Neonate* 2005;87:288–95.
2 Elder DE, Campbell AJ, Galletly D. Current definitions for neonatal apnoea: are they evidence based? *J Paediatr Child Health* 2013;49:E388–96.
3 Finer NN, Higgins R, Kattwinkel J, *et al*. Summary proceedings from the apnea-of-prematurity group. *Pediatrics* 2006;117:S47–51.
4 Yamamoto A, Yokoyama N, Yonetani M, *et al*. Evaluation of change of cerebral circulation by SpO2 in preterm infants with apneic episodes using near infrared spectroscopy. *Pediatr Int* 2003;45:661–4.
5 Janvier A, Khairy M, Kokkotis A, *et al*. Apnea is associated with neurodevelopmental impairment in very low birth weight infants. *J Perinatol* 2004;24:763–8.
6 Pillekamp F, Hermann C, Keller T, *et al*. Factors influencing apnea and bradycardia of prematurity - implications for neurodevelopment. *Neonatology* 2007;91:155–61.
7 Gallacher DJ, Hart K, Kotecha S. Common respiratory conditions of the newborn. *Breathe* 2016;12:30–42.
8 Fairchild K, Mohr M, Paget-Brown A, *et al*. Clinical associations of immature breathing in preterm infants: part 1-central apnea. *Pediatr Res* 2016;80:21–7.
9 Moschino L, Zivanovic S, Hartley C, *et al*. Caffeine in preterm infants: where are we in 2020? *ERJ Open Res* 2020;6:00330-2019.
10 Schmidt B, Roberts RS, Davis P, *et al*. Caffeine therapy for apnea of prematurity. *N Engl J Med* 2006;354:2112–21.
11 Hartley C, Moultrie F, Hoskin A, *et al*. Analgesic efficacy and safety of morphine in the procedural pain in premature infants (Poppi) study: randomised placebo-controlled trial. *Lancet* 2018;392:2595–605.
12 Mitchell AJ, Green A, Jeffs DA, *et al*. Physiologic effects of retinopathy of prematurity screening examinations. *Adv Neonatal Care* 2011;11:291–7.
13 Lee H, Rusin CG, Lake DE, *et al*. A new algorithm for detecting central apnea in neonates. *Physiol Meas* 2012;33:1–17.
14 Vergales BD, Paget-Brown AO, Lee H, *et al*. Accurate automated apnea analysis in preterm infants. *Am J Perinatol* 2014;31:157–62.
15 Lim K, Eastwood-Sutherland C, Marshall AP, *et al*. Limitations of thoracic impedance monitoring for central apnoea detection in preterm infants. *Acta Paediatr* 2021;110:2550–2.
16 Philips K. IntelliVue patient monitor [online], 2019. Available: https://www.fda.gov/media/137229/download [Accessed 8 Oct 2020].
17 Draeger. Infinity acute care system: instructions for use [online], 2017. Available: https://www.draeger.com/Products/Content/iacs-vg7-monitoring-applications-ifu-ms34093-en.pdf [Accessed 8 Oct 2020].
18 Welch Allyn 1500 patient monitor [online], 2013. Available: https://www.welchallyn.com/content/dam/welchallyn/documents/upload-docs/Training-and-Use/User-Manual/Welch-Allyn-1500-Patient-Monitor-Software-version-1.4.X_User-Manual.pdf [Accessed 8 Oct 2020].
19 Jorge J. *Non-contact monitoring of respiration in the neonatal intensive care unit [PhD Thesis]*. University of Oxford, 2018.
20 Lim K, Jiang H, Marshall AP, *et al*. Predicting apnoeic events in preterm infants. *Front Pediatr* 2020;8:1–7.
21 Jorge J, Villarroel M, Chaichulee S, *et al*. Assessment of signal processing methods for measuring the respiratory rate in the neonatal intensive care unit. *IEEE J Biomed Health Inform* 2019;23:2335–46.
22 Zong W, Moody GB, Jiang D. A robust open-source algorithm to detect onset and duration of QRS complexes. *Computers Cardiol* 2003:737–40.
23 Zong W, Heldt T, Moody GB. An open-source algorithm to detect onset of arterial blood pressure pulses. *Computer Cardiol* 2003:259–62.
24 Karlen W, Ansermino JM, Dumont G. Adaptive pulse segmentation and artifact detection in photoplethysmography for mobile applications. in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2012:3131–4.
25 Winkler AM, Ridgway GR, Webster MA, *et al*. Permutation inference for the general linear model. *Neuroimage* 2014;92:381–97.
26 National Institute for Health and Care Excellence. Specialist neonatal respiratory care for babies born preterm. NICE guideline NG124 2019.
27 Haddad W, Sajous C, Hummel P, *et al*. Discontinuing caffeine in preterm infants at 33-35 weeks corrected gestational age: failure rate and predictive factors. *J Neonatal Perinatal Med* 2015;8:41–5.
28 Brockmann PE, Wiechers C, Pantalitschka T, *et al*. Under-recognition of alarms in a neonatal intensive care unit. *Arch Dis Child Fetal Neonatal Ed* 2013;98:F524–7.
29 Razi NM, Humphreys J, Pandit PB, *et al*. Predischarge monitoring of preterm infants. *Pediatr Pulmonol* 1999;27:113–6.
30 Griffin MP, O'Shea TM, Bissonette EA, *et al*. Abnormal heart rate characteristics preceding neonatal sepsis and sepsis-like illness. *Pediatr Res* 2003;53:920–6.
31 Kumar N, Akangire G, Sullivan B, *et al*. Continuous vital sign analysis for predicting and preventing neonatal diseases in the twenty-first century: big data to the forefront. *Pediatr Res* 2020;87:210–20.
32 Hartley C, Baxter L, Moultrie F, *et al*. Predicting severity of adverse cardiorespiratory effects of morphine in premature infants: a post hoc analysis of procedural pain in premature infants trial data. *Br J Anaesth* 2021;126:e133–5.
33 Belda S, Pallás CR, De la Cruz J, *et al*. Screening for retinopathy of prematurity: is it painful? *Biol Neonate* 2004;86:195–200.
34 Wood MG, Kaufman LM. Apnea and bradycardia in two premature infants during routine outpatient retinopathy of prematurity screening. *J Aapos* 2009;13:501–3.

Supplementary Material: A new method to measure inter-breath intervals in infants for the assessment of apnoea and respiration

Tricia Adjei, Ryan Purdy, João Jorge, Eleri Adams, Miranda Buckle, Ria Evans Fry, Gabrielle Green, Chetan Patel, Richard Rogers, Rebeccah Slater, Lionel Tarassenko, Mauricio Villarroel and Caroline Hartley

Supplementary Methods

Study design and participants

Full details of data collection for the MONITOR study (Data set 1) are given in Jorge et al. 2019 (1). These data were collected in five infants aged between 30 – 34 weeks postmenstrual age (PMA) at the time of study. Each infant's physiological data were recorded for between 2-4 days, during which time manual annotations of trains of approximately 40 breaths were recorded by a clinical member of staff who watched the infant. A total of 181 manual count trains from the five infants were assessed in the present study.

Full details of the recruitment, trial design and trial procedures for Data set 2 from the Poppi (Procedural Pain in Premature Infants) trial are given in Hartley and Moultrie et al. 2018 (2–4). Briefly, infants were aged between 34-39 weeks PMA at the time of study. Infants were given either oral morphine (100 µg/kg, n=15 infants) or a placebo (n=15) approximately 1 hour prior to the clinical procedure - a medically-required heel lance for blood sampling and a retinopathy of prematurity (ROP) screening. As part of the trial, physiological data were recorded for approximately 24 hours before and after the clinical procedure (total duration of recordings: 1503 hours).

Data set 3 comprised of data from 7 infants who were aged between 30-37 weeks PMA at the time of study and had been born at less than 32 weeks' gestation or with a birthweight lower than 1501 g**,** fulfilling national criteria for ROP screening. This was a subset of infants collected as part of an ongoing study to investigate changes in brain activity and physiological activity following ROP screening. All 7 infants selected had recordings for at least 12 hours before ROP screening, and a minimum of 1 hour after ROP screening.

For infants in both Data set 2 and 3, ROP screening was performed using binocular indirect ophthalmoscopy. Mydriatic eye drops (tropicamide 1% and phenylephrine 2.5%) were administered approximately 60 minutes and again at approximately 45 minutes before the examination. Prior to the examination the infant was swaddled, and immediately prior topical local anaesthetic (proxymetacaine 0.5%) eye drops were instilled.

Physiological recordings

Heart rate, oxygen saturation and respiratory rate (calculated by the monitor) were downloaded at a sampling rate of 1 Hz; the ECG was recorded at a sampling rate of 250 Hz from 3 electrodes placed on the infant's chest, and the IP at a sampling rate of 62.5 Hz from

Adjei T, *et al. BMJ Open Resp Res* 2021; 8:e001042. doi: 10.1136/bmjresp-2021-001042

the chest electrodes. In Data set 2 the time of drug administration, heel lancing, and ROP screening, and in Data set 3 the time of ROP screening, were electronically marked on the physiological recordings by a researcher in real time.

Development of a new algorithm to identify breaths from the IP signal

The code for this algorithm is available from https://gitlab.com/paediatric_neuroimaging/identify_ibi_from_ip.git.

Removal of noise and artefacts

Cardiac-frequency noise is a particular problem affecting IP signals as the cardiac-synchronous signal can be erroneously detected as respiration, particularly during pauses in breathing (Supplementary Figure 1A, B). To remove this artefact, we used the approach of Lee and colleagues (5), whereby the timing of R-peaks is first identified from the ECG signal and from this the series of the time intervals between consecutive R-peaks, known as RR intervals, is computed and the R-peak interference is removed (see Lee et al. for further details). Empirically we found that the filter regime introduced by Lee et al. did not remove the R-peak interference from the IP signals in our data. This was primarily due to the poor performance of the Pan-Tompkins algorithm (6,7), which was used by Lee et al. to identify ECG R-peaks (5), on our IP data. Instead, we used signal envelopes to identify R-peaks, whereby the envelope of the ECG signal was generated using the MATLAB *envelope* function (which uses a spline interpolation of the signal maxima), in 5-second segments, enabling the identification of the R-peaks as the points of upper intersect between the envelope and the ECG signal. The RR intervals derived from the identified R-peaks were then stretched into 50 equally-sized steps, before resampling the IP signals in accordance with the RR intervals, as described above. Two infinite impulse response (IIR) notch filters were used to remove the 1 Hz and 2 Hz components (representing the R-peak interference) from the IP signal, before resampling the signal at 50 Hz. Following this, low frequency movement artefacts were removed using a high-pass finite impulse response (FIR) filter with a cut-off frequency of 0.5 Hz. Finally, large amplitude artefacts in the ECG recording could lead to large amplitude artefacts in the IP signal when it is filtered using the ECG signal. Extreme outliers in the filtered IP signals were identified through visual assessment as signal values with amplitudes greater than 6 times the 90^{th} percentile of the positive values of the filtered signal or less then 6 times the 10^{th} percentile of the negative values of the filtered signal; these were replaced with these boundary values. This occurred in 0.0009% of the total signal recorded across all infants. The differences between the computation employed by Lee et al. and that used in this study to remove cardiac-frequency noise are summarised in Supplementary Table 2.

The IP signal recorded on the monitor is hard-limited at upper and lower values. Gross movement artefact or periods of interference can lead to the signal being hard-limited for some period of time (example shown in Supplementary Figure 1C). To avoid these segments of hard-limited signal being erroneously detected as pauses in breathing in the filtered signal, segments of IP which corresponded to periods in which the original IP signal remained at a continuous maximum or minimum value for at least 1 second were removed. In addition, the 2.5-second segments either side of these segments were removed.

Gross artefacts which affect the ECG (and so also the IP signal) are detected by the patient monitor, which will not calculate a heart rate value during this period. Therefore, segments of IP which corresponded to instances of missing estimates of HR, and the 2.5-second segments of the IP either side of these episodes, were also removed from the IP signal.

Identification of breaths

The timing of individual breaths was identified from the filtered IP signal as the point at which the signal crossed an amplitude threshold. The threshold, θ, used to identify breaths was set as a proportion of the standard deviation of the IP signal across the previous N breaths. That is $\theta = \alpha\sigma(IP_N)$, where IP_N is the IP signal for the N previous breaths, and α is a scaling factor. We identified optimal values of α and N that achieved a balance between the number of false positives and false negatives in the identification of individual breaths and pauses in breathing (see Threshold optimisation). The first 600 seconds of the IP signal were analysed using a fixed threshold of $\theta = \alpha\sigma(IP|_0^{600})$, and the adaptive threshold was used for the rest of the recording (except for Data set 1 – see Threshold optimisation). Threshold crossings occurring within 0.3 seconds of the previous crossing were removed.

From this time series of breaths, the time intervals between consecutive breaths were computed giving the sequence of inter-breath-intervals (IBIs) for each infant. An IBI of 5 seconds or more was defined as a pause in breathing. Pauses in breathing which occurred within 2 seconds of a previous pause were combined and counted as a single pause. Also, as it is possible that a pause in breathing could begin during a period in which the IP data has been removed or is missing, IBIs of 5 seconds or more occurring immediately before or after a gap in IP data were included in the IBI series.

The time series of breaths was also used to compute the infants' respiratory rate shown in Figure 3. The respiratory rate was calculated from the number of breaths in a 20 second sliding window, moved through the IP signal with a 1-second increment; thus, a respiratory rate of 0 corresponds to an apnoea lasting 20 seconds or more.

Threshold optimisation

To determine the optimal threshold parameters for the identification of breaths in the infants' IP signals we assessed the algorithm's ability to detect individual breaths using Data set 1 for different levels of α and N. The manually annotated breaths were defined as true breaths, and false negatives defined as true breaths not identified by the algorithm. False positive breaths were defined as the breaths identified by the algorithm, but not annotated by clinical staff. Manual annotations were not necessarily annotated at a specific time within the breath wave and so the time of the manual annotations and the threshold crossings would not be expected to match exactly. However, the timings would be expected to be sequential i.e. there should only be one manual breath annotation between any two threshold crossings and vice versa. Therefore, the presence of two successive manual breath annotations without a threshold crossing between them indicated that the algorithm did not identify a breath (false negative), whilst two successive threshold crossings without a manual breath annotation between them indicated that the algorithm identified a false positive breath. The IP signal epochs from Data set 1 were each only approximately 180 seconds long (60 seconds before and approximately 80 seconds after a train of manual counts), so the threshold was initially applied to the first 40 seconds of data, before using an adaptive threshold. Initially, the

optimal value for α was identified as that which achieved a balance (i.e. similar number) between the false positive and false negative rates. In this assessment α was varied from 0.1 to 0.7 (with increments of 0.05) and a fixed value of N=15 was used. Having identified an optimal value for α, N was then varied to identify an optimal value for this parameter (which was identified as N=15).

To verify that the chosen threshold parameters were also able to accurately identify pauses in breathing we compared manual ratings of pauses in breathing with pauses in breathing identified by the algorithm. Pauses in breathing of greater than 5 seconds were identified in the first hour of the recordings in 15 of the infants from Data set 2 (of the 15 infants studied, 8 received morphine and 7 received placebo, though no drug had been given during the section of the recording assessed) by two investigators experienced in reviewing neonatal physiological data. The investigators identified pauses in breathing through visual assessment of the infants' original IP signals recorded from the monitor and the filtered IP signals (with cardiac-frequency noise and movement artefacts removed). Pauses in breathing where there were single oscillations (possibly breaths/gasps or possibly artefact from movement or stimulation) within a pause were identified as one pause. The investigators also reviewed the ECG, heart rate, and oxygen saturation traces for the infant, though the identification of pauses in breathing was primarily based on their visual inspection of the IP signals as short pauses in breathing do not necessarily affect the infant's heart rate or oxygen saturation.

For each threshold parameter, the pauses in breathing identified by the algorithm were compared to those identified by the two investigators; the pauses found by both investigators were defined as true pauses. The false negative rate was defined as the percentage of true pauses not identified by the algorithm. The false positive rate was defined as the percentage of pauses identified by the algorithm but identified by neither investigator.

Occasionally, due to technical difficulties, periods occurred where the ECG signal was not recorded but the IP signal was (0% in Data set 1, 0.3% of the total ECG record in Data set 2, 0.4% of the total record in Data set 3). Thus, during these periods, cardiac-frequency noise could not be removed from the IP signal. To assess the performance of the adaptive threshold without the removal of cardiac-frequency noise we also applied the algorithm to the same sections of the recording (which had ECG recorded throughout), for both the identification of breaths (Data set 1) and the identification of pauses in breathing (first hour of recordings in 15 infants from Data set 2) but did not filter out the cardiac noise.

Accurate identification of apnoeas using machine learning

Shallow breathing or poor electrode placement can lead to a low amplitude IP signal which is erroneously detected as a prolonged pause in breathing. We used machine learning to identify episodes of central apnoea versus periods erroneously detected as a prolonged pause in breathing. To develop the model, two investigators viewed the original IP signal, the filtered IP signal, the ECG, the heart rate and the oxygen saturation for 30 seconds before and after the start of all potential apnoeas identified in the entire recordings from all 30 infants in Data set 2. We defined here a central apnoea as a pause in breathing lasting 20 seconds or longer, so the investigators evaluated all periods for which no threshold crossings occurred for at least 20 seconds. The investigators determined whether they considered each potential apnoea to be a true apnoea or a false alarm (i.e. whether there was low amplitude signal for

another reason such as shallow breathing or poor electrode placement). The investigators viewed a total of 164 episodes of potential apnoea from the 30 infants and had an inter-rater agreement of 76%.

A support vector machine classification model was developed using the data from the 15 morphine-treated infants (training set), with all 69 episodes for which the investigators agreed, with 62% of these episodes classified as true apnoeas and the other 38% being classified as false alarms. Model performance was assessed using leave-one-subject-out cross-validation (i.e. all episodes for a subject were removed in a single fold). Having trained the model on the data from the morphine-treated infants, it was then tested in the data from the 15 placebo-treated infants (test set).

The model was constructed with 5 features:

1. The root-mean-square (RMS) of the IP signal during the episode (possible apnoea). The RMS was calculated in 1-second intervals from 1 second after the last annotated breath, until 19 seconds after the last annotated breath (i.e. up until the point immediately before the pause in breathing is classed as an apnoea according to our definition of at least 20 seconds in length), and the median of the RMS values across all intervals selected. The first and last second were removed as the IP signal can be affected by the edge of the last/first breaths as the time of first/last breath is selected as the point at which the threshold is crossed (Supplementary Figure 2A). The median value from the RMS values of the filtered IP signal was computed for each 1-second interval, as movement artefact in the IP signal could increase a sub-set of the RMS values even if a real apnoea was occurring (Supplementary Figure 2B).
2. The RMS value of the IP signal in the 10 seconds prior to the start of the episode.
3. The RMS value of the IP signal in the 10 seconds after the end of the episode.
4. The change in oxygen saturation calculated as the difference between the mean saturation in the 10 seconds prior to the start of the episode and the minimum value in the 60 seconds following the start of the episode.
5. The change in heart rate calculated as the difference between the heart rate in the 10 seconds prior to the start of the episode and the minimum value in the 60 seconds following the start of the episode.

Features 1-3 were included as it was noticed that true apnoeas often present with a marked decrease in the amplitude of the IP signal during the apnoea, which is not observed in false alarms (Figure 2A, B).

This model included input features computed before and after the apnoea to optimise it for retrospective apnoea identification. To determine whether this method could also potentially be used in prospective apnoea detection, we also adjusted the model to include only features up to 20 seconds from the start of the apnoea (including features 1 and 2, and 4 and 5 up to 20 seconds from the start as opposed to 60 seconds), see Supplementary Results.

Performance of apnoea identification

To compare the accuracy of our approach with the accuracy of the current standard, all periods where the monitor-derived respiratory rate reached 0 were viewed by two investigators and rated according to whether the investigator thought this period was a true central apnoea or a false alarm. The investigators viewed the original IP signal, the filtered IP

signal, the ECG, the heart rate and the oxygen saturation for 30 seconds before and after the start of all periods where the monitor-derived respiratory rate reached 0 from all 30 infants in Data set 2.

Supplementary Results

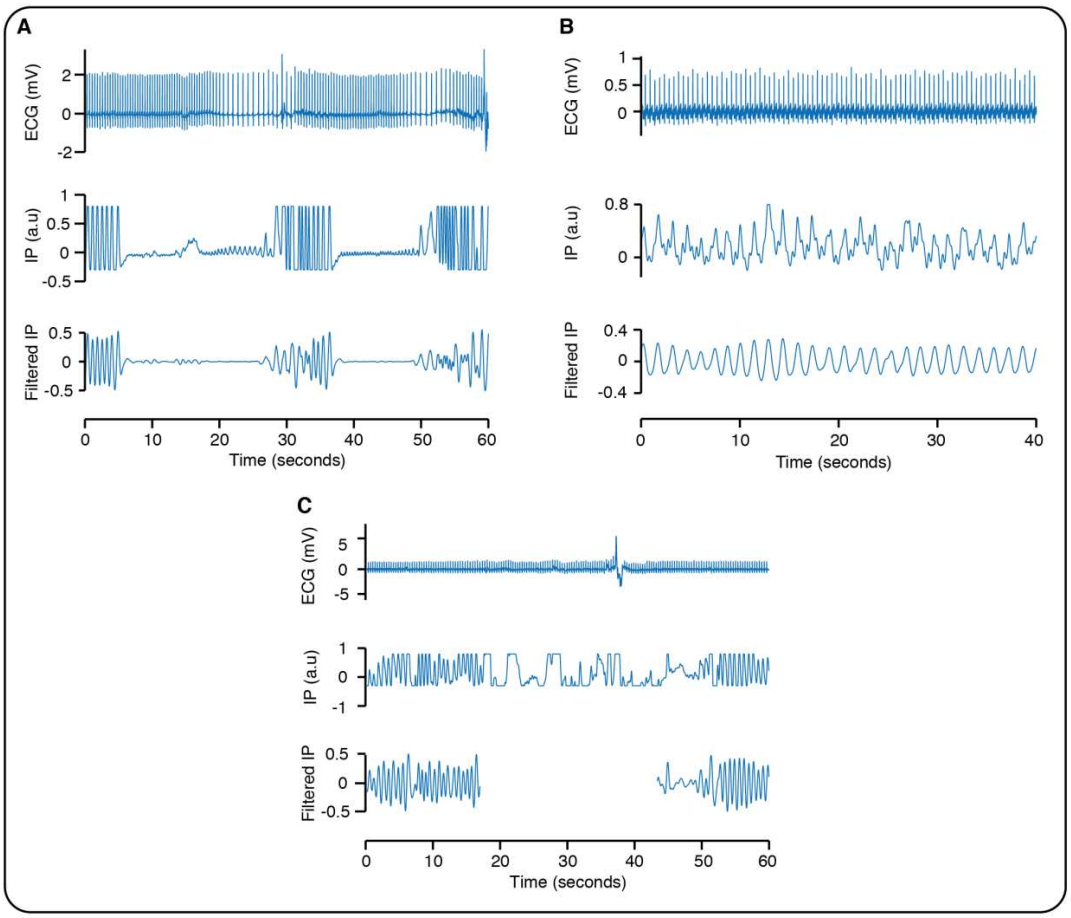
Optimising the adaptive threshold – without removal of ECG artefact

Due to technical problems, the IP signal is occasionally recorded without the ECG signal and so cardiac-frequency noise cannot be removed from the IP signal. We assessed the performance of the adaptive threshold at different levels without the removal of ECG artefact. With the removal of ECG artefact from the signal, the optimal threshold parameters were found to be α = 0.4 and N=15 (see the Results). With these parameters and without the removal of ECG artefact, 9% of the manually-annotated breaths in Data set 1 were missed by the algorithm, whilst 13% of the breaths identified by the algorithm were false positives. However, a threshold of 0.7 times the standard deviation (i.e. α = 0.7) achieved a better balance between false positive and false negative rates without the removal of ECG artefact, 13% in each case (Supplementary Figure 4A). In the assessment of pauses in breathing, at a threshold of 0.4 times the standard deviation of the IP signal computed using the previous 15 breaths, 31% of true pauses were missed by our algorithm but only 12% of pauses identified by our algorithm were false positives. However, a threshold of 0.5 times the standard deviation computed using the previous 15 breaths, without the removal of ECG artefact, achieved a better balance between the false negative and false positive rates, at 21% and 22% respectively (Supplementary Figure 4B). Values of α = 0.5 and N=15 were used in the subsequent analysis in periods where no ECG was present (0.3% of the total ECG record in Data set 2, 0.4% of the total record in Data set 3). Values of α = 0.4 and N=15 were used in the results when the ECG signal was present.

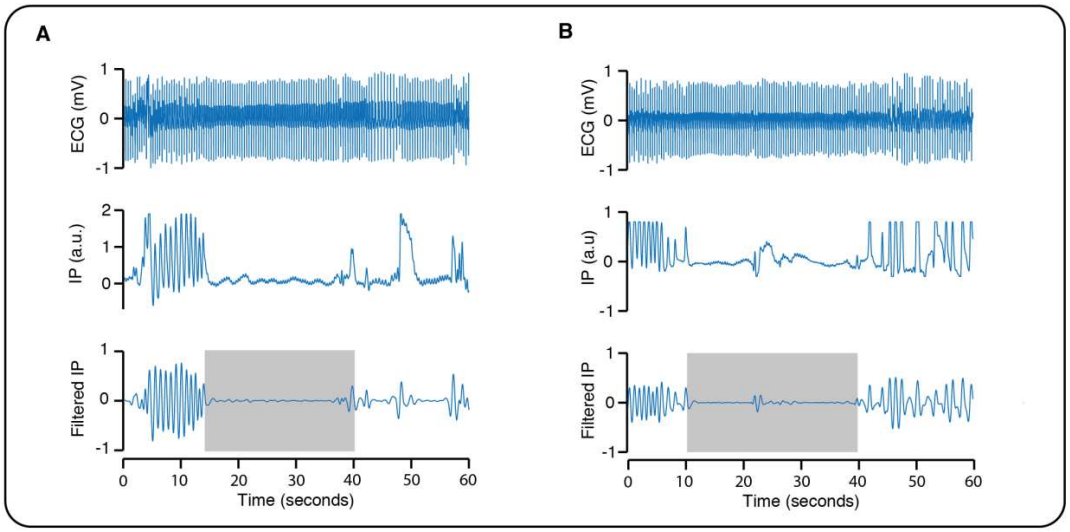
Accurate identification of apnoeas using machine learning

The model presented in the main text included input features computed before and after the apnoea to optimise it for retrospective apnoea identification. Adjusting the model to only include features up to 20 seconds after the start of the episode still achieved an accuracy of 74% (MCC=0.43), suggesting this model could be further developed to improve prospective apnoea identification.

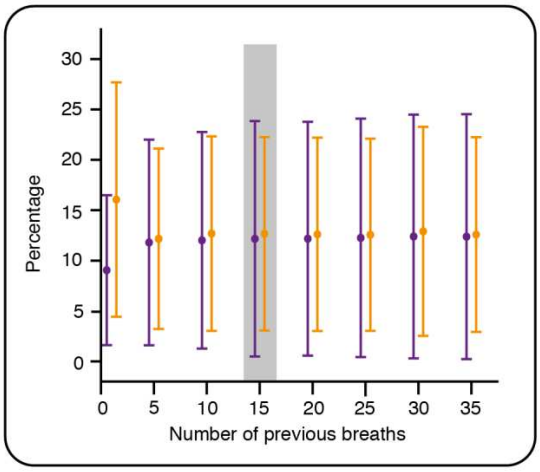
Supplementary Figures



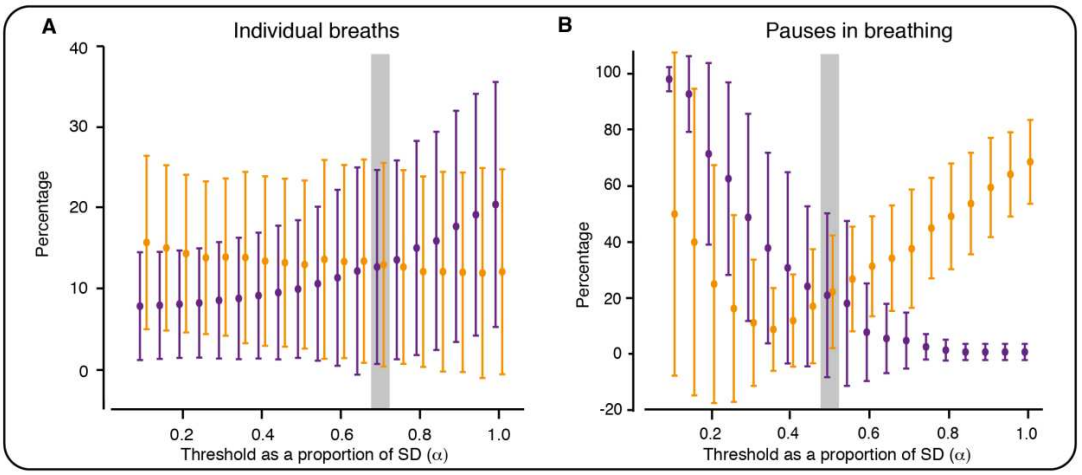
Supplementary Figure 1: Artefact removal. Examples of artefact removal from the IP (impedance pneumograph) signal. IP indicates the original IP signal recorded from the monitor. Filtered IP is the IP signal following removal of movement and cardiac-frequency artefacts. Time is given from the start of the epoch. (A) An example of cardiac-frequency noise which becomes particularly prominent during the period of apnoea – between approximately 15 and 25 seconds, and 40-50 seconds. (B) Ongoing cardiac-frequency noise during a period of regular breathing (the smaller amplitude oscillation on top of the higher amplitude respiratory oscillation). (C) An example of movement artefact. During periods of gross movement artefact or poor signal quality the IP signal hard limits at an upper or lower value, with the signal staying equal to this value for some period. Periods such as this (and 2.5 seconds either side) were removed from the signal for analysis so they were not incorrectly identified as pauses in breathing.



Supplementary Figure 2: Identification of true apnoeas using machine learning. Examples of apnoeas identified from the electrical impedance pneumograph (IP) signal. IP indicates the original IP recorded from the monitor. Filtered IP shows the IP signal following removal of cardiac-frequency noise and movement artefact. Grey shaded boxes indicate the periods identified as apnoeas using the filtered IP. Time is given from the start of the epoch. (A) An example showing clear 'edge' effects of the IP signal at the start and end of the period identified as an apnoea, as the period identified relates to when the signal crosses a threshold (i.e. at the edges of the apnoea the amplitude of the IP signal is higher than during the rest of the apnoea). For the machine learning algorithm the first second and last second of the signal during the apnoea are removed from the analysis to avoid these edge effects. (B) An example where a single oscillation (possibly movement artefact) occurs in the middle of the apnoea. For this reason, in the machine learning analysis, the median value of the filtered IP signal, computed from the RMS values for each 1-second interval during the apnoea, is used.



Supplementary Figure 3: Optimising the adaptive threshold according to the number N of previous breaths. Percentage of false positives (orange) and false negatives (purple) at different threshold values according to the number of previous breaths used to calculate the threshold. The threshold is set as 0.4 times the standard deviation of the signal from the previous N breaths. Error bars indicate mean and standard deviation. Values are jittered slightly on the x-axis so that false positive and false negative bars do not overlap. Grey shading indicates selected threshold parameters; at this threshold there was the optimal balance between the numbers of false positives and negatives.



Supplementary Figure 4: Optimising the adaptive threshold for the impedance pneumograph signal without removal of cardiac-frequency noise. Percentage of false positives (orange) and false negatives (purple) for different values of α (with N=15). (A) Values calculated in relation to identification of individual breaths and compared with breaths manually annotated at the time of the recording by visual observation of the infant (Data set 1). (B) Values calculated in relation to identification of pauses in breathing compared with pauses (of length at least 5 seconds) manually annotated by two investigators from viewing the signals (Data set 2, first hour of every recording). Unlike Figure 2 cardiac-frequency noise was not removed from the signals prior to the analysis shown here. In this case, slightly higher threshold parameters achieved the optimal balance between the percentages of false positives and false negatives (see Supplementary Results). Error bars indicate mean and standard deviation. Values are jittered slightly on the x-axis so that false positive and false negative bars do not overlap. Grey shading indicates optimal value of α without removal of cardiac-frequency noise in both cases – the best compromise between the numbers of false positives and negatives.

Supplementary Tables

Procedure	Data set used
Optimising the adaptive threshold (i) by comparison with manually annotated breaths	Data set 1
Optimising the adaptive threshold (ii) by comparison with pauses in breathing of at least 5 seconds	First hour of recording from 15 infants (8 allocated to receive morphine, 7 allocated to receive placebo) in Data set 2. The first hour of recording was chosen as this was not included in further analysis investigating morphine or ROP-induced changes in the IBI distribution. No drug had been given during this period.
Training of classification algorithm to identify true apnoeas versus false alarms	Morphine-treated infants in Data set 2 (n=15)
Test set for classification algorithm to identify true apnoeas versus false alarms	Placebo-treated infants in Data set 2 (n=15)
Testing the algorithm - evaluating respiratory depression following morphine administration	Morphine-treated infants in Data set 2 (n=15)
Using the algorithm - evaluating changes in IBIs following ROP screening	Placebo-treated infants in Data set 2 and Data set 3 (n=22)

Supplementary Table 1: Summary of where each data set was used.

Step	Current Study	Lee *et al.* (2012)
1	Derive RR intervals from the ECG through the application of envelopes to the ECG signal. The R-peaks are found as the points of intersect between the envelope and the ECG signal.	Derive RR intervals using the Pan-Tompkins R-peak detection algorithm (6,7).
2	Stretch or compress each RR interval such that each interval is made up of 50 steps.	Stretch or compress each RR interval such that each interval is made up of 30 steps.
3	Resample the IP in accordance with the modified RR interval series.	Resample the IP in accordance the modified RR interval series.
4	Apply a notch filter to remove the 1 Hz components from the IP, which now represent the R-peak artefact.	Apply Butterworth band-stop filters to remove R peak artefacts.
5	Apply a notch filter to remove the 2 Hz components from the IP, which represent the second harmonic of the R-peak artefact.	
6	Resample the IP to 50 Hz	Resample the IP to 60 Hz
7	High-pass filter the IP to remove <0.5 Hz noise.	High-pass filter the IP to remove <0.4 Hz noise

Supplementary Table 2: Filtering of the impedance pneumograph. Differences between the computations employed in the filtering of the IP signals in this study, and those employed by Lee et al. (5), primarily required due to the poor performance of the Pan-Tompkins R-peak detection in our data. The algorithm described by Lee and colleagues was constructed to remove cardiac interference from the IP signals and detect apnoeas. We use this algorithm to remove cardiac interference in our data. Additionally, we trained an SVM classifier to distinguish between true apnoeas and low amplitude artefacts and validated an adaptive threshold to identify individual breaths.

Gestational age at birth (weeks) – median (IQR)	28.1 (27.1 – 29.3)
Postmenstrual age at study (weeks) – median (IQR)	34.9 (34.3 – 36.1)
Weight at birth (g) – median (IQR)	1100 (889 – 1228)
Weight at study (g) – median (IQR)	2080 (1923 – 2206)
Sex – Female/Male – number (%)	12 (55) / 10 (45)
Normal vaginal delivery – number (%)	7 (32)
Assisted delivery – number (%)	1 (4)
Caesarean delivery – number (%)	14 (64)
Apgar score at 10 minutes – mean (SD)*	8.9 (1.5)
Infants ventilated during admission – number (%)	15 (68%)
Number of days ventilated in those ventilated – median (IQR)	2 (1.5-11.5)

Supplementary Table 3: Demographic details for the 22 infants where inter-breath interval distributions are compared before and after ROP screening. IQR – interquartile range, SD – standard deviation, * Apgar scores were missing from the medical records of 2 infants.

Supplementary References

1. Jorge J, Villarroel M, Chaichulee S, Green G, McCormick K, Tarassenko L. Assessment of signal processing methods for measuring the respiratory rate in the neonatal intensive care unit. IEEE J Biomed Heal Informatics. 2019;23(6):2335–46.
2. Hartley C, Moultrie F, Hoskin A, Green G, Monk V, Bell JL, et al. Analgesic efficacy and safety of morphine in the Procedural Pain in Premature Infants (Poppi) study: randomised placebo-controlled trial. Lancet. 2018;392(10164):2595–605.
3. Monk V, Moultrie F, Hartley C, Hoskin A, Green G, Bell JL, et al. Oral morphine analgesia for preventing pain during invasive procedures in non-ventilated premature infants in hospital: the Poppi RCT. Effic Mech Eval. 2019;6(9).
4. Slater R, Hartley C, Moultrie F, Adams E, Juszczak E, Rogers R, et al. A blinded randomised placebo-controlled trial investigating the efficacy of morphine analgesia for procedural pain in infants: Trial protocol [version 2; peer review: 3 approved]. Wellcome Open Res. 2017;1(7).
5. Lee H, Rusin CG, Lake DE, Clark MT, Guin L, Smoot TJ, et al. A new algorithm for detecting central apnea in neonates. Physiol Meas. 2012 Jan 1;33(1):1–17.
6. Pan J, Tompkins WJ. A Real-Time QRS Detection Algorithm. IEEE Trans Biomed Eng. 1985;BME-32(3):230–6.
7. Sedghamiz H. Complete Pan Tompkins Implementation ECG QRS detector [Internet]. 2014. Available from: https://uk.mathworks.com/matlabcentral/fileexchange/45840-complete-pan-tompkins-implementation-ecg-qrs-detector