

Assessment of Extubation Readiness Using Spontaneous Breathing Trials in Extremely Preterm Neonates

Wissam Shalish, MD; Lara Kanbar, MSc; Lajos Kovacs, MD; Sanjay Chawla, MD; Martin Keszler, MD; Smita Rao; Samantha Latremouille, MSc; Doina Precup, PhD; Karen Brown, MD; Robert E. Kearney, PhD; Guilherme M. Sant'Anna, MD, PhD

[+ Supplemental content](#)

IMPORTANCE Spontaneous breathing trials (SBTs) are used to determine extubation readiness in extremely preterm neonates (gestational age ≤ 28 weeks), but these trials rely on empirical combinations of clinical events during endotracheal continuous positive airway pressure (ET-CPAP).

OBJECTIVES To describe clinical events during ET-CPAP and to assess accuracy of comprehensive clinical event combinations in predicting successful extubation compared with clinical judgment alone.

DESIGN, SETTING, AND PARTICIPANTS This multicenter diagnostic study used data from 259 neonates seen at 5 neonatal intensive care units from the prospective Automated Prediction of Extubation Readiness (APEX) study from September 1, 2013, through August 31, 2018. Neonates with birth weight less than 1250 g who required mechanical ventilation were eligible. Neonates deemed to be ready for extubation and who underwent ET-CPAP before extubation were included.

INTERVENTIONS In the APEX study, cardiorespiratory signals were recorded during 5-minute ET-CPAP, and signs of clinical instability were monitored.

MAIN OUTCOMES AND MEASURES Four clinical events were documented during ET-CPAP: apnea requiring stimulation, presence and cumulative durations of bradycardia and desaturation, and increased supplemental oxygen. Clinical event occurrence was assessed and compared between extubation pass and fail (defined as reintubation within 7 days). An automated algorithm was developed to generate SBT definitions using all clinical event combinations and to compute diagnostic accuracies of an SBT in predicting extubation success.

RESULTS Of 259 neonates (139 [54%] male) with a median gestational age of 26.1 weeks (interquartile range [IQR], 24.9-27.4 weeks) and median birth weight of 830 g (IQR, 690-1019 g), 147 (57%) had at least 1 clinical event during ET-CPAP. Apneas occurred in 10% (26 of 259) of neonates, bradycardias in 19% (48), desaturations in 53% (138), and increased oxygen needs in 41% (107). Neonates with successful extubation (71% [184 of 259]) had significantly fewer clinical events (51% [93 of 184] vs 72% [54 of 75], $P = .002$), shorter cumulative bradycardia duration (median, 0 seconds [IQR, 0 seconds] vs 0 seconds [IQR, 0-9 seconds], $P < .001$), shorter cumulative desaturation duration (median, 0 seconds [IQR, 0-59 seconds] vs 25 seconds [IQR, 0-90 seconds], $P = .003$), and less increase in oxygen (median, 0% [IQR, 0%-6%] vs 5% [0%-18%], $P < .001$) compared with neonates with failed extubation. In total, 41 602 SBT definitions were generated, demonstrating sensitivities of 51% to 100% (median, 96%) and specificities of 0% to 72% (median, 22%). Youden indices for all SBTs ranged from 0 to 0.32 (median, 0.17), suggesting low accuracy. The SBT with highest Youden index defined SBT pass as having no apnea (with desaturation requiring stimulation) or increase in oxygen requirements by 15% from baseline and predicted extubation success with a sensitivity of 93% and a specificity of 39%.

CONCLUSIONS AND RELEVANCE The findings suggest that extremely preterm neonates commonly show signs of clinical instability during ET-CPAP and that the accuracy of multiple clinical event combinations to define SBTs is low. Thus, SBTs may provide little added value in the assessment of extubation readiness.

JAMA Pediatr. 2020;174(2):178-185. doi:10.1001/jamapediatrics.2019.4868
Published online December 20, 2019.

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Guilherme M. Sant'Anna, MD, PhD, Division of Neonatology, Department of Pediatrics, Montreal Children's Hospital, McGill University, 1001 Boul Décarie, Room B05.2714, Montreal, QC H4A 3J1, Canada (guilherme.santanna@mcgill.ca).

Exremely preterm neonates (gestational age ≤ 28 weeks) commonly require mechanical ventilation after birth.¹ Given the known harms associated with prolonged mechanical ventilation, clinicians strive to limit mechanical ventilation exposure by routinely assessing neonates' readiness for extubation.² Currently, the decision to extubate relies on clinical judgment through interpretation of ventilatory support, blood gas values, and overall clinical stability of the neonate.^{3,4} However, clinical judgment is subjective, leads to variable practices, and is often associated with inaccurate decisions.³ In fact, nearly one-third of neonates require reintubation within 7 days after the first extubation attempt.⁵

In recent years, spontaneous breathing trials (SBTs) have increasingly been used to determine extubation readiness.^{3,4} The SBTs entail a 3- to 10-minute period of spontaneous breathing via endotracheal continuous positive airway pressure (ET-CPAP), during which pass or fail is determined from a combination of clinical events (apnea, bradycardias, and desaturations). To date, only 2 small studies^{6,7} investigated SBT accuracies in predicting successful extubation compared with clinical judgment alone; an SBT pass identified almost all successful extubations (excellent sensitivity) but misclassified one-third of failed extubations (low specificity).⁶⁻⁸ Of note, cut-offs to define SBT pass or fail were chosen empirically with no background knowledge about the range of clinical events that normally occur during the trial. Thus, the objectives of this study were to describe the occurrence of clinical events in extremely preterm neonates during ET-CPAP and to evaluate the accuracy of more-comprehensive pass or fail definitions in predicting extubation success compared with clinical judgment alone. We conjectured that such inclusive evaluation would identify an SBT definition with better overall accuracy.

Methods

Study Design and Context

This diagnostic study used data from the Automated Prediction of Extubation Readiness (APEX) prospective multicenter study and was reported using the Standards for Reporting of Diagnostic Accuracy (STARD) reporting guideline. The APEX study aimed to develop an automated predictor of extubation readiness using machine learning tools that integrate clinical variables and quantitative measures of cardiorespiratory behavior in extremely preterm neonates deemed to be ready for extubation.⁹ Enrollment for APEX has been completed, but the development and validation of the predictor are ongoing. As part of APEX, neonates had cardiorespiratory signals acquired electronically during 5 minutes of ET-CPAP immediately preceding extubation. The rationale was to capture intrinsic respiratory behavior without interference from mechanical inflations. Extubation was not predicated on ET-CPAP findings, and no cutoffs for pass or fail were prespecified. However, bedside clinical events that occurred during ET-CPAP were prospectively documented, allowing for this analysis. The institutional review board at each participating institution provided approval for the original APEX study. As for the present study, institutional review board

Key Points

Question Can spontaneous breathing trials improve clinicians' ability to assess extubation readiness in extremely preterm neonates?

Findings In this diagnostic study of 259 extremely preterm neonates, 57% developed signs of clinical instability during a 5-minute period of endotracheal continuous positive airway pressure. In an evaluation of 41 602 combinations of clinical events to define spontaneous breathing trial success or failure, all combinations of clinical events had low accuracies in predicting extubation success compared with clinical judgment alone.

Meaning The findings suggest that spontaneous breathing trials are unwarranted in clinical practice because they may expose neonates to clinical instability without improving the ability to assess extubation readiness.

approval was waived because the analysis involved no additional data collection beyond that which was collected at APEX. Informed consent was written and only obtained for the original APEX study.

Participants

All consecutive neonates admitted to 5 tertiary care neonatal intensive care units in North America between September 1, 2013, and August 31, 2018, were screened for eligibility. The APEX study inclusion and exclusion criteria have been published and are detailed elsewhere.⁹ In brief, extremely preterm neonates with birth weight less than 1250 g and requiring mechanical ventilation who had cardiorespiratory signals acquired before the first elective extubation were included.

Test Methods

Reference Standard: Clinical Judgment

All neonates were extubated using the reference standard of clinical judgment (ie, once they were deemed to be ready for extubation by the treating team). At the time of extubation, data pertaining to postmenstrual age (presented in weeks and referring to the gestational age plus the time elapsed after birth [postnatal age]), postnatal age, weight, ventilator mode, mean airway pressure, fraction of inspired oxygen (FiO₂), and blood gas values (measured within 24 hours before extubation) were collected. Of note, SBTs were not a part of clinical practice at participating sites.

Index Test: ET-CPAP

All included neonates underwent 5-minute ET-CPAP before extubation. The ET-CPAP was equivalent to the positive end-expiratory pressure (PEEP) preset by the clinical team on conventional mechanical ventilation. No pressure support was provided, and the backup rate was turned off. During ET-CPAP, a research investigator (including W.S., M.K., S.R., S.L., and G.M.S.) and respiratory therapist monitored the neonates for apnea, bradycardia, or desaturation and intervened per clinical discretion. Interventions included increasing FiO₂ from baseline, stimulation in case of apnea, and termination of ET-CPAP (ie, resumption of mechanical ventilation) if necessary. Concomitantly, the following data were collected: PEEP

level, baseline oxygen saturation, and FiO_2 just before starting ET-CPAP; presence and cumulative durations of desaturations (baseline oxygen saturation <85%) and bradycardias (heart rate <100 bpm); need for additional oxygen from baseline (and highest amount provided); and total ET-CPAP duration. After ET-CPAP was completed, the clinical team extubated neonates to noninvasive respiratory support. Of note, the study design did not include blinding the treating team from the use of ET-CPAP. Consequently, the treating team members were permitted to change their minds about extubation at any time, in which case neonates would be eligible for the study again (ie, when deemed to be ready for the next extubation by the clinical team). In those cases, only the final ET-CPAP trial would be included for analysis.

Statistical Analysis

The first objective was to describe occurrences of 4 clinical event categories during ET-CPAP: apneas requiring stimulation, bradycardias, desaturations, and increase in oxygen supplementation from baseline. To avoid any confounding, neonates who transitioned to ET-CPAP with a baseline oxygen saturation already below 85% were excluded. The number and proportion of neonates with apneas, bradycardias, desaturations, and increased oxygen needs were determined. Also, medians and interquartile ranges (IQRs) were ascertained for the cumulative durations of bradycardia or desaturation and the additional amount of oxygen needed.

The second objective was to evaluate the accuracy of different SBT definitions in predicting successful extubation compared with clinical judgment alone. Extubation success was defined as not needing reintubation within 7 days. First, clinical events were compared between neonates who succeeded or failed extubation using Wilcoxon rank sum test, χ^2 test, or Fisher exact test as appropriate. For continuous variables, probability density functions were plotted to better visualize the overlap between success and failure groups, and areas under the receiver operating characteristic curve were computed. Continuous variables were transformed into binary variables using equally spaced cutoff points ranging from 0 to 100 seconds for cumulative bradycardia or desaturation durations and from 0% to 30% for supplemental oxygen needed. With use of these variables, an automated algorithm was developed to create multiple combinations of the 4 clinical events with both AND/OR logical operators. Examples of generated SBT definitions are shown in eTable 1 in the Supplement. Of note, neonates with missing data for any clinical event were excluded. Sensitivity, specificity, and positive and negative predictive values of a passed SBT were computed for all derived SBTs along with their respective 95% CIs. Sensitivity referred to the proportion of neonates with successful extubation correctly identified by a passed SBT, whereas specificity referred to the proportion of neonates with failed extubations correctly identified by a failed SBT (definitions and interpretation of all diagnostic terms are provided in the eMethods in the Supplement). The diagnostic performance of each SBT was graphically displayed, and the accuracy of each SBT in predicting extubation success was estimated using the Youden index. The latter is a measure of a test's overall discriminative power

assuming equal weight between sensitivity and specificity and ranges from zero (poor accuracy) to 1 (perfect accuracy). Furthermore, the best SBT definition was identified for each of the following diagnostic goals: (1) achievement of best overall accuracy (ie, highest Youden index), (2) achievement of maximal ability to detect failures (ie, maximal specificity), and (3) achievement of a minimal number of misclassified neonates with extubation success (ie, maximal sensitivity).

A priori, we recognized that SBT accuracy might be influenced by the pretest probability of extubation success in the cohort and the observation window used to define extubation success. To test the former, the diagnostic performance of SBTs was evaluated for neonates above and below the median gestational age (a known marker of extubation success).^{10,11} To test the latter, the diagnostic performance of SBTs was computed for 4 additional definitions of extubation success using observation windows of 24 hours, 48 hours, 72 hours, and 5 days after extubation.

Lastly, based on the sample size and prevalence of successful extubation in the APEX cohort and assuming a 2-tailed $\alpha = .05$, the computed SBT sensitivities would be estimated with approximately 5% precision and specificities would be estimated with approximately 10% precision.¹² All analyses were conducted using MATLAB R2018a (The MathWorks Inc).

Results

Of 605 eligible neonates, 278 underwent ET-CPAP after they were deemed to be ready for extubation (Figure 1). There were 7 circumstances in which clinicians changed their minds about extubation after ET-CPAP; 4 cases were subsequently excluded, and 3 were later restudied after the neonate was deemed to be ready for extubation. Therefore, 274 neonates were extubated after ET-CPAP. After all additional exclusions, a total of 259 neonates (139 [54%] male) were included in this study.

Primary Objective

Characteristics of patients at the time of extubation and during ET-CPAP are presented in Table 1 and Figure 2. The median gestational age was 26.1 weeks (IQR, 24.9-27.4 weeks), median birth weight was 830 g (IQR, 690-1019 g), and median postnatal age was 8 days (IQR, 3-26 days) at extubation. Caffeine was administered in 253 of 259 neonates (98%). Continuous positive airway pressure was the most common postextubation respiratory support (150 of 259 [58%]), followed by nasal intermittent positive pressure ventilation (96 of 259 [37%]) and high-flow nasal cannula (13 of 259 [5%]). The ET-CPAP was performed using a median PEEP of 5 cm H_2O (IQR, 5-6 cm H_2O) and median interval of 32 minutes (IQR, 21-59 minutes) before extubation. Assessors decided to prematurely terminate ET-CPAP in 21 neonates based on variable thresholds of clinical events (eTable 2 in the Supplement). During ET-CPAP, apneas occurred in 10% (26 of 259), bradycardias in 19% (48), desaturations in 53% (138), and increased oxygen needs in 41% (107) of neonates. Cumulative durations of bradycardias occurred from 2 to 114 seconds (median, 15 seconds), and desaturations ranged from 2 to 240 seconds (median, 61 sec-

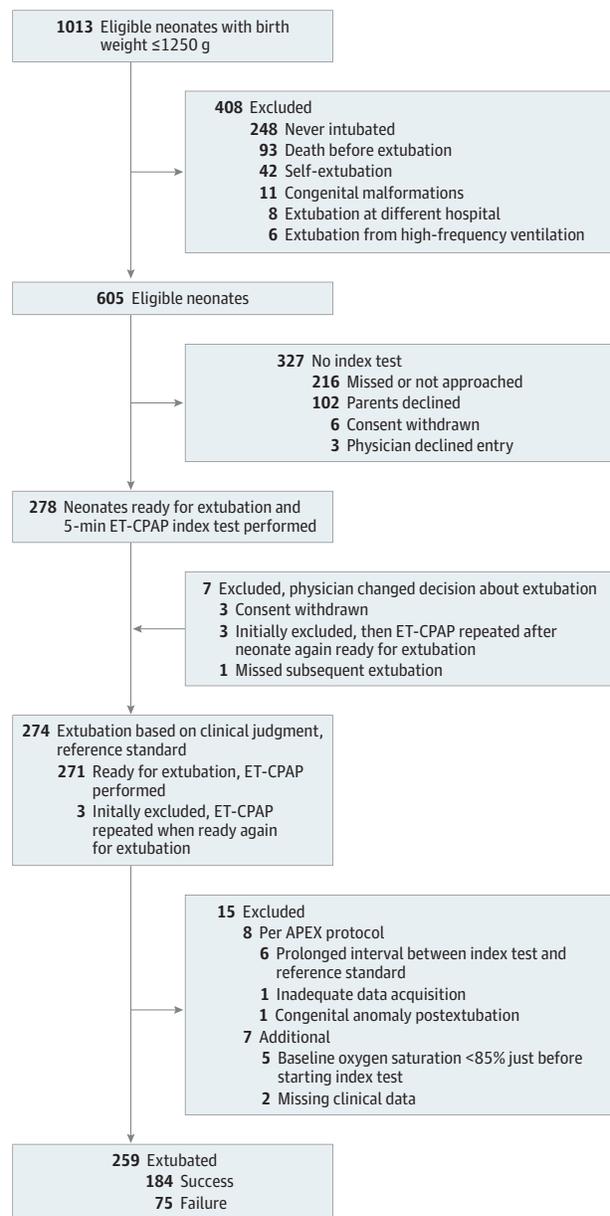
onds), whereas the amount of additional oxygen provided ranged from 2% to 77% (median, 10%). Altogether, 147 neonates (57%) experienced at least 1 clinical event during ET-CPAP, with variable combinations. The combination of desaturation and increased oxygen needs was most common, occurring in 58 of 147 neonates (39%) with a clinical event.

Secondary Objective

A total of 184 neonates (71%) were successfully extubated. Extubation success was significantly associated with older gestational age, increased weight at birth, increased weight at extubation, and less respiratory support (reduced mean airway pressure and FiO_2) at extubation compared with extubation failure (Table 1). During ET-CPAP, neonates with successful extubation were significantly less likely to have early ET-CPAP termination (3% [6 of 184] vs 20% [15 of 75], $P < .001$), had fewer clinical events (51% [93 of 184] vs 72% [54 of 75], $P = .002$), had shorter cumulative durations of bradycardia (median, 0 seconds [IQR, 0 seconds] vs 0 seconds [IQR, 0-9 seconds], $P < .001$) or desaturations (median, 0 seconds [IQR, 0-59 seconds] vs 25 seconds [IQR, 0-90 seconds], $P = .003$), and received smaller amounts of oxygen (median, 0% [IQR, 0%-6%] vs 5% [0%-18%], $P < .001$) compared with neonates with failed extubation (Table 2). When evaluated separately, the absence of each categorical clinical event predicted successful extubation as follows: absence of apnea (sensitivity, 96%; specificity, 24%), absence of bradycardia (sensitivity, 87%; specificity, 32%), absence of desaturations (sensitivity, 53%; specificity, 69%), and absence of increased oxygen needs (sensitivity, 64%; specificity, 53%). Moreover, the diagnostic performance of each continuous clinical event was characterized by low areas under the receiver operating characteristic curve (cumulative bradycardia duration [0.60], cumulative desaturation duration [0.61], and amount of additional oxygen provided [0.63]) and a high degree of overlap between the probability density functions (eFigure 1 in the Supplement).

After excluding 7 neonates with missing data, the automated algorithm was applied to the remaining 252 neonates to create all combinations of the 4 clinical events, thus generating a total of 41 602 SBT definitions. The SBTs had sensitivities ranging from 51% to 100% (median, 96%), specificities from 0% to 72% (median, 22%), positive predictive values from 71% to 82% (median, 75%), and negative predictive values from 33% to 100% (median, 67%) (Figure 3). Youden indices ranged from 0 to 0.32 (median, 0.17), suggesting an overall low accuracy. The best SBT definitions to achieve the highest Youden index, maximal specificity, and maximal sensitivity are provided in eTable 3 of the Supplement. The combination of clinical events with highest Youden index defined a passed SBT as having no apnea (with desaturation requiring stimulation) or increase in oxygen requirements by 15% from baseline. After applying this definition to the cohort, 171 of 184 neonates (93%) with successful extubation would have passed SBT (sensitivity, 93%) and 29 of 75 neonates (39%) with failed extubation would have failed SBT (specificity, 39%). As such, 13 neonates (7%) with successful extubation would have failed the SBT (and would have received mechanical ventilation for longer than necessary), whereas 46 neonates (61%) with failed

Figure 1. Flow Diagram



APEX indicates Automated Prediction of Extubation Readiness; ET-CPAP, endotracheal continuous positive airway pressure.

extubation would have still passed the SBT. The best SBT that achieved maximal specificity resulted in the detection of 72% of failed extubations, while misclassifying 46% of successful extubations. Finally, the best SBT that achieved maximal sensitivity (100%) correctly identified 14 of 75 failed extubations (19%) without misclassifying any successful extubation as a failure.

Analyses of Variability in Diagnostic Accuracy

The SBT accuracies were further evaluated for neonates above the median gestational age (pretest probability of 84% for successful extubation) and below the median gestational age

Table 1. Characteristics of Neonates Before Extubation and During ET-CPAP^a

Clinical Variable	Cohort (N = 259)	Extubation Success (n = 184)	Extubation Failure (n = 75)	P Value
Gestational age, wk	26.1 (24.9-27.4)	26.4 (25.0-27.9)	25.4 (24.5-26.4)	<.001
Birth weight, g	830 (690-1019)	880 (715-1073)	740 (633-872)	<.001
Male, No. (%)	139 (54)	97 (53)	42 (56)	.63
Antenatal steroids, No. (%)	233 (90)	166 (90)	67 (89)	.83
Cesarean delivery, No. (%)	173 (67)	127 (69)	46 (61)	.23
Apgar at 5 min ^b	7 (5-8)	7 (5-8)	7 (5-8)	.93
Delivery room intubation, No. (%)	128 (49)	79 (43)	49 (65)	.001
Surfactant, No. (%)	247 (95)	175 (95)	72 (96)	>.99
Caffeine, No. (%)	253 (98)	179 (97)	74 (99)	.68
Preextubation				
Postmenstrual age, wk	28.0 (26.9-29.4)	28.6 (27.4-29.9)	27.4 (26.6-28.5)	<.001
Day of life	8 (3-26)	7 (3-27)	9 (4-25)	.55
Weight, g	940 (810-1080)	988 (850-1120)	820 (720-950)	<.001
Patient-triggered ventilation, No. (%)	133 (51)	89 (48)	44 (59)	.13
Mean airway pressure, cm H ₂ O ^c	7.1 (6.3-8.0)	6.9 (6.2-7.9)	7.5 (6.6-9.0)	.002
FiO ₂	0.23 (0.21-0.27)	0.21 (0.21-0.26)	0.25 (0.22-0.28)	<.001
pH ^d	7.34 (7.29-7.38)	7.34 (7.30-7.38)	7.32 (7.29-7.37)	.21
pCO ₂ , mm Hg ^d	44 (38-51)	44 (37-50)	46 (38-55)	.10
ET-CPAP				
Duration, min	5 (5-5)	5 (5-5)	5 (5-5)	<.001
PEEP, cm H ₂ O	5 (5-6)	5 (5-6)	5 (5-6)	.19
Starting FiO ₂ ^c	0.23 (0.21-0.27)	0.21 (0.21-0.26)	0.26 (0.23-0.29)	<.001
Starting oxygen saturation, % ^e	0.94 (0.92-0.96)	0.95 (0.92-0.97)	0.94 (0.92-0.95)	.03

Abbreviations:
 ET-CPAP, endotracheal continuous positive airway pressure;
 FiO₂, fraction of inspired oxygen;
 pCO₂, partial pressure of carbon dioxide; PEEP, positive end-expiratory pressure.

^a Data are presented as median (interquartile range) unless otherwise indicated.

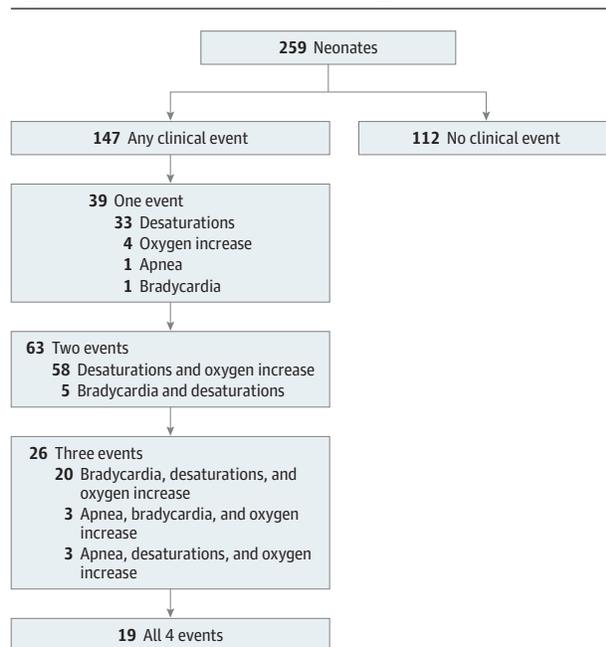
^b Extubation success (n = 183); failure (n = 74).

^c Extubation success (n = 183).

^d Extubation success (n = 152); failure (n = 66).

^e Extubation success (n = 179); failure (n = 74).

Figure 2. Occurrence of Clinical Events During ET-CPAP



Apnea was defined as neonates needing stimulation, bradycardia as heart rate less than 100 beats per minute, desaturations as oxygen saturation less than 85%, and oxygen increase as increase in oxygen requirements from baseline. ET-CPAP indicates endotracheal continuous positive airway pressure.

(pretest probability of 61% for successful extubation) and using different observation windows to define extubation success (eFigures 2 and 3 in the Supplement). Both analyses yielded poor SBT accuracies, as reflected by the low Youden indices (none of which exceeding 0.36).

Discussion

In this diagnostic study of data from a large prospective cohort of extremely preterm neonates, we found that 57% of neonates exhibited at least 1 clinical event while undergoing 5-minute ET-CPAP immediately before extubation. Evaluation of multiple clinical event combinations to define passed or failed SBT revealed that none could distinguish between extubation success and failure with sufficient accuracy to justify their routine use. Together, these results provide additional information on the safety and value of SBTs as currently performed.

Assessment of extubation readiness during a period of spontaneous breathing during ET-CPAP has been done for several years. In the 1980s, some preterm neonates were extubated after passing a 6-hour to 24-hour ET-CPAP trial using PEEP levels of 2 to 3 cm H₂O.¹³⁻¹⁵ This practice was abandoned once evidence showed increased risks of apnea, respiratory acidosis, and extubation failure likely because of low levels of support provided for long periods.¹⁶ Years later, SBTs using shorter time frames (3-5 minutes) and higher PEEP

Table 2. Clinical Events During ET-CPAP and Respective Diagnostic Accuracies at Predicting Successful Extubation

Clinical Event	Extubation Success (n = 184) ^a	Extubation Failure (n = 75) ^a	Rate (95% CI), % ^b				Area Under ROC Curve
			Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	
Apnea needing stimulation, No. (%)	8 (4)	18 (24)	96 (93-99)	24 (14-34)	76 (70-81)	69 (51-87)	NA
Bradycardia, No. (%)	24 (13)	24 (32)	87 (82-92)	32 (21-43)	76 (70-82)	50 (36-64)	NA
Desaturation (oxygen saturation <85%), No. (%)	86 (47)	52 (69)	53 (46-60)	69 (59-80)	81 (74-88)	38 (30-46)	NA
Increase in oxygen requirement from baseline, No. (%)	67 (36)	40 (53)	64 (57-71)	53 (42-65)	78 (70-84)	37 (28-47)	NA
Early ET-CPAP termination, No. (%)	6 (3)	15 (20)	97 (94-99)	20 (11-29)	75 (69-80)	71 (52-91)	NA
Any clinical event, No. (%)	93 (51)	54 (72)	49 (42-57)	72 (62-82)	81 (74-88)	37 (29-45)	NA
All 4 clinical events, No. (%)	4 (2)	15 (20)	98 (96-100)	20 (11-29)	75 (70-80)	79 (61-97)	NA
Desaturation, median (IQR), s	0 (0-59)	25 (0-90)	NA	NA	NA	NA	0.61
Bradycardia, median (IQR), s	0 (0-0)	0 (0-9)	NA	NA	NA	NA	0.60
Supplemental oxygen, median (IQR), %	0 (0-6)	5 (0-18)	NA	NA	NA	NA	0.63

Abbreviations: ET-CPAP, endotracheal continuous positive airway pressure; IQR, interquartile range; NA, not applicable; ROC, receiver operating characteristic.

^b Diagnostic accuracy of the absence of each clinical event during ET-CPAP in predicting extubation success.

^a All comparisons were statistically significant, with $P \leq .01$.

levels (5-6 cm H₂O) were attempted to lessen the risks of lung derecruitment and respiratory fatigue. In 2 small studies,^{6,7} the diagnostic accuracy of SBTs was evaluated among neonates deemed to be ready for extubation using empirical and/or combinations of clinical events to define SBT pass or fail. Both studies showed excellent SBT sensitivities (97% and 92%) but only modest specificities (73% and 50%) at predicting extubation success. Of interest, the only study to our knowledge to prospectively audit the consequences of incorporating routine SBTs into clinical practice showed that SBT-driven extubation was not associated with improvements in extubation success rates or mechanical ventilation durations compared with clinical judgment alone.¹⁷ Furthermore, a randomized clinical trial¹⁸ comparing the effects of SBT vs clinical judgment on time to successful extubation was terminated on grounds of futility. Nonetheless, an increasing number of clinicians worldwide have reported using SBTs for preterm neonates, either as an adjunct to clinical judgment or as part of mechanical ventilation weaning protocols.^{3,4}

A major limitation with current SBTs is that they were defined without foreknowledge of how neonates normally react to a trial of ET-CPAP. In the APEX study, neonates were exposed to a 5-minute ET-CPAP recording without predefined SBT pass or fail criteria, which allowed us to pragmatically describe their clinical behavior in the present study. We found that episodes of apneas, bradycardias, desaturations, and increased oxygen needs frequently occurred during ET-CPAP in various combinations and wide ranges of durations and severities. Although these findings highlight the important heterogeneity in patient behaviors during ET-CPAP, they also reflect a certain degree of variability in the way assessors reacted to clinical events. For example, by allowing assessors to stop ET-CPAP at their discretion, we noted important variations in thresholds for early termination. Thus, SBTs may still leave ample room for subjective interpretation, which unavoidably leads to difficult test reproducibility among assessors. A similar phe-

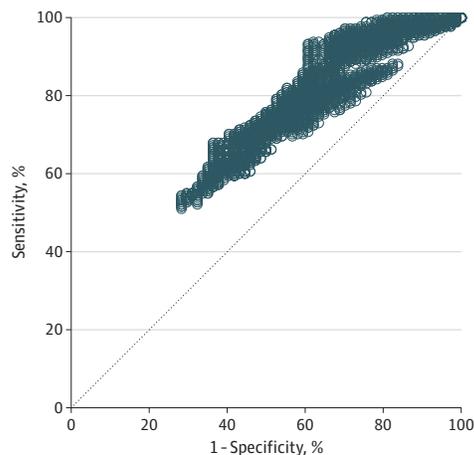
nomenon of variability in SBT performance and reporting practices has been described in adults.¹⁹

Arguably, the documentation of clinical events during ET-CPAP would be justifiable if it could accurately predict which neonates would succeed or fail extubation. In this cohort, although neonates with extubation failure were significantly more likely to have clinical events compared with those with successful extubation, there was considerable overlap between the 2 groups. Consequently, when computing the diagnostic performance of all possible SBT definitions, none had an acceptable trade-off between sensitivity and specificity, as reflected by the low Youden indices. In fact, given that nearly one-third of neonates who failed extubation had an uneventful ET-CPAP recording, they would have been automatically misclassified by any SBT definition. The addition of a 5-minute SBT to clinical judgment appears to be unwarranted because it exposes neonates to clinical instability without improving the ability to identify extubation failures.

Before initiating this analysis, we recognized that the observation window used to define extubation success might influence the diagnostic performance of the SBT. We chose an observation window of 7 days based on the rationale that it would capture most reintubations associated with respiratory factors.⁵ However, it was conceivable that SBTs would be better suited for detecting reintubations occurring within shorter time frames after extubation. Similarly, we recognized that the pretest probability of extubation success in the cohort might be significantly associated with alterations in the test's sensitivity and specificity, as previously described.^{20,21} For those reasons, we explored whether neonates with higher or lower probabilities of successful extubation (based on their gestational age) or those reintubated within shorter time frames after extubation would benefit differently from an SBT. However, no improvements in SBT accuracies could be uncovered.

There are several possible explanations why SBTs evaluated in our study did not accurately capture neonates' likelihood of successful extubation. First, confounding factors such

Figure 3. Diagnostic Performance of 41 602 Spontaneous Breathing Trial (SBT) Definitions for Predicting Successful Extubation



Sensitivity represents the proportion of neonates with successful extubation that passed the SBT; 1 - specificity represents the proportion of neonates with failed extubation that were inaccurately misclassified by the SBT.

as endotracheal tube size, length, and partial obstruction (owing to respiratory secretions or biofilm formation) may have influenced clinical event occurrences during ET-CPAP.²² Second, the 5-minute trial duration may have been too short to accurately assess the neonates' ability to sustain spontaneous breathing without significant apneas, especially considering that apneas are the most commonly reported cause of reintubation in this population.⁵ Third, it is unclear from the available literature whether a PEEP of 5 to 6 cm H₂O during ET-CPAP would accurately match the patient's physiologic conditions after extubation while receiving noninvasive respiratory support. The decrease in mean airway pressure and increased resistance associated with ET-CPAP may have further contributed to clinical instability.

Limitations

An integral part of the study was that clinical events were captured continuously and subjectively through direct observation of the patient and bedside monitor. Although the pragmatic nature of the assessment may be considered a limitation, it likely reflected the way that SBTs are actually

evaluated in clinical practice, thereby adding external validity to the study. Nonetheless, more precise information on the timing, number, duration, and depth of each event and more direct associations between individual events may have provided further understanding of the neonates' clinical behavior during ET-CPAP.

The study has some other limitations. The ET-CPAP, performed only after neonates were deemed to be ready for extubation, introduced test-referral bias, meaning that inevitably more stable patients (likely to be successfully extubated) were preselected for the diagnostic test. This phenomenon has been well described to overestimate sensitivity, underestimate specificity, and compromise test generalizability.^{23,24} Furthermore, because of lack of blinding of the ET-CPAP recording, extubation was postponed for 7 neonates who would have otherwise been extubated per clinical judgment. Whereas these neonates only represented 3% of the cohort, their inclusion may have marginally improved the specificity of the evaluated SBTs. Also, a considerable number of eligible neonates were not approached or missed. Our results may not be generalizable to neonates weighing more than 1250 g who received mechanical ventilation or extubations beyond the first elective attempt.

Conclusions

In this study, extremely preterm neonates commonly showed signs of clinical instability during a 5-minute ET-CPAP trial. Although neonates with extubation failure had significantly more clinical events compared with those with extubation success, the accuracy of more than 41 000 evaluated SBT pass or fail definitions remained low. As such, SBTs as currently performed may provide little to no added value in the assessment of extubation readiness (especially in the identification of extubation failures) compared with clinical judgment alone. Future studies appear to be needed to evaluate the role of SBT duration and the provided PEEP levels necessary in improving the accuracy of the test. Furthermore, ongoing analysis of the APEX study aims to evaluate the value of more complex and automated analyses of cardiorespiratory behavior during mechanical ventilation and ET-CPAP in better predicting extubation success.

ARTICLE INFORMATION

Accepted for Publication: July 30, 2019.

Published Online: December 20, 2019.
doi:10.1001/jamapediatrics.2019.4868

Author Affiliations: Division of Neonatology, Department of Pediatrics, Montreal Children's Hospital, McGill University Health Center, Montreal, Quebec, Canada (Shalish, Rao, Latremouille, Sant'Anna); Division of Biomedical Engineering, McGill University, Montreal, Quebec, Canada (Kanbar, Kearney); Department of Neonatology, Jewish General Hospital, Montreal, Quebec, Canada (Kovacs); Division of Neonatal-Perinatal Medicine, Hutzel Women's Hospital, Wayne State University, Detroit, Michigan (Chawla); Division of Neonatology, Women and Infants Hospital of

Rhode Island, Brown University, Providence (Keszler); Department of Computer Science, McGill University, Montreal, Quebec, Canada (Precup); Department of Anesthesia, Montreal Children's Hospital, McGill University Health Center, Montreal, Quebec, Canada (Brown).

Author Contributions: Drs Shalish and Sant'Anna had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Shalish, Kanbar, Kovacs, Chawla, Precup, Kearney, Sant'Anna.

Acquisition, analysis, or interpretation of data: All authors.

Drafting of the manuscript: Shalish, Kovacs, Rao, Sant'Anna.

Critical revision of the manuscript for important intellectual content: Shalish, Kanbar, Kovacs, Chawla, Keszler, Latremouille, Precup, Brown, Kearney, Sant'Anna.

Statistical analysis: Shalish, Precup, Kearney.

Obtained funding: Precup, Kearney, Sant'Anna.

Administrative, technical, or material support: Shalish, Chawla, Rao, Sant'Anna.

Supervision: Keszler, Precup, Kearney, Sant'Anna.

Conflict of Interest Disclosures: Dr Chawla reported receiving grant R1-2015-82 from Children's Research Center of Michigan and Children's Hospital of Michigan Foundation during the conduct of the study. Drs Precup, Brown, Kearney, and Sant'Anna were all recipients of the grant for Automated Prediction of Extubation Readiness

[APEX]) from the Canadian Institute of Health Research during the conduct of the study. No other disclosures were reported.

REFERENCES

1. Stoll BJ, Hansen NI, Bell EF, et al; Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network. Trends in care practices, morbidity, and mortality of extremely preterm neonates, 1993-2012. *JAMA*. 2015;314(10):1039-1051. doi:10.1001/jama.2015.10244
2. Jensen EA, DeMauro SB, Kornhauser M, Aghai ZH, Greenspan JS, Dysart KC. Effects of multiple ventilation courses and duration of mechanical ventilation on respiratory outcomes in extremely low-birth-weight infants. *JAMA Pediatr*. 2015;169(11):1011-1017. doi:10.1001/jamapediatrics.2015.2401
3. Al-Mandari H, Shalish W, Dempsey E, Keszler M, Davis PG, Sant'Anna G. International survey on perextubation practices in extremely preterm infants. *Arch Dis Child Fetal Neonatal Ed*. 2015;100(5):F428-F431. doi:10.1136/archdischild-2015-308549
4. Beltempo M, Isayama T, Vento M, et al; International Network for Evaluating Outcomes of Neonates. Respiratory management of extremely preterm infants: an international survey. *Neonatology*. 2018;114(1):28-36. doi:10.1159/000487987
5. Shalish W, Kanbar L, Keszler M, et al. Patterns of reintubation in extremely preterm infants: a longitudinal cohort study. *Pediatr Res*. 2018;83(5):969-975. doi:10.1038/pr.2017.330
6. Kamlin CO, Davis PG, Morley CJ. Predicting successful extubation of very low birthweight infants. *Arch Dis Child Fetal Neonatal Ed*. 2006;91(3):F180-F183. doi:10.1136/adc.2005.081083
7. Chawla S, Natarajan G, Gelmini M, Kazzi SN. Role of spontaneous breathing trial in predicting successful extubation in premature infants. *Pediatr Pulmonol*. 2013;48(5):443-448. doi:10.1002/ppul.22623
8. Shalish W, Latremouille S, Papenburg J, Sant'Anna GM. Predictors of extubation readiness in preterm infants: a systematic review and meta-analysis. *Arch Dis Child Fetal Neonatal Ed*. 2019;104(1):F89-F97. doi:10.1136/archdischild-2017-313878
9. Shalish W, Kanbar LJ, Rao S, et al. Prediction of extubation readiness in extremely preterm infants by the automated analysis of cardiorespiratory behavior: study protocol. *BMC Pediatr*. 2017;17(1):167. doi:10.1186/s12887-017-0911-z
10. Chawla S, Natarajan G, Shankaran S, et al; Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network. Markers of successful extubation in extremely preterm infants, and morbidity after failed extubation. *J Pediatr*. 2017;189:113-119.e2. doi:10.1016/j.jpeds.2017.04.050
11. Manley BJ, Doyle LW, Owen LS, Davis PG. Extubating extremely preterm infants: predictors of success and outcomes following failure. *J Pediatr*. 2016;173:45-49. doi:10.1016/j.jpeds.2016.02.016
12. Flahault A, Cadiilhac M, Thomas G. Sample size calculation should be performed for design accuracy in diagnostic test studies. *J Clin Epidemiol*. 2005;58(8):859-862. doi:10.1016/j.jclinepi.2004.12.009
13. Kim EH. Successful extubation of newborn infants without preextubation trial of continuous positive airway pressure. *J Perinatol*. 1989;9(1):72-76.
14. Kim EH, Boutwell WC. Successful direct extubation of very low birth weight infants from low intermittent mandatory ventilation rate. *Pediatrics*. 1987;80(3):409-414.
15. Tapia JL, Bancalari A, González A, Mercado ME. Does continuous positive airway pressure (CPAP) during weaning from intermittent mandatory ventilation in very low birth weight infants have risks or benefits? a controlled trial. *Pediatr Pulmonol*. 1995;19(5):269-274. doi:10.1002/ppul.1950190505
16. Davis PG, Henderson-Smart DJ. Extubation from low-rate intermittent positive airways pressure versus extubation after a trial of endotracheal continuous positive airways pressure in intubated preterm infants. *Cochrane Database Syst Rev*. 2001;(4):CD001078. doi:10.1002/14651858.CD001078
17. Kamlin CO, Davis PG, Argus B, Mills B, Morley CJ. A trial of spontaneous breathing to determine the readiness for extubation in very low birth weight infants: a prospective evaluation. *Arch Dis Child Fetal Neonatal Ed*. 2008;93(4):F305-F306. doi:10.1136/adc.2007.129890
18. ClinicalTrials.gov. A randomized trial of the spontaneous breathing trial to extubate very low birthweight neonates (NCT01471431). <https://clinicaltrials.gov/ct2/show/NCT01471431?term=NCT01471431&draw=2&rank=1>. Accessed on November 6, 2019.
19. Godard S, Herry C, Westergaard P, et al. Practice variation in spontaneous breathing trial performance and reporting. *Can Respir J*. 2016;2016:9848942. doi:10.1155/2016/9848942
20. Usher-Smith JA, Sharp SJ, Griffin SJ. The spectrum effect in tests for risk prediction, screening, and diagnosis. *BMJ*. 2016;353:i3139. doi:10.1136/bmj.i3139
21. Leeflang MM, Rutjes AW, Reitsma JB, Hooft L, Bossuyt PM. Variation of a test's sensitivity and specificity with disease prevalence. *CMAJ*. 2013;185(11):E537-E544. doi:10.1503/cmaj.121286
22. Kenaley KM, Blackson T, Boylan L, et al. Impact of endotracheal tube biofilm and respiratory secretions on airway resistance and mechanics of breathing in a neonatal lung model [published online July 26, 2018]. *J Appl Physiol*. 1985. doi:10.1152/jappphysiol.00083.2018
23. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med*. 2004;140(3):189-202. doi:10.7326/0003-4819-140-3-200402030-00010
24. Tobin MJ. The new irrationalism in weaning. *J Bras Pneumol*. 2011;37(5):571-573.