Clinical usefulness of reintubation criteria in extremely preterm infants: a cohort study

Tugba Alarcon-Martinez,[1,2] Samantha Latremouille ,[1] Lajos Kovacs,[3] Robert E Kearney,[4] Guilherme M Sant'Anna ,[1] Wissam Shalish [1]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/archdischild-2022-325245).

[1]Pediatrics, McGill University Health Centre, Montreal, Quebec, Canada
[2]Neonatal Services, The Royal Women's Hospital, Melbourne, VIC, Australia
[3]Department of Neonatology, Jewish General Hospital, Montreal, Quebec, Canada
[4]Biomedical Engineering, McGill University, Montreal, Quebec, Canada

Correspondence to
Dr Wissam Shalish, McGill University Health Centre, Montreal, QC H4A 3J1, Canada; wissam.shalish@mcgill.ca

Received 15 December 2022
Accepted 2 May 2023

To cite: Alarcon-Martinez T, Latremouille S, Kovacs L, *et al*. *Arch Dis Child Fetal Neonatal Ed* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/archdischild-2022-325245

ABSTRACT

Objective To describe the thresholds of instability used by clinicians at reintubation and evaluate the accuracy of different combinations of criteria in predicting reintubation decisions.

Design Secondary analysis using data obtained from the prospective observational Automated Prediction of Extubation Readiness study (NCT01909947) between 2013 and 2018.

Setting Multicentre (three neonatal intensive care units).

Patients Infants with birth weight ≤1250 g, mechanically ventilated and undergoing their first planned extubation were included.

Interventions After extubation, hourly O_2 requirements, blood gas values and occurrence of cardiorespiratory events requiring intervention were recorded for 14 days or until reintubation, whichever came first.

Main outcome measures Thresholds at reintubation were described and grouped into four categories: increased O_2, respiratory acidosis, frequent cardiorespiratory events and severe cardiorespiratory events (requiring positive pressure ventilation). An automated algorithm was used to generate multiple combinations of criteria from the four categories and compute their accuracies in capturing reintubated infants (sensitivity) without including non-reintubated infants (specificity).

Results 55 infants were reintubated (median gestational age 25.2 weeks (IQR 24.5–26.1 weeks), birth weight 750 g (IQR 640–880 g)), with highly variable thresholds at reintubation. After extubation, reintubated infants had significantly greater O_2 needs, lower pH, higher pCO_2 and more frequent and severe cardiorespiratory events compared with non-reintubated infants. After evaluating 123 374 combinations of reintubation criteria, Youden indices ranged from 0 to 0.46, suggesting low accuracy. This was primarily attributable to the poor agreement between clinicians on the number of cardiorespiratory events at which to reintubate.

Conclusions Criteria used for reintubation in clinical practice are highly variable, with no combination accurately predicting the decision to reintubate.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Extremely preterm infants (gestational age ≤28 weeks) are commonly reintubated after their first extubation attempt.
⇒ The decision to reintubate is typically based on the clinicians' assessment of the infant's oxygen needs, gas exchange and frequency/severity of cardiorespiratory events.
⇒ Thresholds of clinical instability used by clinicians to determine the need for reintubation are not well established.

WHAT THIS STUDY ADDS

⇒ In this large multicentre observational study, thresholds used for reintubation varied widely in clinical practice.
⇒ After evaluating 123 374 combinations of reintubation criteria, none could accurately predict which infants were reintubated.
⇒ The low accuracy was primarily attributable to the poor agreement between clinicians on the number of cardiorespiratory events at which to reintubate.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Future research is needed to develop and validate new bedside tools for quantitatively detecting cardiorespiratory events and objectively assessing their clinical significance.

INTRODUCTION

Extremely preterm infants (gestational age ≤28 weeks) commonly require reintubation after extubation. Given that reintubation prolongs exposure to mechanical ventilation and increases adverse outcomes, it is important to understand the factors involved in this complex decision.[1 2] In practice, the decision to reintubate is typically made by the medical team based on evaluation of the infants' clinical instability, as determined by the frequency/severity of cardiorespiratory events (apnoeas, desaturations and bradycardias), oxygen (O_2) requirements and gas exchange.[3 4] However, the exact thresholds of clinical instability at which infants should be reintubated are not well established. In an international survey, only 10% of neonatal intensive care units (NICUs) reported having guidelines with pre-established criteria for reintubation, and most units relied on clinical judgement.[4] Moreover, while some clinical trials have attempted to define extubation failure based on fulfilling predetermined reintubation criteria,[5 6] these definitions may not reflect everyday practice.

Recently, a longitudinal cohort study described the patterns of reintubation in extremely preterm infants.[7] In that cohort, the most stated reason for reintubation was frequent and/or severe

cardiorespiratory events, followed by increased O_2 and respiratory acidosis. In this secondary analysis, we sought to describe in more detail the exact thresholds of clinical instability used by clinicians to determine the need for reintubation, and aimed to comprehensively evaluate the accuracy of different combinations of criteria in predicting reintubation decisions. We hypothesised that such inclusive evaluation would identify a set of reintubation criteria that would best reflect clinical practice.

METHODS

Context

This was a secondary analysis of a prospective multicentre study that aimed to develop an automated predictor of extubation readiness in extremely preterm infants (Automated Prediction of Extubation Readiness (APEX), ClinicalTrials.gov NCT01909947). The study protocol and main results of APEX have been previously published.[8 9] APEX was conducted in five NICUs between 2013 and 2018 and enrolled infants with birth weight ≤1250 g undergoing their first planned extubation. All infants were extubated to non-invasive respiratory support. For the study, extubation failure was defined based on fulfilling one of the following criteria: (1) O_2 requirements >50% for 2 consecutive hours; (2) partial pressure of carbon dioxide (pCO_2) >55 mm Hg with pH <7.25 on two blood gases; (3) one cardiorespiratory event requiring positive pressure ventilation; (4) ≥6 cardiorespiratory events in 6 hours requiring tactile stimulation. However, those criteria were not mandated for reintubation and all decisions were left to the clinicians' discretion. Study team members recorded whether infants fulfilled extubation failure criteria by prospectively inspecting their hourly O_2, blood gases and hourly occurrence of cardiorespiratory events as documented in the nursing charts for the first 14 days after extubation. In three out of the five participating NICUs, the hourly data were also recorded. Therefore, only infants recruited from those three NICUs were included in this analysis.

Data collection

The following variables were collected hourly from the nursing charts, starting from the time of extubation until reintubation or midnight on the 14th postextubation day, whichever came first: (1) O_2 requirements; (2) type of respiratory support; (3) pH and pCO_2 values, whenever available; (4) presence and number of clinically significant cardiorespiratory events, defined as bradycardias (heart rate ≤100 beats per minute) and/or desaturations (O_2 saturation ≤85%) needing tactile stimulation, supplemental O_2 and/or positive pressure ventilation; (5) lowest heart rate and O_2 saturation values; in cases where the frequency of cardiorespiratory events was charted qualitatively, numerical conversions were established a priori ('occasional' or 'few'=0.5 events/hour; 'multiple', 'frequent', 'clustered', 'many' or 'back-to-back'=2 events/hour). Additionally, the reasons stated by clinicians to proceed with reintubation were noted, and non-respiratory-related diagnoses thought to explain the reintubation (infection, necrotising enterocolitis, elective surgery) were documented.

Data analysis

This study aimed to (1) describe the thresholds of clinical instability used by clinicians for reintubation during the first 14 days after extubation, and (2) evaluate the accuracy of different combinations of criteria in predicting reintubation decisions. Given that infants reintubated in the context of infection, necrotising enterocolitis or elective surgery likely had distinct reasons for reintubation, they were excluded. All analyses were performed using MATLAB (R2018, MathWorks) and GraphPad Prism (9.5.1, GraphPad Software). Birth demographics and periextubation characteristics of reintubated and non-reintubated infants were compared using Wilcoxon rank-sum or χ^2 tests for continuous and categorical variables, respectively. Moreover, the daily median O_2, daily number of cardiorespiratory events and daily median pH and pCO_2 after extubation were compared using repeated measures two-way analysis of variance. A p value <0.05 was considered statistically significant.

For the first objective, thresholds at reintubation pertaining to O_2 needs, gas exchange and cardiorespiratory events were described by averaging the hourly values over the 6 hours preceding reintubation. These thresholds were presented for all reintubated infants and per subgroup of infants specifically stated to be reintubated for increased O_2, respiratory acidosis and frequent/severe cardiorespiratory events. Moreover, reintubation thresholds specific to each participating centre were presented.

For the second objective, reintubation criteria were first generated using different thresholds of clinical instability and grouped into four categories: increased O_2, respiratory acidosis, frequent cardiorespiratory events and severe cardiorespiratory events (defined as desaturations below 60% and/or events requiring positive pressure ventilation). A detailed list of selected thresholds is provided in online supplemental table 1. Next, an automated algorithm was developed to create multiple combinations of reintubation criteria from those four categories using logical OR operators. The percentages of reintubated infants that met criteria (sensitivity) and non-reintubated infants that met criteria (1-specificity) were determined for each combination. The performance of each combination was graphically displayed in a receiver operating characteristic curve, and the accuracy of each combination in predicting reintubation was estimated using the Youden index, which could range from 0 (poor accuracy) to 1 (perfect accuracy).[10] The combination of criteria with the highest Youden index, which maximised inclusion of reintubated infants while minimising inclusion of non-reintubated infants, was described for the overall cohort and for each participating centre.

Post hoc analysis

Using all generated combinations of reintubation criteria, we explored the accuracy of different criteria proposed in the literature at predicting reintubation decisions within this cohort. Criteria were identified from a recent meta-analysis of randomised controlled trials that evaluated different interventions to improve extubation success in preterm infants.[11] The accuracy of the criteria proposed to define extubation failure in the APEX study was also determined.

RESULTS

A total of 266 infants were enrolled in APEX, of which 193 were included for this secondary study (online supplemental figure 1): 55 (28%) were reintubated for respiratory-related reasons and 138 (72%) were not reintubated during the 14 days after extubation. Included infants had a median gestational age of 26.1 weeks (IQR 25–27.6), birth weight of 840 g (IQR 710–1030) and postnatal age at extubation of 8 days (IQR 3–26). Reintubated infants were significantly smaller and more immature at birth and at extubation compared with non-reintubated infants and were extubated from significantly higher O_2 and mean airway pressures (table 1). After extubation, reintubated infants had significantly greater O_2 needs, lower pH, higher pCO_2 and

Table 1 Characteristics of the cohort

Clinical variable	Cohort (n=193)	Not reintubated (n=138)	Reintubated (n=55)	P value
Birth demographics				
Gestational age (weeks)	26.1 [25–27.6]	26.6 [25.2–27.9]	25.2 [24.5–26.1]	<0.001
Birth weight (g)	840 [710–1030]	885 [720–1090]	750 [640-880]	<0.001
Male (%)	98 (51)	67 (49)	31 (56)	0.41
Antenatal steroids (%)	175 (91)	126 (91)	49 (89)	0.77
Caesarean section (%)	132 (68)	99 (72)	33 (60)	0.15
Apgar at 5 min	7 [5–8]	7 [5–8]	6 [5–8]	0.23
Pre-extubation				
Day of life	8 [3–26]	6 [3–27]	14 [5–25]	0.15
PMA (weeks)	28.1 [26.9–29.4]	28.8 [27.6–30]	27.4 [26.6–28.4]	<0.001
Weight (g)	950 [815–1090]	1020 [850–1137]	840 [770–950]	<0.001
Received caffeine (%)	187 (97)	133 (96)	54 (98)	0.67
MAP (cmH_2O)	6.9 [6.2–7.8]	6.7 [6.1–7.5]	7.3 [6.5–8.3]	0.003
FiO_2	0.22 [0.21–0.26]	0.21 [0.21–0.25]	0.25 [0.21–0.28]	<0.001
pH	7.33 [7.29–7.37]	7.33 [7.29–7.37]	7.32 [7.29–7.37]	0.73
pCO_2 (mm Hg)	45 [39–51]	44 [39–49]	46 [40–55]	0.10
Immediate postextubation				
HFNC (%)	7 (4)	7 (5)	0	0.19
CPAP (%)	123 (64)	94 (68)	29 (53)	0.04
NIPPV (%)	63 (33)	37 (27)	26 (47)	0.01

Values are expressed as median [IQR] or number (%) for continuous and categorical variables, respectively.
CPAP, continuous positive airway pressure; FiO_2, fraction of inspired oxygen; HFNC, high flow nasal cannula; MAP, mean airway pressure; NIPPV, nasal intermittent positive pressure ventilation; pCO_2, partial pressure of carbon dioxide; PMA, postmenstrual age.

more frequent cardiorespiratory events compared with non-reintubated infants (figure 1).

Thresholds of clinical instability at reintubation

At reintubation, 52 infants (95%) were on nasal intermittent positive pressure ventilation, with median positive end-expiratory pressure of 7 cmH_2O (IQR 6–8), peak inflation pressure at 17 cmH_2O (IQR 15–18) and rate of 40 inflations/min (IQR 33–50). Median O_2 needs were 35% (IQR 28–44) for all infants and 42% (IQR 36–47) for infants specifically reintubated for increased O_2 (table 2). Median pH and pCO_2 were 7.24 (IQR 7.19–7.3) and 57 mm Hg (IQR 47–68) for all infants, and 7.17 (IQR 7.09–7.23) and 67 mm Hg (IQR 65–82) for infants specifically reintubated for respiratory acidosis. Median number of cardiorespiratory events were 3 (IQR 1–5) for all infants and 3 (IQR 2–5) for infants specifically reintubated for frequent cardiorespiratory events. Site-specific reintubation thresholds are presented in online supplemental table 2.

Reintubation criteria

The proportion of reintubated and non-reintubated infants that fulfilled select clinical instability criteria is shown in table 3. The presence of $O_2>50\%$ for 2 hours, respiratory acidosis ($pH<7.25$ and $pCO_2>55$ mm Hg) and ≥2 events needing positive pressure ventilation were highly specific, being fulfilled by only 2%, 12% and 9% of non-reintubated infants, respectively; however, they lacked sensitivity as they were fulfilled by only 16%, 22% and 22% of reintubated infants, respectively. In contrast, the presence of ≥2 events/hour, ≥4 events in 6 hours and ≥6 events in 6 hours had higher sensitivity (being fulfilled by 62%, 56% and 38% of reintubated infants, respectively) but lacked specificity (being also fulfilled by 75%, 60% and 34% of non-reintubated infants, respectively). After evaluating 123 374 combinations of criteria from the four categories, all had low accuracies in predicting reintubation with Youden indices ranging from 0 to 0.46 (figure 2). The combination with highest accuracy proposed reintubation when $O_2>55\%$ for 1 hour, or $pH<7.2$ with $pCO_2>50$ mm Hg, or ≥2 events needing positive pressure ventilation (sensitivity 64%, specificity 82%, Youden index 0.46). When evaluating each site separately, the combinations of criteria with highest accuracy had Youden indices of 0.48, 0.51 and 0.6, respectively (online supplemental table 3).

Post hoc analysis

Reintubation criteria were reported in 26 randomised controlled trials.[11] Criteria were highly variable across studies using different thresholds of O_2, blood gas values and frequency/severity of cardiorespiratory events to define extubation failure. Youden indices could be determined for 16 trials, ranging from 0.06 to 0.35 (online supplemental table 4). The combination of criteria used to define extubation failure for APEX had a Youden index of 0.14.

DISCUSSION

In this secondary analysis from a large cohort study, specific criteria used for reintubation in extremely preterm infants were described. Thresholds of clinical instability at reintubation were highly variable both within and between participating centres. After evaluating multiple combinations of criteria, none could accurately discriminate between reintubated and non-reintubated infants. Together, these findings highlight the limitations surrounding reintubation decisions in this population.

With increasingly more premature infants exposed to non-invasive respiratory support, the decision to reintubate has become a complex process that carries important consequences. On one hand, a hastened reintubation may expose infants to unnecessary exposure to mechanical ventilation, which is associated with increased risk of long-term morbidities.[1 2] On the

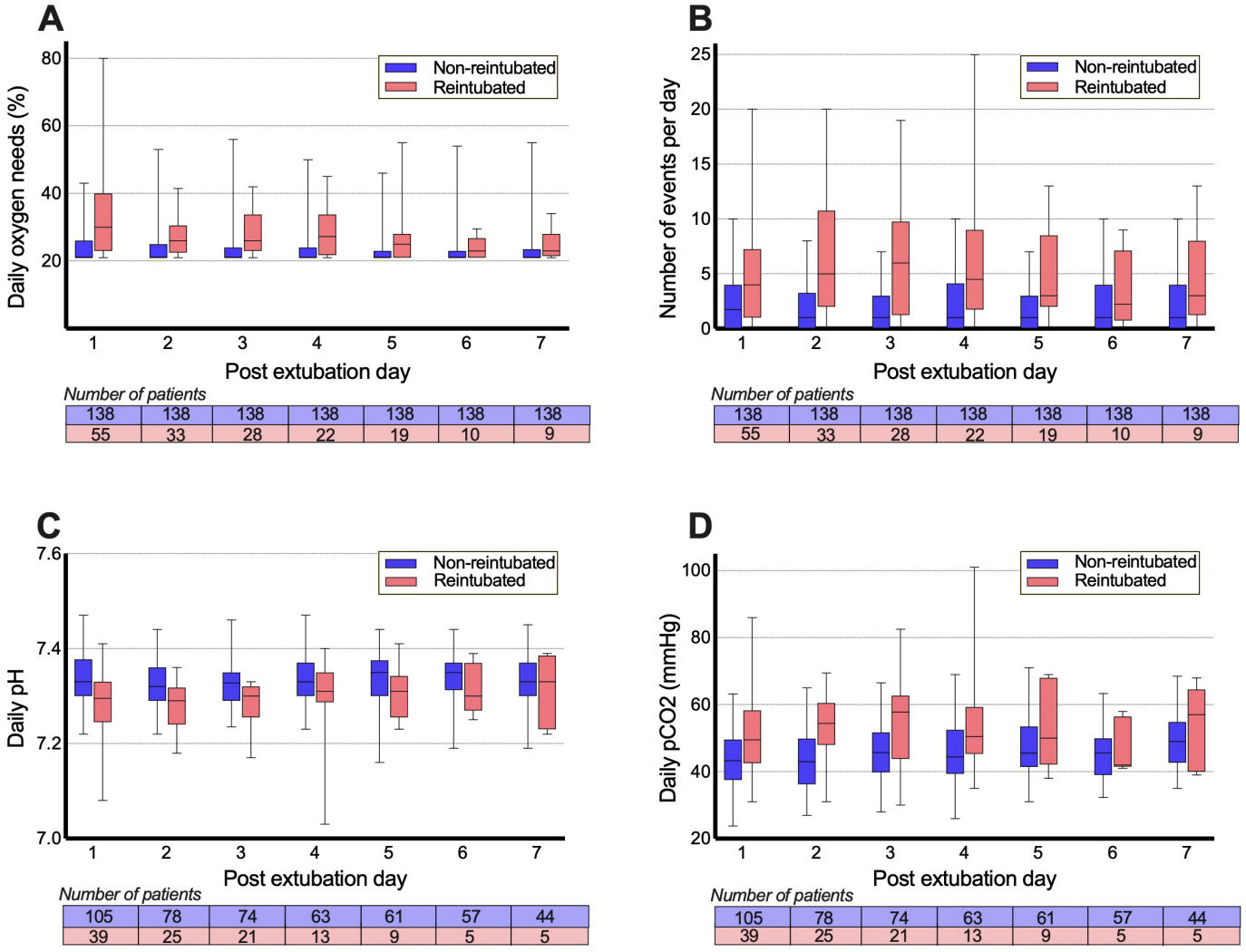


Figure 1 Clinical stability of reintubated and non-reintubated infants after extubation. The figure presents the daily median oxygen needs (A), the daily number of bradycardias/desaturations needing intervention (B), the daily median pH (C) and the daily median pCO_2 (D) among reintubated and non-reintubated infants in the 7 days after extubation. Below each panel, the number of reintubated and non-reintubated patients for which data were available on each postextubation day is shown in the blue and pink rows, respectively. Comparisons between reintubated and non-reintubated infants were performed using repeated measures two-way analysis of variance (ANOVA) and were all statistically significant, with $p \leq 0.05$.

other hand, withholding or delaying reintubation in infants with clinical instability may expose them to unnecessary harm from frequent or prolonged intermittent hypoxia, hyperoxia and/or impaired gas exchange.[12 13] In balancing these risks and benefits, the evidence is scarce regarding the optimal thresholds of clinical instability at which to reintubate extremely preterm infants. As a starting point, we sought to better understand which elements lead clinicians to reintubate in practice and tried to identify a set of reintubation criteria for which there would be agreement in clinical decision-making.

Unfortunately, we found important variations in O_2 needs, pH/pCO_2 values and number of cardiorespiratory events at reintubation, which reaffirms that reintubation decisions are often based on variable degrees of tolerance to clinical instability, both at the individual and unit levels. This variability is further heightened by the absence of unit-specific guidelines for cardiorespiratory event documentation, indications for blood gas sampling and criteria for reintubation. Moreover, it is likely that other unmeasured clinical variables (the infant's physical examination or chest X-ray) and organisational factors (nurse staffing, unit occupancy and presence of trained in-house staff at night) play important roles in those decisions.

Notwithstanding the variable thresholds at reintubation, we evaluated over 120 000 combinations of criteria in the hopes of identifying a set that could discriminate between reintubated and non-reintubated infants. While some thresholds pertaining to increased O_2, respiratory acidosis and severe cardiorespiratory events were highly specific to reintubated infants, there was no threshold pertaining to the frequency of clinically significant cardiorespiratory events that could separate the two groups. In other words, there was poor agreement between clinicians on the number of cardiorespiratory events at which to reintubate. Considering that the most commonly stated reason for reintubation was for frequent cardiorespiratory events, these findings highlight critical shortcomings of cardiorespiratory event monitoring in current practice. First, detection of cardiorespiratory events based on nursing documentation is known to underestimate the true occurrence and nature of events.[14 15] To mitigate this limitation, we only evaluated cardiorespiratory events that required intervention, under the presumption that serious events

Table 2 Clinical instability in the 6 hours preceding reintubation

	All reintubated (n=55)	Subgroups reintubated for:
O_2 needs		Increased O_2 (n=22)
FiO_2 at reintubation	0.35 [0.28–0.44]	0.42 [0.36–0.47]
Additional FiO_2 above baseline	0.09 [0.02–0.18]	0.15 [0.1–0.21]
Gas exchange		Respiratory acidosis (n=8)
pH*	7.24 [7.19–7.3]	7.17 [7.09–7.23]
pCO_2 (mm Hg)*	57 [47–68]	67 [65–82]
Cardiorespiratory events		Frequent/severe events (n=47)
Need for stimulation/O_2 (%)†	36 (68)	35 (76)
Need for PPV (%)†	14 (26)	13 (28)
Total number of events‡	3 [1–5]	3 [2–5]
Total number of events with PPV†	0 [0–1]	0 [0–1]
Lowest desaturation value (%)§	59 [47–65]	59 [45–66]
Lowest bradycardia value (bpm)¶	61 [54–69]	61 [54–69]

Values are expressed as median [QR] or number (%) for continuous and categorical variables, respectively. Cardiorespiratory events were defined as bradycardias and/or desaturations requiring stimulation, oxygen supplementation and/or positive pressure ventilation.
*Data available for 23 out of 55 reintubated infants and 7 out of 8 infants specifically reintubated for respiratory acidosis.
†Data available for 54 out of 55 reintubated infants and 46 out of 47 infants specifically reintubated for frequent/severe events.
‡Data available for 53 out of 55 reintubated infants and 45 out of 47 infants specifically reintubated for frequent/severe events.
§Data were available from 36 out of 55 reintubated infants and 33 out of 47 infants specifically reintubated for frequent/severe events.
¶Data were available from 28 out of 55 reintubated infants and 28 out of 47 infants specifically reintubated for frequent/severe events.
bpm, beats per minute; FiO_2, fraction of inspired oxygen; pCO_2, partial pressure of carbon dioxide; PPV, positive pressure ventilation.

Table 3 Proportion of reintubated and non-reintubated infants that fulfilled select clinical instability criteria

Clinical event category	Reintubated infants that met criteria (n=55)	Non-reintubated infants that met criteria (n=138)
Increased O_2 needs		
FiO_2>0.3 for 1 hour	47 (85)	80 (58)
FiO_2>0.4 for 1 hour	32 (58)	35 (25)
FiO_2>0.5 for 1 hour	20 (36)	17 (12)
FiO_2>0.6 for 1 hour	10 (18)	4 (3)
FiO_2>0.4 for 2 hours	23 (42)	18 (13)
FiO_2>0.5 for 2 hours	9 (16)	3 (2)
Respiratory acidosis		
pH<7.3 and pCO_2>60	13 (24)	20 (15)
pH<7.25 and pCO_2>55	12 (22)	17 (12)
pH<7.2 and pCO_2>65	6 (11)	4 (3)
Frequent events		
≥2 events/hour	34 (62)	104 (75)
≥4 events in 6 hours	31 (56)	83 (60)
≥6 events in 6 hours	21 (38)	47 (34)
≥8 events in 6 hours	12 (22)	28 (20)
≥6 events in 24 hours	33 (60)	87 (63)
Events needing PPV		
≥1 event in 24 hours	18 (33)	39 (28)
≥2 events in 24 hours	12 (22)	12 (9)

Values are represented as number (%). Cardiorespiratory events were defined as bradycardias and/or desaturations requiring stimulation, oxygen supplementation and/or positive pressure ventilation.
FiO_2, fraction of inspired oxygen; pCO_2, partial pressure of carbon dioxide; PPV, positive pressure ventilation.

would be documented more reliably than self-recovering events. However, some of those events may still have been under-reported or misreported, especially when they were clustered, or during the acute period immediately preceding reintubation. Second, even if the number of cardiorespiratory events could be better quantified, the long-term effects of apnoeas, bradycardias and desaturations of varying frequencies, durations and severities on the preterm lung and brain remain unknown. For instance, while studies have shown an association between intermittent hypoxias and retinopathy of prematurity,[16 17] bronchopulmonary dysplasia[1 18 19] and neurodevelopmental impairment,[20 21] it is unclear which patterns of intermittent hypoxias translate into compromised regional tissue oxygenation and poor outcomes.[22] Thus, in the absence of bedside tools for quantitatively detecting cardiorespiratory events and objectively assessing their clinical significance, it appears justifiably difficult for clinicians to guide their decisions about reintubation.

The variability and low predictability of reintubation decisions also call into question the results of randomised controlled trials evaluating different interventions to improve chances of extubation success in extremely preterm infants. Indeed, we observed significant variations across trials in the proposed criteria to define extubation failure, with no combination of criteria accurately discriminating between reintubated and non-reintubated infants when applied to the APEX cohort. These findings further reinforce the subjective nature of the outcome of 'reintubation' or 'extubation failure' when evaluating different respiratory care interventions in preterm infants.

The study had some limitations. First, information pertaining to hourly O_2 needs and cardiorespiratory event occurrence was imprecise, as it was abstracted from subjective and non-standardised nursing documentation. Given that nursing charts are still the primary documentation source in most NICUs, this limitation may add external validity by providing a snapshot of real-world practice. While direct data extraction from the bedside monitor could have provided a more precise account of the number and duration of cardiorespiratory events, this would have possibly come at the expense of including more self-limited events, for which the clinical impact is unknown.[21] Second, in spite of the significant intercentre variations observed in this study, data were derived from three NICUs with a similar university affiliation, which suggests the possibility of relatively more homogeneous practices than would be the case if participating units were completely unrelated. Third, the different clinical approaches used by participating centres lead to inconsistent blood gas sampling prior to reintubation. Fourth, several variables that could potentially influence reintubation decisions were not evaluated, including the infant's work of breathing, chest X-ray findings, presence of haemodynamically significant patent ductus arteriosus and feeding intolerance. Fifth, our results may not be generalisable to bigger or more mature neonates and may not apply to infants after an accidental extubation or beyond the first attempt.

In conclusion, criteria used for reintubation in extremely preterm infants remain highly variable, with no combination accurately predicting the decision to reintubate. These results appear primarily attributable to the suboptimal monitoring and

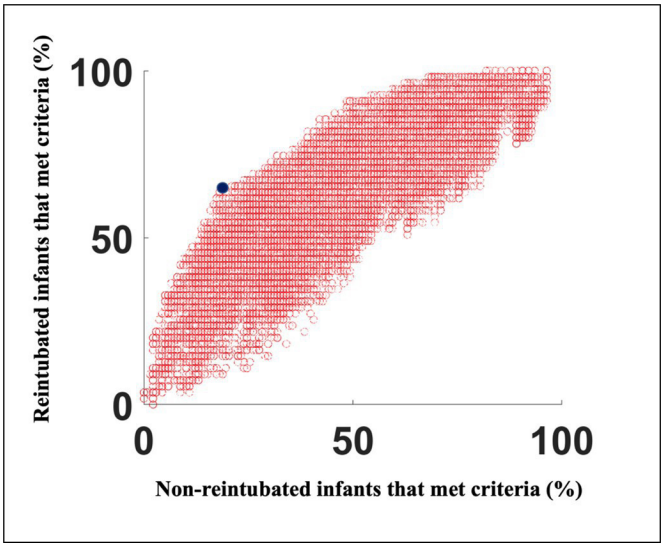


Figure 2 Accuracy of multiple combinations of reintubation criteria in predicting reintubation. This figure presents the accuracy of 123 374 combinations of reintubation criteria from four clinical event categories in discriminating between reintubated and non-reintubated infants. The four clinical event categories included different thresholds for (1) increased oxygen requirements, (2) respiratory acidosis, (3) frequent bradycardias/desaturations requiring intervention (stimulation, supplemental oxygen and/or positive pressure ventilation) and (4) profound desaturations (desaturations below 60%) or bradycardias/desaturations needing positive pressure ventilation. Each combination of reintubation criteria is represented by one red circle. The filled blue circle represents the combination of reintubation criteria with the highest balanced accuracy.

characterisation of cardiorespiratory events in clinical practice. Until such bedside tools are developed and validated, efforts should focus on standardising postextubation monitoring and management to reduce practice variations related to reintubation. This could be achieved through implementation of unit-specific guidelines for non-invasive respiratory support and standardised approaches to documentation and management of cardiorespiratory events.

Acknowledgements We are grateful to all the families who agreed to participate in the APEX study and the health professionals who were involved in the care of participants.

Contributors The research question for this study was conceived by WS and developed by WS, TAM and GMS'A. TAM undertook the data extraction. WS and TAM undertook the data analysis. WS and TAM drafted the first draft of the manuscript. All authors were involved in interpretation, review and revision of the draft manuscript and approval of the final version. WS had full access to all the data in the study and take full responsibility for the integrity of the data and accuracy of the data analysis. WS is the guarantor of the study.

Funding This project received funding via an operational grant from the Canadian Institutes of Health Research.

Disclaimer The funding body did not have a role in the design and collection, analysis or interpretation of the data.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and the institutional review board at each participating institution provided approval for the original APEX study. The principal site's ethics approval was granted by the McGill University Health Centre's Research Ethics Board (MP-37-2013-510, 12-387-PED, eReviews_3030, MP-37-12-387-PED). For the present study, institutional review board approval was waived because the analysis involved no additional data collection beyond that which was collected at for APEX. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

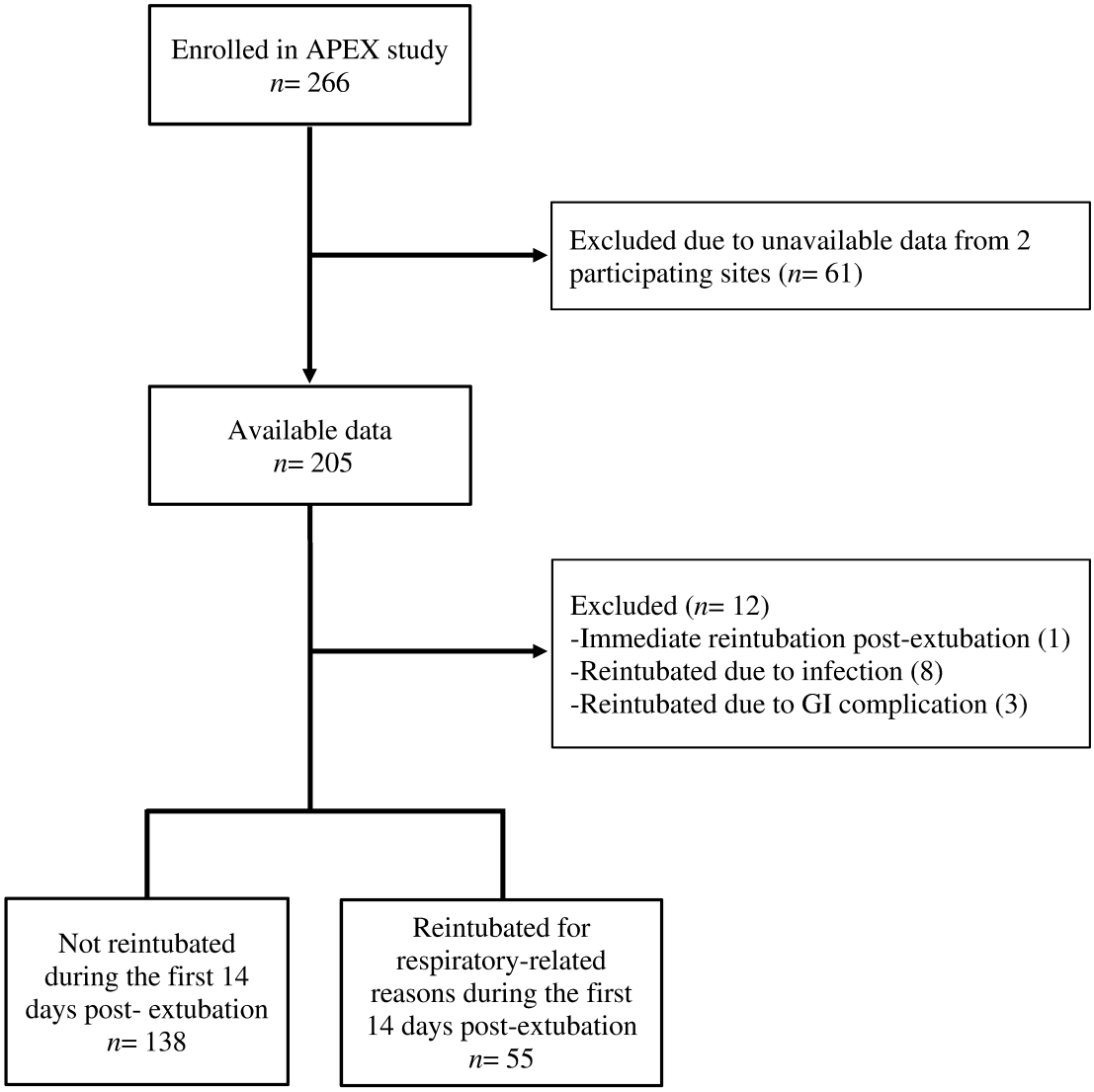
Data availability statement Data are available upon reasonable request. All data relevant to the study are included in the article or uploaded as supplementary information.

ORCID iDs
Samantha Latremouille http://orcid.org/0000-0002-1504-4271
Guilherme M Sant'Anna http://orcid.org/0000-0003-1633-3654
Wissam Shalish http://orcid.org/0000-0002-7435-1466

REFERENCES

1 Jensen EA, DeMauro SB, Kornhauser M, *et al*. Effects of multiple ventilation courses and duration of mechanical ventilation on respiratory outcomes in extremely low-birth-weight infants. *JAMA Pediatr* 2015;169:1011–7.
2 Miller JD, Carlo WA. Pulmonary complications of mechanical ventilation in neonates. *Clin Perinatol* 2008;35:273–81.
3 Beltempo M, Isayama T, Vento M, *et al*. Respiratory management of extremely preterm infants: an international survey. *Neonatology* 2018;114:28–36.
4 Al-Mandari H, Shalish W, Dempsey E, *et al*. International survey on periextubation practices in extremely preterm infants. *Arch Dis Child Fetal Neonatal Ed* 2015;100:F428–31.
5 Finer N, Carlo WA. Early CPAP versus surfactant in extremely preterm infants. *N Engl J Med* 2010;362:1970–9.
6 Morley CJ, Davis PG, Doyle LW, *et al*. Nasal CPAP or intubation at birth for very preterm infants. *N Engl J Med* 2008;358:700–8.
7 Shalish W, Kanbar L, Keszler M, *et al*. Patterns of reintubation in extremely preterm infants: a longitudinal cohort study. *Pediatr Res* 2018;83:969–75.
8 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;17:167:167.:.
9 Kanbar LJ, Shalish W, Onu CC, *et al*. Automated prediction of extubation success in extremely preterm infants: the apex multicenter study. *Pediatr Res* 2023;93:1041–9.
10 Hughes G. Youden's index and the weight of evidence. *Methods Inf Med* 2015;54:198–9.
11 Ferguson KN, Roberts CT, Manley BJ, *et al*. Interventions to improve rates of successful extubation in preterm infants: a systematic review and meta-analysis. *JAMA Pediatr* 2017;171:165–74.
12 Di Fiore JM, MacFarlane PM, Martin RJ. Intermittent hypoxemia in preterm infants. *Clin Perinatol* 2019;46:553–65.
13 Di Fiore JM, Poets CF, Gauda E, *et al*. Cardiorespiratory events in preterm infants: interventions and consequences. *J Perinatol* 2016;36:251–8.
14 Vergales BD, Paget-Brown AO, Lee H, *et al*. Accurate automated apnea analysis in preterm infants. *Am J Perinatol* 2014;31:157–62.
15 Brockmann PE, Wiechers C, Pantalitschka T, *et al*. Under-Recognition of alarms in a neonatal intensive care unit. *Arch Dis Child Fetal Neonatal Ed* 2013;98:F524–7.
16 Di Fiore JM, Bloom JN, Orge F, *et al*. A higher incidence of intermittent hypoxemic episodes is associated with severe retinopathy of prematurity. *J Pediatr* 2010;157:69–73.
17 Di Fiore JM, Kaffashi F, Loparo K, *et al*. The relationship between patterns of intermittent hypoxia and retinopathy of prematurity in preterm infants. *Pediatr Res* 2012;72:606–12.
18 Raffay TM, Dylag AM, Sattar A, *et al*. Neonatal intermittent hypoxemia events are associated with diagnosis of bronchopulmonary dysplasia at 36 weeks postmenstrual age. *Pediatr Res* 2019;85:318–23.
19 Fairchild KD, Nagraj VP, Sullivan BA, *et al*. Oxygen desaturations in the early neonatal period predict development of bronchopulmonary dysplasia. *Pediatr Res* 2019;85:987–93.
20 Di Fiore JM, Raffay TM. The relationship between intermittent hypoxemia events and neural outcomes in neonates. *Exp Neurol* 2021;342:113753.
21 Poets CF, Roberts RS, Schmidt B, *et al*. Association between intermittent hypoxemia or bradycardia and late death or disability in extremely preterm infants. *JAMA* 2015;314:595–603.
22 Di Fiore JM, Martin RJ, Raffay TM. Intermittent hypoxemia and bronchopulmonary dysplasia: manifestations of immature respiratory control and the preterm lung. *Am J Respir Crit Care Med* 2021;204:1126–7.

Supplemental Figure 1. Flow of participants



Abbreviations: APEX – Automated Predictor of extubation readiness, GI – gastrointestinal

Supplemental Table 1. Thresholds used for the development and evaluation of reintubation criteria

Clinical event categories	Range of thresholds used
Increased Oxygen needs (24 reintubation criteria)	FiO_2 > 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, or 0.6 for 1hr FiO_2 > 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, or 0.6 for 2 hrs Additional FiO_2 > 0.05, 0.1, 0.15, or 0.2 for 1 hr Additional FiO_2 > 0.05, 0.1, 0.15, or 0.2 for 2 hrs
Respiratory acidosis (46 reintubation criteria)	pH < 7.2, 7.25 or 7.3 pCO_2 > 50, 55, 60, 65, or 70 pH < 7.2, 7.25, or 7.3 with pCO_2 > 50, 55, 60, 65, or 70 mmHg pH < 7.2, 7.25 or 7.3 in 2 consecutive blood gases CO_2 > 50, 55, 60, 65 or 70 mmHg in 2 consecutive blood gases pH < 7.2, 7.25, or 7.3 with pCO_2 > 50, 55, 60, 65, or 70 mmHg in 2 consecutive blood gases
Frequent events (14 reintubation criteria)	> 1, 2, 3, or 4 events per hour in 6 hrs ≥ 2, 4, 6, 8, 10, or 12 events in 6 hrs > 5, 10, 15, or 20 events in 24 hrs
Severe events (6 reintubation criteria)	> 1, ≥ 1, or ≥ 2 events needing PPV in 6 hrs > 1, 2, or 3 desaturations below 60% in 6 hrs

Cardiorespiratory events were defined as bradycardias and/or desaturations requiring stimulation, oxygen supplementation, and/or positive pressure ventilation.
Abbreviations: FiO_2 – fraction of inspired oxygen, PPV – positive pressure ventilation.

Supplemental Table 2. Clinical instability in the 6 hours preceding reintubation in three participating study centers

	Center 1 ***(n=24)***	**Center 2** ***(n=24)***	**Center 3** ***(n=7)***
O_2 needs			
FiO_2 at reintubation	0.33 [0.28-0.38]	0.43 [0.3-0.47]	0.29 [0.25-0.41]
Additional FiO_2 above baseline	0.08 [0.02-0.15]	0.13 [0.04-0.21]	0.03 [0-0.18]
Gas exchange			
pH	7.21 [7.13-7.27]	7.27 [7.2-7.3]	7.26 [7.22-7.33]
pCO_2, mmHg	65 [56-87]	55 [44-66]	55 [44-57]
Cardiorespiratory events			
Need for Stimulation/O_2, %	13 (56)	20 (83)	3 (43)
Need for PPV, %	4 (17)	9 (38)	1 (14)
Number of events needing intervention	2 [0-3]	3 [2-7]	4 [1-6]
Number of events needing PPV	0 [0-0]	0 [0-1]	0 [0-0]
Lowest desaturation value, %	59 [47-67]	59 [42-66]	54 [47-61]
Lowest bradycardia value, bpm	59 [50-66]	61 [53-68]	67 [61-72]

Values are expressed as median [interquartile range] or number (%) for continuous and categorical variables, respectively. Cardiorespiratory events were defined as bradycardias and/or desaturations requiring stimulation, oxygen supplementation, and/or positive pressure ventilation. Abbreviations: FiO_2 – fraction of inspired oxygen, PPV – positive pressure ventilation, bpm – beats per minute.

Supplemental Table 3. Reintubation criteria with the highest accuracies

Criteria for Reintubation	% Reintubated infants that met criteria	% Non-Reintubated infants that met criteria	Youden index
Highest accuracy – overall cohort (1) FiO2 > 0.55 for 1h; or (2) pH < 7.2 & pCO_2 > 50 mmHg; or (3) ≥ 2 events needing PPV in 24h.	64	18	0.46
Highest accuracy – study center 1 (1) FiO2 > 0.35 for 1h; or (2) pH < 7.2 & pCO_2 > 50 mmHg.	92	32	0.6
Highest accuracy – study center 2 (1) FiO2 > 0.5 for 2h; or (2) pH < 7.2 & pCO_2 > 50 mmHg; or (3) ≥ 12 events in 6h; or (4) ≥ 2 events needing PPV in 24h.	71	20	0.51
Highest accuracy – study center 3 (1) FiO2 > 0.35 for 1h; or (2) pH < 7.2 & pCO_2 > 50 mmHg; or (3) ≥ 4 events in 1h; or (4) ≥ 1 event needing PPV.	86	37	0.48

Values are represented as percentages. Cardiorespiratory events were defined as bradycardias and/or desaturations requiring stimulation, oxygen supplementation, and/or positive pressure ventilation.
Abbreviations: FiO_2 – fraction of inspired oxygen, PPV – positive pressure ventilation.

Supplemental Table 4. Accuracy of different reintubation criteria previously proposed in the literature in predicting reintubation decisions

Criteria for Reintubation	% Reintubated that met criteria	% Non-Reintubated that met criteria	Youden index
APEX proposed criteria			
Kanbar, 2022 (1) FiO_2 > 0.5 for 2h; or (2) pH < 7.25 & pCO_2 > 55; or (3) ≥ 6 events in 6h; or (4) ≥ 1 event needing PPV.	51	37	0.14
Criteria proposed in the literature			
Engelke, 1982 (1) Additional O_2 needs > 15%; or (2) pH < 7.2 & pCO_2 > 60.	64	28	0.35
Higgins, 1991 (1) FiO_2 > 0.6; or (2) pH ≤ 7.23 & pCO_2 > 60; or (3) ≥ 3 events in 1h; or (4) ≥ 1 event needing PPV.	73	49	0.24
Chan, 1993 (1) FiO_2 > 0.6; or (2) pH ≤ 7.25 & pCO_2 > 50; or (3) Recurrent minor events;* or (4) ≥ 1 major event.	N/A	N/A	N/A
Annibale, 1994 (1) FiO_2 ≥ 0.8; or (2) pH ≤ 7.20; or (3) ≥ 6 events in 1h; or (4) ≥ 1 event needing PPV.	62	33	0.29
So, 1995 (1) FiO_2 ≥ 0.7; or (2) pH ≤ 7.25 & pCO_2 > 60; or (3) ≥ 3 events in 1h; or (4) ≥ 1 event needing PPV.	75	51	0.24

Davis, 1998 and 2001 (1) Additional O_2 needs > 15%; or (2) pH < 7.25 & pCO_2 > 50; or (3) > 6 events in 6 hours; or (4) ≥ 1 event needing PPV.	85	60	0.25
Friedlich, 1999 (1) Additional O_2 needs > 15%; or (2) pH ≤ 7.25 x 2; or (3) pCO_2 ≥ 25% above baseline;* or (4) ≥ 1 event needing PPV.	N/A	N/A	N/A
Dimitriou, 2000 (1) FiO_2 > 0.6; or (2) pH < 7.25; or (3) Recurrent minor events;* or (4) ≥ 1 major event.	N/A	N/A	N/A
Barrington, 2001 (1) FiO_2 > 0.7; or (2) pCO_2 > 70; or (3) > 6 events in 24h; or (4) > 2 events needing PPV in 24h.	73	67	0.06
Khalaf, 2001 (1) pH < 7.25 & pCO_2 > 60; or (2) > 2 events in 1h; or (3) ≥ 1 event needing PPV.	64	49	0.15
Stefanescu, 2003 (1) FiO_2 ≥ 0.5; or (2) pH < 7.25 & pCO_2 ≥ 70; or (3) recurrent significant events.*	N/A	N/A	N/A
Peake, 2005 (1) FiO_2 ≥ 0.7; or (2) pH < 7.25 & pCO_2 > 55 x 2; or (3) ≥ 3 events; or (4) ≥ 1 event needing PPV.	44	35	0.09
Campbell, 2006 (1) FiO_2 > 0.6; or (2) pH < 7.25; or (3) ≥ 3 events in 1h; or (4) ≥ 1 event needing PPV.	76	51	0.25

Moretti, 2008 (1) $FiO_2 \geq 0.7$; or (2) pH < 7.25 & pCO_2 > 70; or (3) Recurrent events needing PPV.*	N/A	N/A	N/A
Gupta, 2009 (1) pH < 7.2 & pCO_2 > 60; or (2) ≥ 1 event needing PPV.	47	30	0.17
Miller, 2010 (1) FiO_2 persistently > 0.7; or (2) pH < 7.25 & pCO_2 > 65; or (3) > 3 events in 12 hours;* or (4) > 2 major events in 8 hours.*	N/A	N/A	N/A
Yadav, 2012 (1) $FiO_2 \geq 0.7$; or (2) pH < 7.2 & pCO_2 > 60; or (3) Recurrent events needing PPV.*	N/A	N/A	N/A
Kumar, 2011 (1) pH < 7.25 & pCO_2 > 60; or (2) > 3 events in 1h; or (3) ≥ 1 event needing PPV.	53	33	0.19
O'Brien, 2012 (1) FiO_2 > 0.6; or (2) pH < 7.25; or (3) ≥ 4 events in 1h; or (3) ≥ 1 event needing PPV.	69	38	0.31
Collins, 2013 (1) Additional O_2 needs > 15%; or (2) pH < 7.25 & pCO_2 > 66; or (3) > 6 events in 6 hours; or (4) ≥ 1 event needing PPV.	82	59	0.23
Kahramener, 2013 (1) FiO_2 > 0.6; or (2) pH < 7.25 & pCO_2 > 60; or (3) Frequent or severe events.*	N/A	N/A	N/A
Kirpalani, 2013 (1) > 6 events in 6 h; or (2) > 1 event needing PPV.	49	38	0.11

Manley, 2013 (1) Additional O_2 needs ≥ 20%; or (2) pH < 7.2 & pCO_2 > 60; or (3) ≥ 6 events in 6 hours; or (4) > 1 event needing PPV in 24h.	80	46	0.34
Yoder, 2013 (1) FiO_2 > 0.6; or (2) pCO_2 > 65; or (3) Frequent or severe events;* or (4) > 1 event needing PPV in 12h; or (5) Persistent severe retractions;* or (6) Base excess > -10.*	N/A	N/A	N/A
Buzzella, 2014 (1) FiO2 > 0.6 with increased O_2 needs ≥ 20% above baseline;* or (2) pCO_2 ≥ 65 x 2 with an increase > 15 above baseline;* or (3) Repeated events.*	N/A	N/A	N/A

Legend: The table presents the accuracy of different reintubation criteria proposed in the literature in predicting the need for reintubation amongst infants enrolled in the Automated Prediction of Extubation readiness (APEX) study. Criteria were identified from a recent systematic review of all randomized controlled trials that evaluated interventions to improve extubation success rates in preterm infants (Ferguson KN, Roberts CT, Manley BJ, Davis PG. Interventions to Improve Rates of Successful Extubation in Preterm Infants: A Systematic Review and Meta-analysis. *JAMA Pediatr* 2017;171(2):165-174). For the purpose of this table, criteria were systematically arranged in the following sequence of categories of clinical instability: (1) increased oxygen requirements, (2) respiratory acidosis, (3) frequent cardiorespiratory events (referring to any apneas, bradycardias or desaturations requiring some form of tactile stimulation, oxygen supplementation, or positive pressure ventilation), (4) major cardiorespiratory event(s) requiring positive pressure ventilation or cardiopulmonary resuscitation, and (5) any other criteria that did not fit the above categories. The accuracy of each proposed set of reintubation criteria was presented if it matched one of the 123,374 combinations of criteria evaluated in this manuscript. Criteria that did not match any of the evaluated thresholds of clinical instability are marked by an asterisk. The Youden index was computed as the percentage of reintubated infants that fulfilled the criteria minus the percentage of non-reintubated infants that fulfilled the criteria, and ranges from 0 (poor accuracy) to 1 (perfect accuracy).

Abbreviations: APEX – Automated prediction of extubation readiness, FiO_2 – fraction of inspired oxygen, PPV – positive pressure ventilation.