

CLINICAL RESEARCH ARTICLE


Automated prediction of extubation success in extremely preterm infants: the APEX multicenter study

Lara J. Kanbar^{1,9}, Wissam Shalish^{2,9}, Charles C. Onu^{3,9}, Samantha Latremouille⁴, Lajos Kovacs⁵, Martin Keszler⁶, Sanjay Chawla⁷, Karen A. Brown⁸, Doina Precup³, Robert E. Kearney¹ and Guilherme M. Sant'Anna²✉

© The Author(s), under exclusive licence to the International Pediatric Research Foundation, Inc 2022

BACKGROUND: Extremely preterm infants are frequently subjected to mechanical ventilation. Current prediction tools of extubation success lacks accuracy.

METHODS: Multicenter study including infants with birth weight ≤ 1250 g undergoing their first extubation attempt. Clinical data and cardiorespiratory signals were acquired before extubation. Primary outcome was prediction of extubation success. Automated analysis of cardiorespiratory signals, development of clinical and cardiorespiratory features, and a 2-stage Clinical Decision-Balanced Random Forest classifier were used. A leave-one-out cross-validation was done. Performance was analyzed by ROC curves and determined by balanced accuracy. An exploratory analysis was performed for extubations before 7 days of age.

RESULTS: A total of 241 infants were included and 44 failed (18%) extubation. The classifier had a balanced accuracy of 73% (sensitivity 70% [95% CI: 63%, 76%], specificity 75% [95% CI: 62%, 88%]). As an additional clinical-decision tool, the classifier would have led to an increase in extubation success from 82% to 93% but misclassified 60 infants who would have been successfully extubated. In infants extubated before 7 days of age, the classifier identified 16/18 failures (specificity 89%) and 73/105 infants with success (sensitivity 70%).

CONCLUSIONS: Machine learning algorithms may improve a balanced prediction of extubation outcomes, but further refinement and validation is required.

Pediatric Research (2023) 93:1041–1049; <https://doi.org/10.1038/s41390-022-02210-9>

IMPACT:

- A machine learning–derived predictive model combining clinical data with automated analyses of individual cardiorespiratory signals may improve the prediction of successful extubation and identify infants at higher risk of failure with a good balanced accuracy.
- Such multidisciplinary approach including medicine, biomedical engineering and computer science is a step forward as current tools investigated to predict extubation outcomes lack sufficient balanced accuracy to justify their use in future trials or clinical practice.
- Thus, this individualized assessment can optimize patient selection for future trials of extubation readiness by decreasing exposure of low-risk infants to interventions and maximize the benefits of those at high risk.

INTRODUCTION

Despite significant technological advances, mechanical ventilation (MV) remains associated with lung injury and important short- and long-term morbidities in extremely preterm infants.^{1–4} Therefore, timely weaning and disconnection from MV is critical. Unfortunately, due to the inability to accurately judge extubation readiness in this fragile population, newborn infants face the highest rates of extubation failure of all intensive care settings.^{5–8}

Currently, all tools investigated to predict extubation success lack sufficient balanced accuracy to justify their use in clinical practice.⁹ Given that these infants can develop significant morbidities and rarely experience severe adverse events after extubation,¹⁰ an enhanced decision-making process is highly desirable. Recent collaborations between medicine, biomedical engineering and computer science have proven to be paramount in health care.^{11,12} Indeed, the use of automated analyses of biological signals and artificial intelligence has allowed complex

¹Department of Biomedical Engineering, McGill University, Montreal, QC, Canada. ²Department of Pediatrics, Neonatology, McGill University Health Center, Montreal, QC, Canada. ³School of Computer Science, McGill University, Montreal, QC, Canada. ⁴Department of Experimental Medicine, McGill University, Montreal, QC, Canada. ⁵Department of Neonatology, Jewish General Hospital, McGill University, Montreal, QC, Canada. ⁶Department of Pediatrics, Women and Infants Hospital of Rhode Island, Brown University, Providence, RI, USA. ⁷Division of Neonatal-Perinatal Medicine, Hutzel Women's Hospital, Children's Hospital of Michigan, Central Michigan University, Pleasant, MI, USA. ⁸Department of Anesthesia, McGill University Health Center, Montreal, QC, Canada. ⁹These authors contributed equally: Lara J. Kanbar, Wissam Shalish, Charles C. Onu. ✉email: guilherme.santanna@mcgill.ca

Received: 28 September 2021 Revised: 29 June 2022 Accepted: 8 July 2022
Published online: 29 July 2022

investigations of physiologic variables characterized by a highly elaborate, apparently random output, arising from nonlinear biological systems.^{13,14}

These advanced signal analyses have also been used as prediction tools of clinical outcomes, including ventilation weaning and extubation readiness.^{15,16} Moreover, in preterm infants analyses of some biological signals have also improved the understanding of MV weaning and post-extubation care.^{17–20} Thus, the objective of this multicenter study was to develop and evaluate a prediction model with balanced accuracy for extubation success in extremely preterm infants using machine learning algorithms that combine clinical data with automated analyses of cardiorespiratory signals.

METHODS

This prospective multicenter study was performed from September 2013 to August 2018 and is presented using the TRIPOD statement. The study protocol was registered in Clinicaltrials.gov (NCT01909947) and published.²¹ Ethics approval was attained from each Institutional Review Board and written informed consent was obtained from parents prior to enrollment.

Participating centers

Neonatal Intensive Care Units of the Royal Victoria Hospital, Montreal Children's Hospital, Jewish General Hospital (Montreal, QC, Canada), Detroit Medical Center (Detroit, MI, USA); and Women and Infant's Hospital (Providence, RI, USA).

Eligibility criteria and clinical management

Infants with birth weight (BW) \leq 1250 g receiving MV and undergoing their first extubation attempt were eligible. Exclusion criteria were unplanned extubation, major congenital anomalies, major heart defects, cardiac arrhythmia, use of vasopressor or sedative drugs at the time of extubation, and patients extubated from high frequency ventilation (due to interference by the oscillations on the cardiorespiratory signals) or directly to oxyhood or low-flow nasal cannula. Guidelines to consider a patient 'ready' for extubation were proposed²¹ but all decisions concerning extubation and reintubation were made by the medical team. Caffeine administration was part of standard of care at all participating centers.

Outcome

The original, per-protocol outcome for the prediction model was extubation success, defined by the absence of specific criteria including oxygen needs, blood gases, and apneas in the 72 h post-extubation.²¹ However, during the conduct of the study it became evident that the data were inconsistently recorded within and across centers. Therefore, definition of extubation success was changed to the absence of reintubation within 72 h post-extubation. In addition, data on the occurrence, timing and reasons for reintubation were recorded. Blinding of health care providers to this outcome was not possible.

Predictors

Candidate predictors for the development of the classifier included 109 clinical parameters pertaining to patient demographics and pre-extubation characteristics.²¹ Cardiorespiratory signals included: (a) Ribcage (RCG) movements, using uncalibrated Respiratory Inductance Plethysmography (RIP; Viasys® Healthcare); (b) Abdominal (ABD) movements, using RIP; (c) Electrocardiogram (ECG), using three electrodes (Vermed®); (d) Photothymogram (PPG) and oxygen saturation (SAT), using a pulse oximeter monitor (Masimo Radical®). Details on data acquisition set up have been previously published.^{10,21} Briefly, all cardiorespiratory signals were continuously acquired before extubation, for a period of 60 min while receiving MV at pre-extubation settings, followed by 5 min during endotracheal tube continuous positive airway pressure (ET-CPAP) to capture the spontaneous breaths without interference from mechanical inflations. Clinical instability during the recording period was not considered, as this was not a spontaneous breathing trial. Thus, all involved patients were extubated thereafter and remained eligible for inclusion in the development of the prediction model. All signals were anti-alias filtered at 500 Hz and sampled at 1000 Hz using the PowerLab 16/30 analog-digital data acquisition system (ADInstruments, Australia) with 16-

bit analog-to-digital resolution. All health care providers were blinded to the cardiorespiratory signals during the study. Clinical and cardiorespiratory data were stored in a cloud-based repository and an automated anonymization protocol was developed.²² Moreover, an algorithm for quality control and validation was systematically applied.²³

Statistical analysis methods

Stages 1 and 2 of the analyses were conducted using MATLAB (R2018a, The MathWorks) and stage 3 with the Python Scikit-Learn v0.19[©] (scikit-learn developers - BSD License). A simplified flow diagram for the development of the classifier is outlined in Fig. 1.

Stage 1. Characterization of clinical data and cardiorespiratory signals

For clinical data, continuous variables were presented as median [interquartile range] and categorical variables as n (percentage). Comparisons between infants with successful and failed extubation were performed using Wilcoxon rank sum test, Chi square test or Fisher exact test, as appropriate. For cardiorespiratory data, the signals were first processed to compute sample-by-sample metrics describing power, respiratory and cardiac frequency, and thoraco-abdominal synchrony between RCG and ABD.²⁴ Next, the RCG and ABD metrics were analyzed using AUREA, an Automated Unsupervised Respiratory Event Analysis algorithm originally developed in older infants²⁵ and revised for the extremely preterm infant population (r-AUREA, Fig. E1). r-AUREA assigned each sample to one of the following unique respiratory patterns: Pause (PAU), Movement Artifact (MVT), Synchronous Breathing (SYB), Asynchronous Breathing (ASB), and Unknown (UNK). Furthermore, the sequence of pattern classification was revised and the probability of switching from one pattern to another was quantified using a Semi-Markov model.²⁶ Cardiac frequency was computed from either the ECG or PPG and algorithms were designed to detect bradycardia (heart rate $<$ 100 bpm). PPG was computed as a backup strategy in case of loss of the ECG signal. Desaturation events (oxygen saturation $<$ 80%) were computed from the PPG. Lastly, the ECG was processed to derive inter-beat intervals and compute measures of heart rate variability (HRV). A detailed description of this methodology is provided (Methods E1).

Stage 2. Development of the clinical and cardiorespiratory feature set

Summary statistics describing the properties of the metrics, patterns, and inter-beat intervals during MV and ET-CPAP were computed to create a set of features. The median [inter-quartile range] of each metric and its power were calculated. For patterns, measures of their frequency of occurrence, duration, and the transition probability from one pattern to another were included. HRV features included time-domain, frequency-domain, and non-linear analyses. This resulted in a total of 224 cardiorespiratory and 109 clinical features. To reduce the size of the set, a principal component analysis (PCA) was used to transform the original features into a set of linearly uncorrelated Principal Component (PC) features that concisely explained the variance in the original set. In other words, PCA reduces the size of the feature set by removing unessential or highly correlated features while preserving as much information as possible. The contribution of each feature to the clinical (Fig. E2) and cardiorespiratory principal component were evaluated by heat maps (Fig. E3) and the percentage of total variance explained by every PC was assessed (Fig. E4).²⁷ PC features that explained $<$ 1% of the total variance were excluded resulting in a set of 95 PC features that explained 95% of the variance. A detailed description of this methodology is provided (Methods E2).

Stage 3. Classification

The overall population was imbalanced since less than 20% of infants failed extubation. Therefore, a two-stage Clinical Decision - Balanced Random Forest (CD-BRF) classifier was used.^{28–31} In the CD stage, the best-performing clinical parameters were the gestational age (GA) at birth and weight at extubation and the optimal boundary was selected to include all failure infants and the fewest number of infants with success (Fig. E5). Infants outside of this boundary were automatically classified as success, while those in the high-risk group were passed to the BRF stage, which used clinical and cardiorespiratory PC features to complete the classification. In simple terms, a random forest classifier averages the predictions made by many decision trees (which individually may not be as accurate) into a single more robust estimate. Each decision tree uses a random

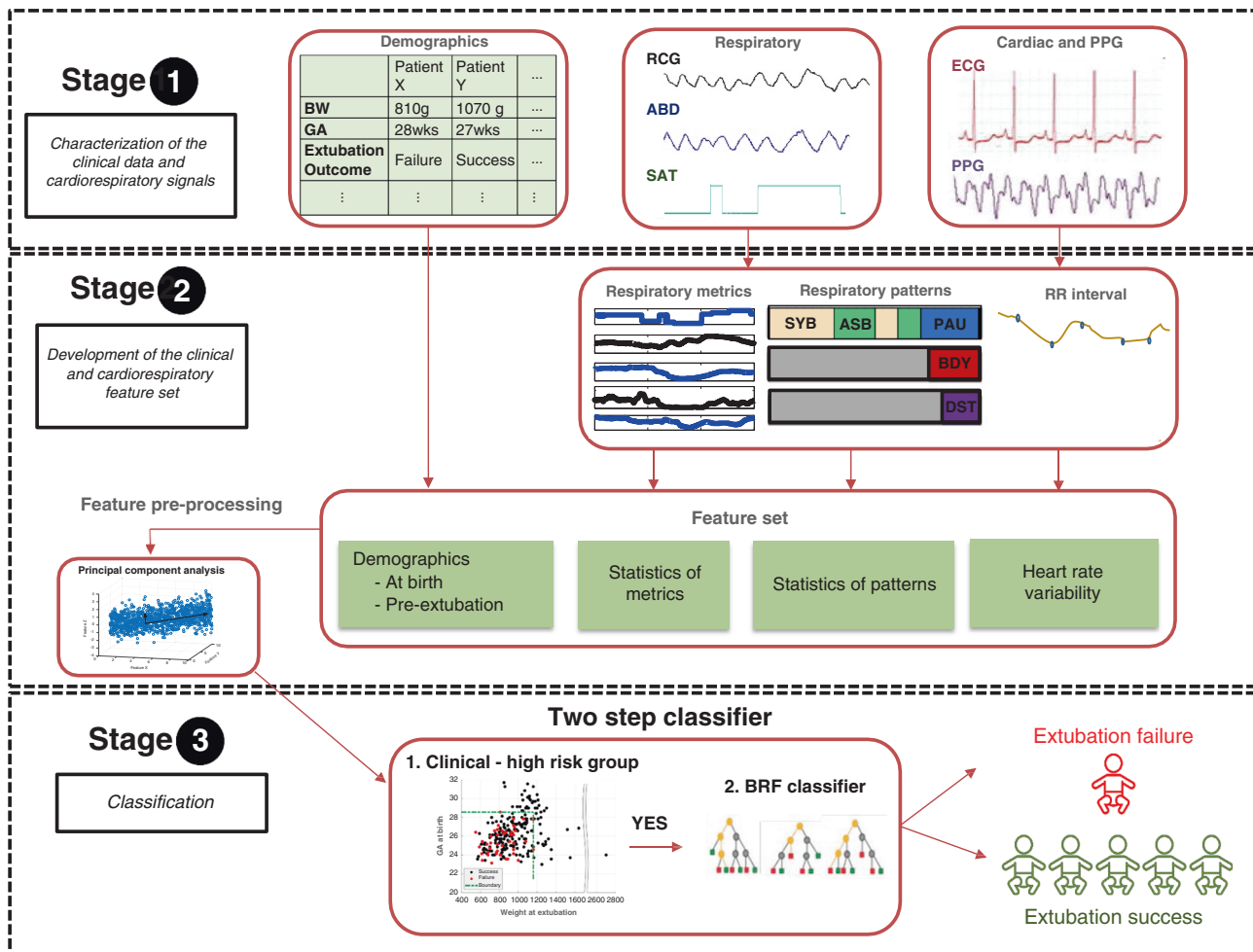


Fig. 1 Flow diagram for the development of the classifier of extubation readiness. The classifier was developed in 3 stages. Stage 1 - Clinical parameters were collected, and cardiorespiratory signals acquired; Stage 2 - Cardiorespiratory signals were processed into metrics describing power, frequency, and thoraco-abdominal synchrony. Using the metrics, a revised automated unsupervised respiratory event analysis algorithm (*r*-AUREA) categorized signals into respiratory patterns (pause [PAU], synchronous [SYB] or asynchronous [ASB] breathing, and movement artifact [MVT]) and patterns of desaturation [DST] and bradycardia [BDY]. Statistical representation of metrics, patterns and RR intervals led to a feature set characterizing the cohort's cardiorespiratory behavior. Next, a Principal Component Analysis (PCA) was applied to transform the original features into PC features. Stage 3 - Machine-learning methods used the PC features to classify infants into extubation success or failure. First, the high-risk population was selected by using an automatically derived threshold for gestational age at birth and weight at extubation that minimized successful extubation while capturing all failures. Then, the most discriminatory combination of clinical and cardiorespiratory features (selected with five-fold cross validation) was used to train and test a balanced random forest classifier using leave-one-out procedure for validation. BW = birth weight, GA = gestational age, RC = ribcage, ABD = abdominal movements, SAT = oxygen saturation, ECG = electrocardiogram, PPG = pulse plethysmography.

subset of features to make its prediction, and the overall random forest has a fixed set of parameters, called hyperparameters, which determine the properties of the model. These BRF parameters included the number of trees in the forest, maximum depth of the trees, minimum number of patients to split a node and maximum number of features at each node. (Methods E3 and Fig. E6). Three BRFs classifiers were designed: Clinical (C) classifier using clinical PC features alone, a Cardiorespiratory (CR) classifier using cardiorespiratory PC features alone, and a Clinical and Cardiorespiratory (CCR) classifier using both. Due to the relatively limited number of failure patients, a leave-one-out cross-validation was used to evaluate the performance of the classifiers. A detailed description of this methodology is provided (Methods E3).

For the performance of the classifiers (C, CR and CCR), receiver operating characteristics (ROC) curves were generated, and the following measures computed: sensitivity (detection rate of successful extubation), specificity (detection rate of failed extubation), positive predictive value (PPV), negative predictive value (NPV), and area under the curve (AUC). The best performance was determined by the balanced accuracy (sensitivity and specificity/2). Lastly, the relative *importance* of each PC feature selected by the best classifier was estimated to quantify the most discriminative role.

Missing features

Some clinical data were not collected, or cardiorespiratory signals were not acquired (disconnection of leads or too short signal acquisition periods). If a feature was missing ≤ 10 infants, the missing values were imputed from the median value of the outcome group to which the infant belonged. If a feature was missing in > 10 infants, the feature was excluded, which occurred with 29 clinical and 25 cardiorespiratory features.

Sample size

Sample size was estimated using the determination method published by Obuchowski and McClish.³² This calculation requires estimation of the prevalence of reintubation in the target population and variance of the ROC curve based on a pilot study¹⁷, and a determined precision for the Area Under the Curve (AUC).²¹ Using a conservative extubation failure rate of 20%, we anticipated that a minimum of 170 infants would be necessary to estimate the AUC of the ROC curve with a precision of 0.1. Accounting for low quality signals and practice changes over time, the sample size was increased to 250 patients. A prospective validation on a sample of 50 infants was initially planned but a *posteriori* decision was made to

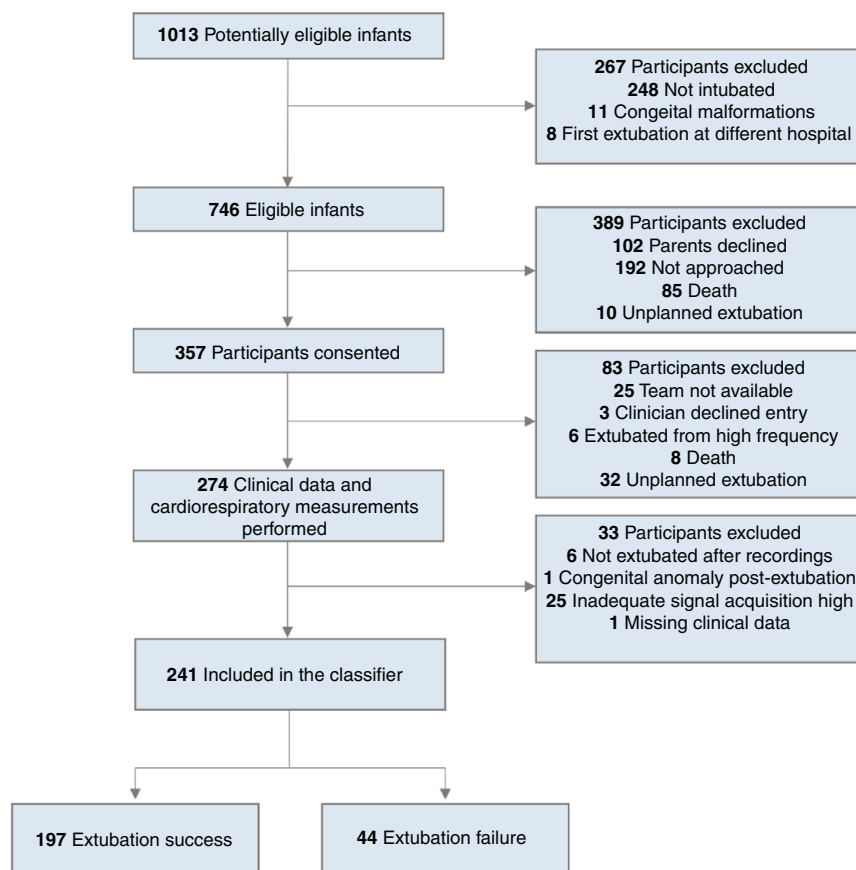


Fig. 2 Patient Flowchart.

proceed with leave-one-out cross-validation because a sample of 50 infants was likely to be insufficient given the relatively low rate of extubation failure.

Risk groups

Considering this was a pragmatic study, included infants were extubated at variable postnatal ages. However, it is plausible that predictors of extubation success may vary between infants extubated early (<7d of age) or late (≥ 7 d of age) as prolonged MV has been associated with lung injury². Also, these infants are more prone to complications such as IVH or PH during the first week of life, where an extubation success or failure might have a different impact. Consequently, a post-hoc analysis was carried out to evaluate the accuracy of the predictor in these groups. Furthermore, the classifier accuracy was also tested for re-intubations occurring ≤ 7 days (168 h) following extubation as longer observation windows can capture more cases of respiratory related extubation failures.^{33,34}

RESULTS

Participants

A total of 266 infants were enrolled and 241 included for the development of the classifier (Fig. 2). The median time between the ET-CPAP recording and extubation was 31 min [IQR 20–55 min] and all involved patients were extubated thereafter and remained eligible for inclusion in the development of the prediction model. Extubation success occurred in 197 infants (82%). Table 1 summarizes pre- and post-extubation characteristics of the population, including timing and reasons for reintubation in the 44 infants who failed extubation.

Model development and specification

In the CD stage, a threshold of GA at birth >28.6 weeks and weight at extubation >1160 g (Fig. E5) automatically classified 52 infants

(22%) as successful extubations (no failures). The remaining 189 infants (78%) passed to the BRF stage. In this group, 44 (23%) failed extubation. For this stage, the optimal set of hyperparameters selected for the C, CR, and CCR classifiers are presented on Table E1.

Model performance

A total of 241 leave-one-out cross-validation tests were performed using 42 PC features selected across all tests: 11 clinical and 31 cardiorespiratory (Fig. E7). Nine of the ten PC features with greatest contribution to the classifier were cardiorespiratory. ROC curves and diagnostic performances for the three classifiers are shown in Fig. 3 and data presented on Table 2. The best performance was obtained using the CCR classifier, with a total of 18 PC features (Fig. E8). The highest three features for each cardiorespiratory PC used in the final CCR classifier are presented on Table E3. These top features spanned a wide range of categories, including measures of metrics (pertaining to respiratory and cardiac frequencies, power in the RCG and ABD signals, phase angle, and oxygen saturation), respiratory patterns (particularly the UNK and bradycardia patterns), and heart rate variability.

The diagnostic and clinical values of the best classifier are presented on Fig. 4. From the diagnostic standpoint, 137/197 infants with successful and 33/44 infants with failed extubation were correctly identified by the classifier (70% sensitivity and 75% specificity, respectively) with a balanced accuracy of 73%. Clinically, when used as an adjunct tool, the classifier agreed with the decision to extubate in 148 infants (61%). Of those infants, 137 were correctly classified as extubation success (93%). However, the classifier predicted that 93 infants (39%) would fail. Of those, 60 were successfully extubated.

Table 1. Population demographics.

Variables	Total (n = 241)	Extubation Success (n = 197)	Extubation Failure (n=44)
<i>At birth</i>			
Gestational age, weeks	26.1 [24.9–27.4]	26.3 [25.0–27.8]	25.4 [24.3–26.5]*
Birth weight, g	830 [708–1016]	870 [710–1050]	725 [610–830]*
Male	127 (53)	106 (54)	21 (47)
Antenatal steroids	217 (90)	179 (91)	38 (86)
Apgar at 5 min	7 [5–8]	7 [5–8], n=196	7 [5–8], n=43
Surfactant	229 (95)	185 (94)	44 (100)
Caffeine	235 (98)	192 (97)	43 (98)
<i>At extubation</i>			
Corrected gestational age	28 [26.9–29.4]	28.4 [27–29.7]	27.4 [26.6–28.5]*
Weight	940 [810–1080]	960 [830–1100]	820 [715–950]*
Day of life	8 [3–25]	7 [3–25]	14 [4–26]
<i>Ventilator mode</i>			
Patient-triggered†	130 (54)	103 (52)	27 (61)
Volume guarantee	94 (39)	71 (36)	23 (52)
Ventilator rate, inflations/min	22 [20–30]	22 [20–30], n=180	23 [20–31], n=41
Tidal volume, ml/kg	4.9 [4.2–5.3]	4.9 [4.4–5.3], n=71	4.7 [4.1–5.3], n=23
MAP, cm H ₂ O	7.1 [6.3–8]	7 [6.3–7.9], n=196	8 [6.5–9.3], n=44*
PEEP, cm H ₂ O	5 [5–6]	5 [5–6]	6 [5–7]‡
PIP, cm H ₂ O	14 [12–15]	13 [12–15], n=110	15 [14–17], n=19‡
FiO ₂	0.22 [0.21–0.27]	0.21 [0.21–0.26]	0.26 [0.22–0.32]*
pH	7.34 [7.29–7.38]	7.34 [7.29–7.38], n=161	7.32 [7.28–7.37], n=38
pCO ₂ , mm Hg	44 [38–51]	44 [37–47], n=161	47 [40–56], n=38
<i>Post-extubation</i>			
<i>Respiratory support post-extubation</i>			
CPAP	145 (60)	119 (60)	26 (59)
CPAP level, cm H ₂ O	6 [5–7]	6 [5–6]	7 [6–8]*
FiO ₂	0.23 [0.21–0.3]	0.21 [0.21–0.27]	0.3 [0.25–0.4]*
NIPPV	85 (35)	67 (34)	18 (41)
PIP, cm H ₂ O	15 [14–19]	15 [14–18]	17 [15–20]‡
PEEP, cm H ₂ O	6 [5–7]	5 [5–7]	7 [6–8]
Rate, inflations/min	30 [29–40]	30 [26–40]	33 [30–40]
FiO ₂	0.28 [0.25–0.35]	0.27 [0.23–0.35]	0.35 [0.27–0.45]‡
HFNC	11 (5)	11 (6)	0
<i>Reintubation</i>			
≤ 72 h	44 (18)	0	44 (100)
≤ 168 h (7 days)	66 (27)	22 (11)	—
≤ 336 h (14 days)	83 (34)	39 (20)	—
Any time during hospitalization	108 (45)	64 (32)	44 (100)*
<i>Respiratory support pre-reintubation</i>			
CPAP	n=44	—	—
CPAP level, cm H ₂ O	6 (13)	—	—
CPAP level, cm H ₂ O	8 [6–9]	—	—
FiO ₂	0.45 [0.31–1]	—	—
NIPPV	38 (86)	—	—
PIP, cm H ₂ O	18 [15–22]	—	—
PEEP, cm H ₂ O	7 [6–9]	—	—
Rate, inflations/min	40 [35–40]	—	—
FiO ₂	0.46 [0.32–0.6]	—	—
<i>Reasons for reintubation within 72h</i>			
Increased work of breathing	n=44	—	—
Increased work of breathing	13 (30)	—	—
Apneas and bradycardias	24 (55)	—	—

Table 1. continued

Variables	Total (n = 241)	Extubation Success (n = 197)	Extubation Failure (n=44)
Increased oxygen requirements	6 (14)		

Values are expressed as *n* (%), or median [interquartile range], *n* = number of patients. MAP = mean airway pressure; PEEP = positive end expiratory pressure; PIP = peak inflation pressure; FiO₂ = fraction of inspired oxygen; CPAP = continuous positive airway pressure; NIPPV = nasal intermittent positive pressure ventilation; HFNC = high flow nasal cannula.

**p* value < 0.01, †mode of ventilation where the newborn respiratory effort is used to trigger the ventilator promoting patient-ventilator synchrony, ‡*p* value < 0.05.

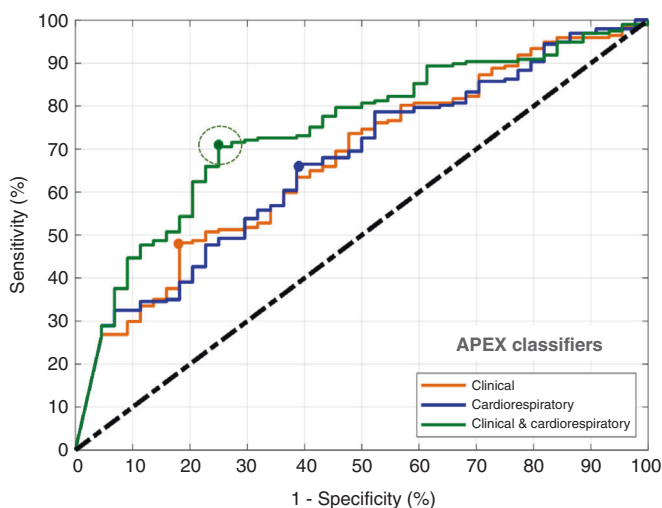


Fig. 3 Performance of the APEX classifier. The APEX classifiers included all infants extubated at any age during the study. The Receiver Operating Characteristic (ROC) curves shows the trade-off between sensitivity (Y axis) and 1-specificity (X axis) for the Clinical (orange), Cardiorespiratory (blue) and Clinical and Cardiorespiratory (green) classifiers at all thresholds. For each curve, the filled dot marker indicates the performance with the highest balanced accuracy. The black 45-degree diagonal dotted line represents the baseline model/random classifier. The closer an ROC curve comes to this diagonal line, the less powerful is the model. Note that the best accuracy was obtained with the combination of clinical and cardiorespiratory signals (green circled dot).

Post-hoc analysis

The performance of the APEX CCR classifier was computed for the subgroup of infants extubated at <7d (*n* = 123) or ≥ 7d of age (*n* = 118), as shown in Table E2. In the group <7d, the classifier correctly identified 16/18 failures (Specificity 89%, 95% CI: 74%, 100%) compared to 17/26 in the late extubation group (Specificity 65%, 95% CI: 47%, 84%). The diagnostic and clinical values of this analysis are presented on Fig. E9. The performance of the classifier decreased when reintubation ≤7d following extubation was used as definition of failure (Fig. E10). Indeed, only 11 (50%) infants reintubated between 72 h and 168 h were correctly classified as failures.

DISCUSSION

The use of machine learning algorithms to combine automated analyses of cardiorespiratory signals with clinical data improved prediction of extubation success in a high-risk population. More importantly, by using a two-stage CD-BRF approach the final classifier excluded a group of infants successfully extubated and was developed using extremely preterm infants with higher risk of extubation failure (Fig. E5). This precision medicine approach can help in the selection of a targeted population in future studies of extubation failure. Indeed, only 2 infants classified as extubation success failed. Unfortunately, the final classifier's improved

identification of extubation failures at the expense of misclassifying nearly one-third of infants who were successfully extubated making it not suitable for clinical decisions at this point. Notably, the classifier performed best amongst infants extubated before 7 days of age, identifying 70% of success and 89% of failures. Although a post-hoc analysis, this is an important finding as early and successful extubation may mitigate complications associated with prolonged MV while limiting the adverse effects caused by reintubation during a critical age.^{5,10}

The decision to extubate extremely preterm infants is complex and subjected to substantial variability.^{35,36} Furthermore, extubation failure inevitably prolongs MV duration and has been associated with increased risks of respiratory morbidities and mortality, even after adjusting for the cumulative MV duration and other known confounders.^{7,8,10} Thus, accurate predictors of extubation success are needed but only a few have been evaluated in preterm infants.^{5,9,37,38} Most studies included limited number of patients from a single center and incorporated a particular clinical or physiological parameter. As a result, the accuracy in detecting extubation failures was low when compared to clinical judgment.¹³ In the current study, the combination of clinical parameters with automated analysis of cardiorespiratory signals generated the highest performance classifier, thereby demonstrating the added value in acquiring those signals.¹⁷

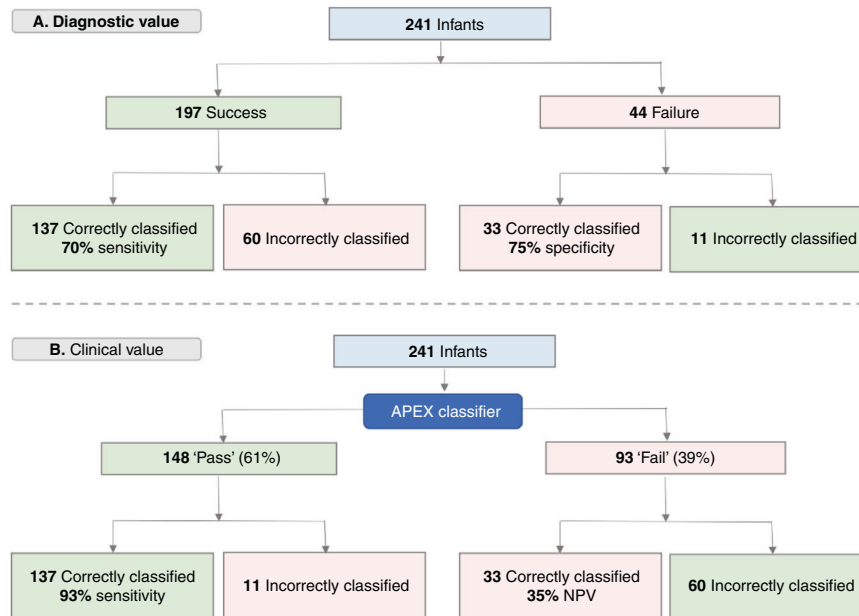
The APEX CCR classifier performed with a balanced accuracy of 73% (sensitivity of 70% and specificity of 75%). Thus, if APEX were used alone in clinical practice, it would correctly identify 3/4 infants that would go on to succeed extubation while misclassify as failure about 2 infants that could have been otherwise successfully extubated. Whether this trade-off is acceptable remains unknown, as it would require direct comparisons of the costs of postponing extubation vs. preventing reintubation. Surprisingly, only one study has investigated the impact of delaying extubation by 36 h in preterm infants.³⁹ No differences in rates of extubation failure or bronchopulmonary dysplasia were noted, but amongst infants <1000 g delayed extubation was associated with significantly shorter cumulative MV duration. Therefore, it is unknown if a short delay can be harmful. Furthermore, the experience of adverse and sometimes severe events after extubation must be considered. Indeed, trade-offs in which significant morbidity reductions are offset by a marginal increase in mortality have often shaped recommendations in Neonatology.⁴⁰⁻⁴² In the APEX cohort, 3 infants died within hours following an electively planned extubation. The causes of death were massive pulmonary hemorrhage minutes after extubation, withdrawal of life-sustaining therapy after a diagnosis of grade 4 intraventricular hemorrhage made 7 h post-extubation, and fulminant necrotizing enterocolitis diagnosed ≤8 h post-extubation. All these patients were correctly identified by the classifier as extubation failures.

More importantly is the use of the classifier as an adjunct tool to clinical judgment. In this study, when a patient 'ready' for extubation was classified as success, the probability of successful extubation increased from 82% to 93%. On the other hand, this probability decreased from 82% to 65% if the patient was classified as failure. In infants extubated <7 days of age, the classifier identified 89% of failures (95% CI: 74%, 100%) and would

Table 2. Performance of the Clinical, Cardiorespiratory and combined Clinical and Cardiorespiratory classifiers.

	Clinical classifier	Cardiorespiratory classifier	Clinical and Cardiorespiratory classifier
Sensitivity	48% (95% CI: 41%, 55%)	66% (95% CI: 59%, 73%)	70% (95% CI: 63%, 76%)
Specificity	82% (95% CI: 70%, 93%)	61% (95% CI: 47%, 76%)	75% (95% CI: 62%, 88%)
PPV	92% (95% CI: 87%, 97%)	88% (95% CI: 83%, 94%)	93% (95% CI: 88%, 97%)
NPV	26% (95% CI: 19%, 33%)	29% (95% CI: 20%, 38%)	35% (95% CI: 26%, 45%)
Balanced accuracy	65%	65%	73%
AUC	0.67	0.67	0.75

Sensitivity, detection rate of successful extubation; Specificity, detection rate of failed extubation, PPV, positive predictive value; NPV, negative predictive value; Balanced accuracy, the average of sensitivity and specificity) and AUC, the area under the curve for each classifier; CI, coefficient interval.

**Fig. 4** Diagnostic and clinical values of the clinical and cardiorespiratory classifier in extremely preterm infants.

have increased the probability of successful extubation from 85% to 97% (95% CI: 94%, 100%; Fig. E10).

Interestingly, the performance of the APEX CCR classifier varied considerably depending on the age at extubation and observation window used to define extubation failure. With regards to age at extubation, the classifier accurately identified 16 out of 18 (89%) failed extubations in the first 7 days of life compared to only 17 out of 26 (65%) failed extubations beyond the first week. On one hand, these findings suggest that the classifier may not be as beneficial to the more immature and sicker infants, who tend to be extubated beyond the first week of life. On the other hand, an enhanced prediction of extubation readiness in the first week of life may prevent the clinical instability that may occur with a failed extubation during this critical period. As for the observation window used to define extubation failure, the classifier only correctly identified 50% of failures that occurred between 72 h and 7 days post-extubation. Considering that reintubations between 72 h and 7 days post-extubation are rarely caused by new non-respiratory pathologies (ex: infection or necrotizing enterocolitis), the classifier failed to capture a large number of true, respiratory-related reintubations. However, given that reintubations occurring within 48-72 h after extubation may be associated with the highest risk of mortality/morbidity, the classifier may be justifiable as a targeted tool for those infants at highest risk of complications.

The study had some limitations. The predictor was applied when the medical team has made the decision to extubate which

tends to increase sensitivity and decrease specificity.⁴³ Also, it may be possible that infants were 'ready' before which is difficult to precisely determine in clinical practice. Furthermore, it is unknown how the predictor would perform if it were conducted in infants on higher ventilatory settings or prior to being recognized as 'ready' by the clinical team. Also, although most respiratory features in the final classifier were recorded during ET-CPAP (Table. E3), it is unknown if reliable and useful recordings of respiratory data would be possible in infants under high frequency ventilation. Due to the small number of patients, a separate analysis of the performance of the predictive tool across centers was not possible. The signal acquisition set-up involved instrumentation and the possibility of recordings of low-quality signals, but the use of small multimodal wireless sensors may make data acquisition more practical and reliable.^{44,45} The classifier was developed using a pre-specified observation window of 72 h which may not capture all respiratory-related reintubations.³³ However, the performance of the classifier decreased in the post-hoc analysis using a longer period of observation. Lastly, the pragmatic design of the study introduced significant heterogeneity in patient characteristics and pretest probabilities of extubation success which may have decreased the accuracy of the final classifier. Indeed, restriction to infants extubated <7 days of age showed an improvement on its performance. Although the classifier demonstrated better prediction of extubation failure, this was achieved at the expense of misclassifying a percentage of successful extubations. Therefore, at this point this tool should not

be used to avoid extubation as it would prolong mechanical ventilation in some infants unnecessarily. However, by pointing out a population with a higher probability of failure in some infants considered 'ready' for elective extubation, the classifier can identify the patients requiring a more intense and continuous monitoring after disconnection from MV and that could maybe benefit from different modes of non-invasive support. Moreover, the classifier can be used to more precisely select a high-risk group to be enrolled in studies testing interventions to decrease extubation failure. Therefore, at the moment this tool should not be used for clinical decisions as this novel approach requires further fine-tuning adjustments and a more friendly instrumentation. The study has several important strengths. Novel cardiorespiratory features and detailed clinical information during the peri-extubation period were acquired from a multi-institutional population. Results are generalizable given the pragmatic design, large sample size, and heterogeneity of clinical practices to determine extubation readiness. A principal component analysis was used to provide only the features that explained the most variability. Finally, the effects of class imbalance were mitigated by pre-classifying the population during the CD stage and randomly under-sampling success patients during training, thereby giving an equal number of success and failure patients to the classifier. Therefore, the advanced analytical methods and interdisciplinary collaboration used in the APEX study are a unique and important step forward towards development of tools able to objectively and effectively expedite extubation^{46,47}. Indeed, sustainable collaborations between disciplines are changing the focus towards multimodal data recording and analysis to enable the right choice of treatment, for the right patient, and at the right time.

CONCLUSION

The combined use of clinical data with automated analyses of cardiorespiratory signals by using machine learning algorithms may provide an adjunct tool to improve prediction of extubation outcomes, but still requires further refinement before adoption into clinical practice. This is critical for planning targeting trials in this understudied high-risk population. As an automated and objective method that requires no human intervention, APEX requires further investigation in larger populations from varied settings to understand its effect on patient outcomes, safety, and generalizability.

DATA AVAILABILITY

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

REFERENCES

1. Stoll, B. J. et al. Trends in care practices, morbidity, and mortality of extremely preterm neonates, 1993-2012. *JAMA* **314**, 1039-1051 (2015).
2. Walsh, M. C. et al. Extremely low birthweight neonates with protracted ventilation: mortality and 18-month neurodevelopmental outcomes. *J. Pediatr.* **146**, 798-804 (2005).
3. Keszler, M. & Sant'Anna, G. Mechanical ventilation and bronchopulmonary dysplasia. *Clin. Perinatol.* **42**, 781-796 (2015).
4. Harris, C., Bisquera, A., Lunt, A., Peacock, J. L. & Greenough, A. Outcomes of the neonatal trial of high-frequency oscillation at 16 to 19 years. *N. Engl. J. Med.* **383**, 689-691 (2020).
5. Shalish, W., Keszler, M., Davis, P. G., Sant'Anna, G. M. Decision to extubate extremely preterm infants: art, science or gamble? *Arch Dis Child Fetal Neonatal Ed.* **107**, 105-112 (2021).
6. Jensen, E. A. et al. Effects of multiple ventilation courses and duration of mechanical ventilation on respiratory outcomes in extremely low-birth-weight infants. *JAMA Pediatr.* **169**, 1011-1017 (2015).
7. Chawla, S. et al. Markers of successful extubation in extremely preterm infants, and morbidity after failed extubation. *J. Pediatr.* **189**, 113-119 (2017).
8. Manley, B. J., Doyle, L. W., Owen, L. S. & Davis, P. G. Extubating extremely preterm infants: predictors of success and outcomes following failure. *J. Pediatr.* **173**, 45-49 (2016).
9. Shalish, W., Latremouille, S., Papenburg, J. & Sant'Anna, G. M. Predictors of extubation readiness in preterm infants: a systematic review and meta-analysis. *Arch. Dis. Child Fetal Neonatal Ed.* **104**, F89-F97 (2019).
10. Shalish, W. et al. The impact of time interval between extubation and reintubation on death or bronchopulmonary dysplasia in extremely preterm infants. *J. Pediatr.* **205**, 70-76 (2019).
11. Milea, D. et al. Artificial intelligence to detect papilledema from ocular fundus photographs. *N. Engl. J. Med.* **382**, 1687-1695 (2020).
12. Wijnberge, M. et al. Effect of a machine learning-derived early warning system for intraoperative hypotension vs standard care on depth and duration of intraoperative hypotension during elective noncardiac surgery: the HYPE randomized clinical trial. *JAMA* **323**, 1052-1060 (2020).
13. Komorowski, M., Celi, L. A., Badawi, O., Gordon, A. C. & Faisal, A. A. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nat. Med.* **24**, 716-720 (2018).
14. Moorman, J. R. et al. Mortality reduction by heart rate characteristic monitoring in very low birth weight neonates: a randomized trial. *J. Pediatr.* **159**, 900-6.e1 (2011).
15. Wysocki, M. et al. Reduced breathing variability as a predictor of unsuccessful patient separation from mechanical ventilation. *Crit. Care Med.* **34**, 2076-2083 (2006).
16. Knox, K. E. et al. High breath-by-breath variability is associated with extubation failure in children. *Crit. Care Med.* **48**, 1165-1174 (2020).
17. Precup, D. et al. Prediction of extubation readiness in extreme preterm infants based on measures of cardiorespiratory variability. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2012**, 5630-5633 (2012).
18. Robles-Rubio, C. A. et al. Automated analysis of respiratory behavior in extremely preterm infants and extubation readiness. *Pediatr. Pulmonol.* **50**, 479-486 (2015).
19. Kanbar, L. J. et al. Cardiorespiratory behavior of preterm infants receiving continuous positive airway pressure and high flow nasal cannula post extubation: randomized crossover study. *Pediatr. Res.* **87**, 62-68 (2020).
20. Latremouille, S., Bhuller, M., Rao, S., Shalish, W. & Sant'Anna, G. Diaphragmatic activity and neural breathing variability during a 5-min endotracheal continuous positive airway pressure trial in extremely preterm infants. *Pediatr. Res.* **89**, 1810-1817 (2021).
21. Shalish, W. et al. Prediction of extubation readiness in extremely preterm infants by the automated analysis of cardiorespiratory behavior: study protocol. *BMC Pediatrics* **17**, 167 (2017).
22. Kanbar, L. J. et al. Organizational principles of cloud storage to support collaborative biomedical research. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2015**, 1231-1234 (2015).
23. Kanbar, L. J. et al. Automated ongoing data validation and quality control of multi-institutional studies. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2016**, 2504-2507 (2016).
24. Robles-Rubio, C. A., Brown, K. A. & Kearney, R. E. Automated unsupervised respiratory event analysis. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **30**, 3201-3204 (2011).
25. Robles-Rubio, C. A., Kearney, R. E., Bertolizio, G. & Brown, K. A. Automatic unsupervised respiratory analysis of infant respiratory inductance plethysmography signals. *PLoS ONE* **15**, e0238402 (2020).
26. Onu, C. C. et al. A semi-Markov chain approach to modeling respiratory patterns prior to extubation in preterm infants. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2017**, 2022-2026 (2017).
27. Choi, S. Independent component analysis. In: Li S. Z. and Jain A. K. (eds) *Encyclopedia of Biometrics* (Springer, Boston, MA, 2015).
28. Breiman, L. Random forests. *Mach. Learn.* **45**, 5-32 (2001).
29. Chen, C., Liaw, A., Breiman, L. Using random forest to learn imbalanced data. (University of California, Berkeley, 2004).
30. Breiman, L. Bagging predictors. *Mach. Learn.* **26**, 123-140 (1996).
31. Lemaitre, G., Nogueira, F. & Aridas, C. K. Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. *J. Mach. Learn. Res.* **18**, 1-5 (2017).
32. Obuchowski, N. A. & McClish, D. K. Sample size determination for diagnostic accuracy studies involving binormal ROC curve indices. *Stat. Med.* **16**, 1529-1542 (1997).
33. Shalish, W. et al. Patterns of reintubation in extremely preterm infants: a longitudinal cohort study. *Pediatr. Res.* **83**, 969-975 (2018).
34. Giaccone, A., Jensen, E., Davis, P. & Schmidt, B. Definitions of extubation success in very premature infants: a systematic review. *Arch. Dis. Child Fetal Neonatal Ed.* **99**, F124-F127 (2014).
35. Al-Mandari, H. et al. International survey on periextubation practices in extremely preterm infants. *Arch. Dis. Child Fetal Neonatal Ed.* **100**, F428-F431 (2015).

36. Beltempo, M. et al. Respiratory management of extremely preterm infants: an international survey. *Neonatology* **114**, 28–36 (2018).
37. Manley, B. J. & Davis, P. G. Solving the extubation equation: successfully weaning infants born extremely preterm from mechanical ventilation. *J. Pediatr.* **189**, 17–18 (2017).
38. Gupta, D. et al. A predictive model for extubation readiness in extremely preterm infants. *J. Perinatol.* **39**, 1663–1669 (2019).
39. Danan, C. et al. randomized trial of delayed extubation for the reduction of reintubation in extremely preterm infants. *Pediatr. Pulmonol.* **43**, 117–124 (2008).
40. SUPPORT Study Group of the Eunice Kennedy Shriver NICHD Neonatal Research Network. Early CPAP versus surfactant in extremely preterm infants. *N. Engl. J. Med.* **362**, 1970–1979 (2010).
41. Bassler, D. et al. Early inhaled budesonide for the prevention of bronchopulmonary dysplasia. *N. Engl. J. Med.* **373**, 1497–1506 (2015).
42. Kirpalani, H. et al. Effect of sustained inflations vs intermittent positive pressure ventilation on bronchopulmonary dysplasia or death among extremely preterm infants: the SAIL randomized clinical trial. *JAMA* **321**, 1165–1175 (2019).
43. Whiting, P. et al. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann. Intern Med* **140**, 189–202 (2004).
44. Chung, H. U. et al. Binodal, wireless epidermal electronic systems with in-sensor analytics for neonatal intensive care. *Science* **363**, eaau0780 (2019).
45. Chung, H. U. et al. Skin-interfaced biosensors for advanced wireless physiological monitoring in neonatal and pediatric intensive-care units. *Nat. Med.* **26**, 418–429 (2020).
46. Tobin, M. J. Extubation and the myth of “minimal ventilator settings”. *Am. J. Respir. Crit. Care Med.* **185**, 349–350 (2012).
47. Thille, A. W., Richard, J. C. & Brochard, L. The decision to extubate in the intensive care unit. *Am. J. Respir. Crit. Care Med.* **187**, 1294–1302 (2013).

AUTHOR CONTRIBUTIONS

L.J.K. provided substantial contributions to the data acquisition, analysis, and interpretation of data; drafted the work and revised it critically for important intellectual content; gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. W.S. provided substantial contributions to the conception and design of the work, data acquisition, analysis and interpretation; drafted the work and revised it critically for important intellectual content; gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. C.C.O. provided substantial contributions to the data acquisition, analysis, and interpretation of data; revised the manuscript critically for important intellectual content; gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. S.L. provided substantial contributions to the data acquisition, analysis, and interpretation of data; revised the manuscript critically for important intellectual content; gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. M.K. provided substantial contributions to the design of the work, data interpretation; and revised the manuscript critically for important intellectual content; gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. S.C. provided substantial contributions to the design of the work, data interpretation; and revised the manuscript critically for important intellectual content; gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. K.A.B. provided

substantial contributions to the conception and design of the work and interpretation of data; revised the manuscript critically for important intellectual content; gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. D.P. provided substantial contributions to data analysis and interpretation; revised the manuscript critically for important intellectual content; gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. R.E.K. provided substantial contributions to the conception and design of the work; data acquisition, analysis, and interpretation of data; revised the manuscript critically for important intellectual content; gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. G.M.S. provided substantial contributions to the conception and design of the work and interpretation of data; drafted the work and revised it critically for important intellectual content; gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING

This work was supported by a major grant from the *Canadian Institutes of Health Research (CIHR)*. Other sources of individual support were: L.J.K. was supported in part by the *Natural Sciences and Engineering Research Council of Canada*; W.S. received personal support awards from the *Richard and Edith Strauss Clinical Fellowship in Medicine* at McGill University and from the *Fonds de Recherche Santé Québec (FRSQ)*; S.L. received the *Max E. Binz Fellowship salary award* from the Department of Medicine and a doctoral training award from the *Fonds de Recherche Santé Québec (FRSQ)* in partnership with the *Fondation Des Étoiles*, McGill University; S.C. was supported by the *Children’s Research Center of Michigan and CHM Foundation*, Hutzel Women’s Hospital and K.A.B. was supported in part by the *Queen Elizabeth Hospital of Montreal Foundation Chair in Pediatric Anesthesia*.

COMPETING INTERESTS

The authors declare no competing interests.

CONSENT STATEMENT

For all patients written informed consent was obtained from parents prior to enrollment.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41390-022-02210-9>.

Correspondence and requests for materials should be addressed to Guilherme M. Sant’Anna.

Reprints and permission information is available at <http://www.nature.com/reprints>

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

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.