Intensive Care Med (2025) 51:106–114
https://doi.org/10.1007/s00134-024-07758-0

ORIGINAL

Best clinical model predicting extubation failure: a diagnostic accuracy post hoc analysis

Patricia Rodríguez Villamizar[1], Arnaud W. Thille[2,3], Margarita Márquez Doblas[1], Jean-Pierre Frat[2,3], Pilar Leal Sanz[1], Elena Alonso[1], Victoria País[4], Guillermo Morales[1], Laura Colinas[1], Alicia Propín[5], Aida Fernández Olivares[6], María Martínez Balaguer[8], Diego Alvaredo Rodrigo[8] and Gonzalo Hernández[1,7,9,10*]

Abstract

Purpose: Predicting extubation failure remains a clinical challenge. This study aimed to determine diagnostic accuracy of models used at the bed side.

Methods: Post hoc analysis of 2341 patients at all risk included in five multicenter randomized trials. Diagnostic accuracy of three clinical prediction models was compared: 3-factors model including age > 65y, chronic heart or pulmonary disease; 4-factors model adding prolonged mechanical ventilation; and 11-factors model including age > 65 years, ≥ 2 comorbidities, prolonged mechanical ventilation, acute heart failure as the primary indication for mechanical ventilation, moderate-to-severe chronic obstructive pulmonary disease, APACHE II score > 12 on extubation day, airway patency problems, inability to deal with respiratory secretions, not simple weaning, obesity, or hypercapnia at the end of the spontaneous breathing trial. Crude and adjusted for spontaneous breathing trial (SBT) models were compared for all-cause reintubation at 7 days using Youden and Kappa indexes.

Results: The 3-factors model had a very low global prediction capability (Youden index 0.08 and Kappa index 0.04); the 4-factors and 11-factors models had low global prediction capability (Youden index 0.12 and 0.16, and Kappa index 0.06 and 0.07, respectively). Aggressive SBT strategies (pressure support ≥ 7 cm H_2O with or without positive end-expiratory pressure) were associated with extubation failure risk ($p < 0.001$). All adjusted models had low diagnostic capability (0.08/0.03, 0.07/0.03, and 0.06/0.02 respectively).

Conclusion: Based on these results, the 3-factors model reported a very low diagnostic accuracy, and the 4 or 11-factors models showed similar low accuracy. No improvement was observed after adjusting for other aspects of weaning.

Keywords: Weaning, Reintubation, Prediction, Model, Extubation failure, Outcome

Introduction

Predicting extubation failure remains a clinical challenge. Many models have been proposed, but their clinical applicability is limited [1–6]. Reintubation rates reported in published trials allow to calculate positive predictive values, but none of the remaining diagnostic performance parameters (e.g. sensitivity, specificity, and negative predictive value). Comparison of those models or their prospective validation are pending, thus a final model is lacking.

There are many reasons explaining this limited acceptance of a definitive model: first, a large amount of risk factors modify extubation outcome and a complete model is limited at the bed side (e.g. time consuming); second, many risk factors lack strong evidence, were described before the post-extubation respiratory support era or have conflicting results depending on the preventive therapy applied (e.g. obesity, acute or chronic

*Correspondence: ghernandezm@telefonica.net
[1] Critical Care Medicine, Complejo Hospitalario Universitario de Toledo, Toledo, Spain
Full author information is available at the end of the article

Springer

heart failure) [1, 3, 7–9]; third, some risk factors, mainly those related to airway failure lack an objective definition [2, 10]; fourth, clinical management during the acute phase of critical illness is strongly related to weaning (e.g. sedation practices), and many risk factors are treatable markers of clinical conditions associated to extubation outcome (e.g., physiological parameters during separation attempts, level of consciousness), making weaning highly dependent on local clinical practices [1, 3]; fifth, extubation failure is also related to other aspects of weaning (e.g. screening, confirmatory test, and preventive therapy), but they have never been included in prediction models before [11].

Trials on weaning have used models designed for clinical decision making at the bed side, according to two different approaches: a pragmatic 3-factors model including age and most relevant comorbidities (e.g. chronic lung and heart diseases) [12], with the threat of underrating, and a detailed and time consuming 11-factors model, including also characteristics of acute phase of critical illness (e.g. time on ventilator) and the clinical condition at the time of extubation (e.g. risk factors for airway failure) [11, 13–15], with the threat of overrating.

A benefit–risk analysis to select optimal model should also include the clinical consequences of classification errors. Overrating is not associated to meaningful adverse clinical consequences, as it could associated to NIV-based strategy overuse and there is no proven risk with the use of post-extubation preventive NIV [11–15], but underrating could lead to inadequate selection of preventive therapy, with the risk of an increased reintubation rate [13, 15].

The most prevalent risk factors are comorbidities, age, not simple weaning, and prolonged mechanical ventilation, respectively [14]. Not simple weaning is more difficult to define [3, 16, 17] and prolonged mechanical ventilation better detects ICU-acquired weakness [18]. A model defining high risk based on four factors (e.g. age, chronic lung or heart diseases, and prolonged mechanical ventilation), could potentially detect most patients detected by an 11-factors model.

Thus, this study aimed to calculate and compare the diagnostic accuracy of three clinical models predicting extubation failure (e.g. three, four and 11 factors) by a post hoc analysis of five randomized trials including a mixed at-risk population, with a reduced treatment bias secondary to balanced spontaneous breathing trials and post-extubation preventive therapies. The flowchart of the study is presented in Fig 1.

Take-home message

Optimal clinical model predicting extubation failure remains unknown: large clinical trials have used a pragmatic 3-factors or a time consuming 11-factors model; and other aspects of weaning (e.g., screening, spontaneous breathing trial, post-extubation therapy) have never been considered as confounders in models predicting reintubation.

This large post-hoc study including patients with a wide range of risk, showed limited diagnostic capability of all the clinical models used for clinical decision making at the bed side. No meaningful improvement was observed after adjusting for aggressive spontaneous breathing trial (pressure support ≥ 7 cm H_2O with or without positive end-expiratory pressure). A 4 factors model including age, chronic heart or lung disease, and prolonged mechanical ventilation could be used at the bed side to predict extubation failure and plan post-extubation respiratory therapy.

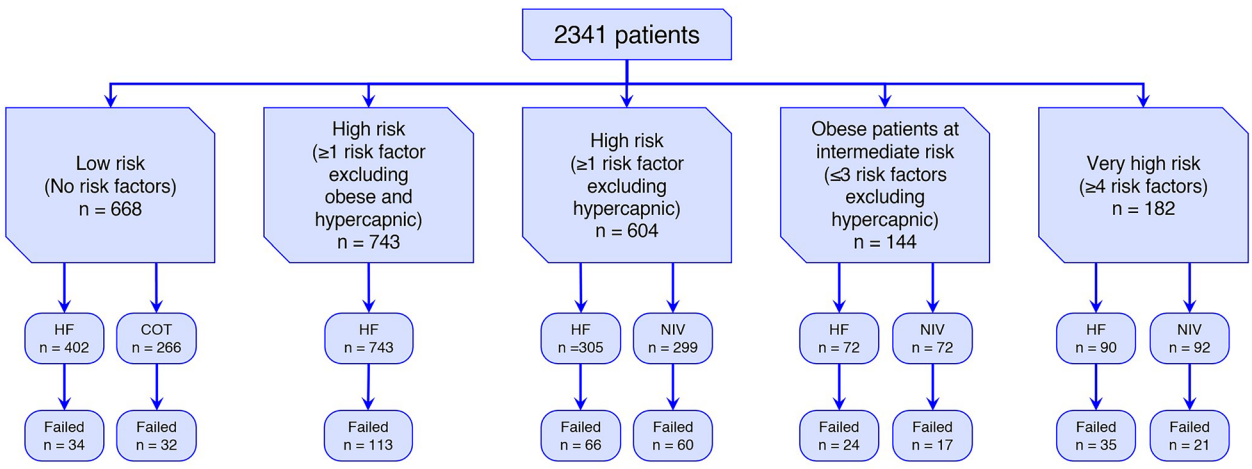

Fig. 1 Flowchart of the patients included in the five randomized clinical trials. *COT* conventional oxygen therapy; *HF* high flow oxygen therapy; *NIV* noninvasive ventilation

Methods

The five multicenter randomized trials used for this post-hoc analysis were conducted in different intensive care units (ICU) in Spain and published in 2016 [14, 15], 2022 [12], and 2024 [11, 19]. The ethics committee at each hospital and the departments of health of the regional governments with which these hospitals are affiliated (Madrid, Castilla–la Mancha, Catalonia, and Illes Balears) approved the study protocols. All patients or their relatives provided written informed consent. The current study was performed according to the STARD guidelines [20].

Patients

The entire population analyzed in the current study included 2341 patients selected using the 11-factors model: 527 low risk patients randomized to preventive conventional oxygen therapy (COT) vs HFNC (ClinicalTrials.gov ID: NCT01191489) [14]; 604 high risk patients, excluding hypercapnia at the end of the spontaneous breathing trial (SBT) randomized to HFNC vs NIV with a non-inferiority design (ClinicalTrials.gov ID: NCT01191489) [15]; 182 patients at very high risk (≥ 4 risk factors) randomized to humidified NIV vs HFNC (ClinicalTrials.gov ID: NCT04125342) [12]; 144 obese at intermediate risk patients (≤ 3 risk factors excluding hypercapnia at the end of the SBT; ClinicalTrials.gov ID: NCT04125342) [19]; and 940 patients at low or intermediate risk (≤ 3 risk factors excluding hypercapnia at the end of the SBT and obesity; ClinicalTrials.gov ID: NCT04758546) with planned HFNC and randomized to aggressive vs conservative screening and spontaneous breathing trial (SBT) [11].

Prediction models

High risk was defined as fulfilling at least one of the risk factors for any model. Three prediction models were analyzed: (1) 3-factors model predicting high risk for reintubation in any patient fulfilling at least one of the following: age > 65y, chronic heart or pulmonary disease defined according to the Charlson Comorbidity index [13]; (2) 4-factors model considering high risk for reintubation any patient fulfilling any of the previous 3-factors model criteria or prolonged mechanical ventilation ≥ 7 days; and (3) 11-factors model, including as a high risk factor any of the following: age > 65 years, acute heart failure as the primary indication for mechanical ventilation, moderate-to-severe chronic obstructive pulmonary disease, ≥ 2 comorbidities defined according to the Charlson Comorbidity index, Acute Physiology and Chronic Health Evaluation II (APACHE II) score > 12 on extubation day, airway patency problems, inability to deal with respiratory secretions, difficult or prolonged weaning (failing the first SBT), length of mechanical ventilation ≥ 7 days, a body mass index > 30 kg/m^2, or hypercapnia at the end of the SBT [14, 15].

Outcomes

The primary outcome analyzed in the current study was all-cause reintubation rate, defined within 7 days after extubation. Two original studies reported reintubation rates at 72 h because the prevention time was limited to 24 h [13, 14], adding up to 403 reintubated patients reported in the original studies. For the current analysis with have included some additional patients reintubated within 7 days as reported in the supplementary material from those trials.

Reasons for reintubation were classified as non-respiratory related, airway failure or weaning failure [2]. Non-respiratory-related reintubations were considered those needed for surgical interventions or neurological deterioration; airway failure related reintubations were considered those recorded from the original studies as inability to clear secretions and persistent postextubation respiratory failure secondary to copious secretions that could not be adequately cleared or that were associated in conscious patients with changes in mental status, acidosis, hypoxemia, and persistent or worsening signs of respiratory-muscle fatigue [13, 14].

Statistical analysis

Descriptive analyses

An initial univariate descriptive analysis of all variables was performed. Categorical variables are presented as absolute and relative frequencies (percentages), while numerical variables are described using their median and interquartile range, due to the non-normal distribution assessed by the Kolmogorov–Smirnov test. Descriptive analysis is provided both globally and according to extubation outcome (success or failure). To evaluate statistically significant differences in the distribution of variables between these two groups, the Chi-Square test for independence was employed for categorical variables. For numerical variables, the non-parametric Mann–Whitney *U* test was used, given the rejection of the normality hypothesis.

Diagnostic accuracy comparison between the models

To assess the validity of predictions made by the three models, various validity measures were evaluated, including sensitivity, specificity, positive predictive value, and negative predictive value, alongside concordance measures such as Youden's and the Kappa index.

Briefly, The Kappa index result can be interpreted as follows: values ≤ 0 as indicate no agreement, values between 0.01 and 0.10 as none to slight, 0.11–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement [21].

The Youden's index ranges from 0 through 1. It combines sensitivity and specificity into a single measure (Sensitivity + Specificity − 1). In a perfect test, Youden's index equals 1 [22].

Exploratory outcomes

Inclusion of others aspects modifying weaning (screening, confirmatory test and post-extubation noninvasive respiratory support)

In order to determine whether the inclusion of the SBT variable is justified in the predictive models, subjects were categorized as 'high risk' if they presented any of the SBT categories statistically associated with failure risk (assessed through a univariate logistic regression model) or any of the initial risk factors from the primary models. Subsequently, the same validity and concordance measures as previously described were calculated. Finally, a multivariate logistic regression model was estimated to confirm whether or not the SBT was significantly associated with reintubation. It included the variables associated with extubation failure analysed in the previous models.

Results

The univariate analysis of general characteristics regarding weaning from mechanical ventilation of the 2341 patients included are presented in Table 1 and Table S1. Briefly, the global reintubation rate was 17.3%, the sample included 1433 (61.2%) male patients, with a mean age (± SD) of 60 (± 20) years, an APACHE II score at ICU admission of 16 (± 6), and 334 (14.3%) with coronavirus disease (COVID-19). No differences were observed regarding the screening strategy ($p = 0.533$), but significant differences were observed regarding the SBT strategy ($p < 0.001$), and post-extubation therapy ($p = 0.015$).

Expectedly, more patients were classified as low risk in the 3-factors model compared to the 4 and 11-factors (1016, 802, and 673, respectively): among those patients predicted at low risk according to the 3-factors model, 214 (21.1%) and 473 (46.5%) patients were defined as high risk with the 4 and 11-factors models, respectively. However, there was an unexpected cross-over classification in the opposite way: among the 673 patients at low risk by 11-factors model, 131 (19.5%) patients were defined as high risk with 3 and 4-factors model.

Ninety five reintubations (23.5%) were secondary to non-respiratory reasons (e.g. surgical interventions or coma not associated to hypercapnia). From all the respiratory-related reintubations, 79 (19.5%) were classified as airway failure, and 231 (57%) as weaning failure (see Table S1).

Patients with hypercapnia at the end of the SBT were included only in one trial [12]. It analyzed 182 patients at very high risk according to the 11-factors model. Forty one of these patients were classified as low risk according to the 3-factors model.

Primary outcome

The diagnostic accuracy comparison is presented in Table 2 and Fig. 2. The 3-factors model had a very low global prediction capability (Youden index 0.08 and Kappa index 0.04); the 4-factors and 11-factors model had similar low global prediction capability (Youden index 0.12 and 0.16, and Kappa index 0.06 and 0.07, respectively).

Exploratory outcomes

The effect of the SBT, screening, and post-extubation therapy on reintubation is presented in Table S2). Finally, the adjusted exploratory analysis of a 3, 4, or 11-factors models included finally the SBT only, considering a risk factor a SBT performed with PS 7 cm H_2O + ZEEP or PS 8 cm H_2O + PEEP 5 cm H_2O. All the adjusted models had a low global prediction capability (Youden index 0.08, 0.07, and 0.06, respectively; and Kappa index 0.03, 0.03, and 0.02, respectively) (Table 3).

The multivariate logistic regression model generated to confirm the significant association of the SBT in the univariate analysis is presented in Table S3. All SBT apart from T-tube trials were associated with a significant increase in the reintubation rate.

Screening strategy lacked statistical significance in the univariate analysis and was not included in the adjusted multivariate analysis.

Discussion

The most important results of this study are that all clinical models predicting extubation failure analyzed have low or very low diagnostic accuracy, the 4-factors and 11-factors models have similar low diagnostic accuracy, and prediction is not improved after adjusting for other aspects of weaning (e.g. SBT).

There are some reasons for this limited diagnostic accuracy: first, our case-mix included a high proportion of non-respiratory reintubations (23.5%) secondary to the large amount of patients with primary non-respiratory failure at ICU admission. This reason for failing extubation is unpredictable with any of the clinical models used, and largely explains the low sensitivity and specificity of all models. Second, reintubations classified as airway

Table 1 Patient´s baseline characteristics

	Entire population *n* = 2341	Successfully extubated *n* = 1936	Reintubated *n* = 405	*p*
General characteristics				
Male, *n* (%)	1433	1170 (60.4)	263 (64.9)	0.361
Age, mean (± SD)	60 (20.0)	60 (20.0)	62 (19.0)	0.004
APACHE II at ICU admission, mean (± SD)[a]	16 (6.0)	16 (6.0)	18 (8.0)	< 0.001
SARS COVID-19, no. (%)[b]	334	274 (14.2)	60 (14.8)	0.789
High risk factors for extubation failure, no. (%)[c]				
Hypercapnia at the end of the SBT	127	86 (4.4)	41 (10.1)	< 0.001
Airway patency problems	140	109 (0.1)	31 (7.6)	0.148
Respiratory secretions	332	248 (12.8)	84 (20.7)	< 0.001
Not simple weaning	602	446 (23.0)	156 (38.5)	< 0.001
APACHE II > 12 at extubation[a]	493	358 (18.5)	135 (33.3)	< 0.001
BMI > 30[d]	379	272 (14.0)	107 (26.4)	< 0.001
CHARLSON index ≥ 2[e]	963	764 (39.5)	199 (10.3)	< 0.001
Prolonged mechanical ventilation	584	446 (23.0)	138 (34.1)	< 0.001
Acute cardiac failure	190	138 (7.1)	52 (12.8)	0.001
At least moderate COPD	239	179 (9.2)	60 (14.8)	0.001
Age > 65 y	713	572 (29.5)	141 (34.8)	0.042
Comorbidities, no. (%)[e]				
Chronic heart disease	489	401 (20.7)	88 (21.7)	0.697
COPD	241	190 (9.8)	51 (12.6)	0.113
Other respiratory diseases	561	449 (23.2)	112 (27.6)	0.064
Vascular diseases	114	84 (4.3)	30 (7.4)	0.013
Hepatic diseases	175	125 (6.5)	50 (12.3)	< 0.001
Other diseases	324	254 (13.1)	70 (17.3)	0.033
Diagnosis at ICU admission, no. (%)[g]				
Primary respiratory failure	1346	1092 (56.4)	254 (62.7)	0.023
Primary non-respiratory failure	1376	1150 (59.4)	226 (55.8)	0.200
Trauma	294	243 (12.5)	51 (12.6)	1
Surgical procedure	781	656 (33.9)	125 (30.8)	0.265
Screening strategy, no. (%)[f]				
Conservative strategy	1980	1633 (84.3)	347 (85.7)	0.533
Aggressive strategy	361	303 (15.6)	58 (14.3)	
SBT, no. (%)				
T-tube trial	382	345 (17.8)	37 (9.1)	< 0.001
Pressure support 5 cm H_2O	517	445 (22.9)	72 (17.8)	
Pressure support 7 cm H_2O	1075	843 (43.5)	232 (57.3)	
Pressure support 8 + PEEP 5 cm H_2O	367	304 (15.7)	63 (15.6)	
Postextubation preventive therapy, no. (%)				
COT	263	231 (11.9)	32 (7.9)	0.015
HFNC	1600	1325 (68.4)	275 (68.0)	
NIV	478	380 (19.6)	98 (24.2)	
Three factors model predicting extubation failure, no. (%)				
Low risk patients	1016	866 (44.7)	150 (37.0)	0.007
High risk patients	1325	1070 (55.3)	255 (63.0)	
Four factors model predicting extubation failure, no. (%)				
Low risk patients	802	702 (36.3)	100 (24.7)	< 0.001
High risk patients	1539	1234 (63.7)	305 (75.3)	

Table 1 (continued)

	Entire population *n* = 2341	Successfully extubated *n* = 1936	Reintubated *n* = 405	*p*
Eleven factors model predicting extubation failure, no. (%)				
Low risk patients	673	609 (31.5)	64 (15.8)	< 0.001
High risk patients	1668	1327 (68.5)	341 (84.2)	

APACHE II Acute Physiology and Chronic Health Evaluation II; *BMI* body mass index; *COPD* chronic obstructive pulmonary disease; *COT* conventional oxygen therapy; *HFNC* high flow nasal cannula; *NIV* noninvasive ventilation; *PEEP* positive end-expiratory pressure; *SARS COVID-19* severe acute respiratory syndrome secondary to coronavirus disease 2019; *SBT* spontaneous breathing trial

[a] APACHE II score was calculated from 17 variables. Scores range from 0 to 71 points, with higher scores indicating more severe disease

[b] Severe hypoxemic respiratory failure secondary to RT-PCR-confirmed COVID-19 pneumonia

[c] Defined according to references [12, 13]

[d] Body mass index calculated as weight in kilograms divided by height in meters squared

[e] Comorbidities were categorized based on the Charlson Comorbidity Index (detailed in the online supplement). Fully detailed in Table S1

[f] Screening strategy defined according to the protocol reported in reference [8]

[g] There can be more than one diagnosis

failure related were also prevalent (19.5%). Surprisingly, risk factors related to airway failure (e.g. airway patency problems and secretions management) showed no significant or slight association (OR 1.43, 95%CI 1.06–1.91) with reintubation, respectively. Although they lack objective criteria, their low prevalence (6 and 14%, respectively) could have contributed to their weak statistical association.

The most important flaw of the 3-factors compared to the 11-factors model was the low sensitivity (62.96% vs 84.20%) with a reduced advantage on specificity (44.75% vs 31.46%), leading to a theoretical difference in optimal selection of post-extubation preventive therapy of 5% (85.24% vs 90.49% NPV, respectively), with an estimated risk of increased reintubation rate from 0.2% [15] to 0.3% [13]. The 4-factors model could have a theoretical increase in the reintubation rate compared to the 11-factors model of 0.12% [15] to 0.18% [13].

There is a risk for collinearity including simultaneously the Charlson index and comorbidities separately, and APACHE II and age. This could partially explain why some patients at low risk using the 11-factors model were classified as high risk using the 3 or 4-factors model.

Adjusting for other aspects of weaning included only the SBT. Post-extubation therapy was not included in the final model to avoid selection bias, as the preventive strategy is planned according to the predicted risk. HFNC and NIV were used in patients with progressive increased risk [12, 15]. The screening strategy did not show a significant result for reintubation and it was also not included [11].

Regarding the SBT, our univariate analysis reported significant differences on reintubation rate, and the multivariate confirmed the result for aggressive SBTs. The association between SBT and reintubation is a traditional controversy. Former physiological studies recommended SBT on PS 0 cm H_2O, continuous positive pressure (CPAP) 0 cm H_2O or T-tube, as they most accurately reflect the physiologic conditions after extubation [23, 24]. Later large clinical trials centered on SBT effects have never confirmed differences in reintubation rate regarding the type of SBT [23, 25]. However, a recent

Table 2 Comparison of crude diagnostic accuracy for reintubation at 7 days of the three models including 2341 patients recruited in five randomized trials [11, 12, 14, 15, 19]

	Se	Sp	PPV	NPV	Youden	Kappa
3 Factors model	62.96%	44.75%	19.26%	85.24%	0.08	0.04
4 Factors model	75.31%	36.31%	19.83%	87.55%	0.12	0.06
11 Factors model	84.20%	31.46%	20.44%	90.49%	0.16	0.07

Se sensitivity; *Sp* specificity; *PPV* positive predictive value; *NPV* negative predictive value

The 3-factors model included the following risk factors: age > 65y, chronic heart or pulmonary disease. The 4-factors model included the following risk factors: age > 65y, chronic heart or pulmonary disease, and prolonged mechanical ventilation. The 11-factors model included the following risk factors: age > 65 years, ≥ 2 comorbidities, prolonged mechanical ventilation, acute heart failure as the primary indication for mechanical ventilation, moderate-to-severe chronic obstructive pulmonary disease, APACHE II score > 12 on extubation day, airway patency problems, inability to deal with respiratory secretions, not simple weaning, obesity, or hypercapnia at the end of the spontaneous breathing trial

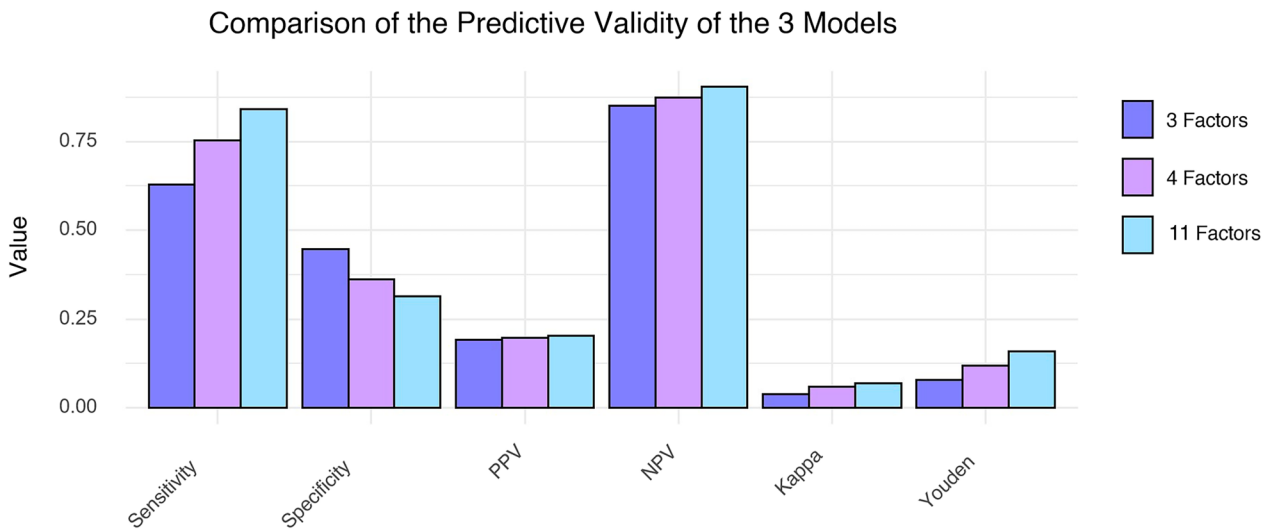

Fig. 2 Diagnostic accuracy comparison of the three models

clinical meta-analysis found possible differences in reintubation [26].

There are some possible clinical explanations for this conflicting results: first, widespread use of post-extubation noninvasive respiratory support after the results of the initial physiologic studies can modify post-extubation work of breathing; second, there could be some physiologic bias not considered when estimating the pre-extubation (e.g., endotracheal intraluminal diameter narrowing secondary to biofilm) [27] or post-extubation work of breathing (e.g., conditioning gases inhaled immediately after extubation) [14]. Third, reconnecting patients to the previous ventilator settings for rest after tolerating the SBT, reduced the reintubation rate irrespective of the length and type of SBT [28, 29], thus any kind of SBT could underestimate the actual post-extubation work of breathing. Both studies by Subira et al. [25] and Thille et al. [30] allowed patients considered at risk to rest and receive prevention. To explain our results supporting the former physiologic studies, there are some possible additional reasons: first, only two trials included in the current analyses were published before the 2017 guidelines and allowed T-tube trials [15, 19, 31]; second, a SBT on PS 8 cm H_2O + 5 cm H_2O PEEP seems to be inadequately high when planned prevention includes HFNC alone therapy [11]; third, other aspects of weaning (e.g., risk stratification, screening and SBT) can buffer the effect of preventive therapy on reintubation. The analysis of patients randomized in clinical trials testing the optimal preventive therapy could have reduced the buffer effect of the type of SBT on reintubation. In addition, non-respiratory reintubations can also modify the effect of SBT on reintubation [15, 19].

Our multivariate analysis rejected some of the most widely accepted factors (e.g., age, acute heart failure, and comorbidities analyzed separately) (see Table S3) [21, 22]. In addition, there is a risk of selection bias for hypercapnia at the end of the SBT, as this risk factor was an exclusion criteria in some of the original trials [10, 14, 15,

Table 3 Comparison of adjusted for aggressive SBT (pressure support 7 cm H_2O or pressure support 8 cm H_2O + 5 cm H_2O PEEP) diagnostic accuracy for reintubation at 7 days of the three models including 2341 patients recruited in five randomized trials [11, 12, 14, 15, 19]

	Se	Sp	PPV	NPV	Youden	Kappa
3 Factors model	90.37%	17.56%	18.65%	89.71%	0.08	0.03
4 Factors model	93.33%	13.58%	18.43%	90.69%	0.07	0.03
11 Factors model	96.05%	10.18%	18.28%	92.49%	0.06	0.02

Se sensitivity; *Sp* specificity; *PPV* positive predictive value; *NPV* negative predictive value

The 3-factors model included the following risk factors: age > 65y, chronic heart or pulmonary disease. The 4-factors model included the following risk factors: age > 65y, chronic heart or pulmonary disease, and prolonged mechanical ventilation. The 11-factors model included the following risk factors: age > 65 years, ≥ 2 comorbidities, prolonged mechanical ventilation, acute heart failure as the primary indication for mechanical ventilation, moderate-to-severe chronic obstructive pulmonary disease, APACHE II score > 12 on extubation day, airway patency problems, inability to deal with respiratory secretions, not simple weaning, obesity, or hypercapnia at the end of the spontaneous breathing trial

19], and low prevalence (e.g. acute heart failure) can also debilitate statistical results.

Even more, it is difficult to compare risk factors with one another from a methodological perspective: on one hand different models including different risk factors in a non-collapsible scale (e.g. odds ratio) could associate methodological pitfalls, on the other hand the comparison of the "strength of association" of two risk factors within a single model should also be avoided due to Table 2 fallacy [7, 32]. Our primary outcome, including all-cause reintubation within 7 days can mitigate the methodological weakness related to differences in strength of association between some risk factors and reintubation [2, 10].

This study has some limitations: first, the post hoc design of the analysis, which was not prespecified, makes impossible to completely rule out bias. Our primary outcome including all-cause reintubations, and the randomized design of the original studies, with prospective classification of reasons for reintubation aimed to reduce these bias. Second, our definition for chronic lung and heart diseases slightly differ from those originally used by Thille et al. [33, 34], as we used the Charlson comorbidity index. This index includes more categories for both chronic cardiac and lung diseases. Thus, there is a risk in the current 3-factors model of overrating compared to the results in the study by Thille et al. [33, 34]. Third, there is a risk for selection bias when including the SBT in the multivariate model, as T-tube SBTs were performed only in two trials [15, 19]. Fourth, the exclusion of hypercapnia at the end of the SBT in our multivariate analysis could be explained by a selection bias, because it was an exclusion criteria in four trials [11, 14, 15, 19]. Fifth, we chose reintubation at 7 days within extubation as primary outcome because most patients received post-extubation preventive therapy, but those receiving conventional oxygen therapy or HFNC for 24 h, could have miss-classified some second-intubation episodes as reintubations. Sixth, from an ethical point of view, it is quite difficult to fully elucidate the effect of post-extubation therapy on the risk for failing extubation with this methodology due to the evidence demonstrating its benefit. However, the inclusion of randomized populations can reduce the effect of this bias in our results.

Conclusions

Based on these results, the 3-factors model reported a very low diagnostic accuracy, and the 4 or 11-factors models showed similar low accuracy. No meaningful improvement was observed after adjusting for other aspects of weaning.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s00134-024-07758-0.

Author details

[1] Critical Care Medicine, Complejo Hospitalario Universitario de Toledo, Toledo, Spain. [2] Centre Hospitalier Universitaire de Poitiers, Médecine Intensive Réanimation, Poitiers, France. [3] INSERM Centre d'Investigation Clinique 1402 ALIVE, Université de Poitiers, Poitiers, France. [4] Unité de Surveillance Continue et Soins Intensifs Respiratoires (USC/USIR), Pôle Soins Critiques, Groupe Hospitalier Est Réunion (GHER), Saint-Benoit, France. [5] Programa Investigo SESCAM, Toledo, Spain. [6] Institut dÍnvestigació i Innovació Parc Tauli (I3PT-CERCA), Grup de Recerca Traslacional del Pacient Crític, Universitat Autònoma de Barcelona, Sabadell, Spain. [7] Ciber Enfermedades Respiratorias (CIBERES), Health Institute Carlos III, Madrid, Spain. [8] Instituto de Investigación Hospital Universitario La Paz (IdiPAZ), Madrid, Spain. [9] Grupo de Investigación en Disfunción y Fallo Orgánico en La Agresión (IdiPAZ), Madrid, Spain. [10] Universidad Alfonso X el Sabio, Madrid, Spain.

Author contributions

Dr. Hernandez had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Concept and design: Thille and Hernandez. Acquisition, analysis, and interpretation of data: Rodriguez, Thille, Marquez, Frat, Leal, Alonso, Pais, Morales, Colinas, Propin, and Hernandez. Drafting of the manuscript: Rodriguez, and Hernandez. Critical revision of the manuscript for important intellectual content: Rodriguez, Thille, Marquez, Frat, Leal, Alonso, Pais, Morales, Colinas, Propin, Fernandez, and Hernandez. Statistical analysis: Martínez Balaguer, and Alvaredo Rodrigo. Administrative, technical, or material support: Hernandez. Studies supervision: Hernandez.

Funding

Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Data availability

Data will be availabe after sensible request to the corresponding author.

Declarations

Conflicts of interest

All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr. Hernandez reported travel expenses and personal fees from Fisher & Paykel Healthcare Ltd.

Role of the funder/sponsor

None.

Publisher's Note

Received: 3 October 2024 Accepted: 4 December 2024
Published: 7 January 2025

References

1. Torrini F, Gendreau S, Morel J et al (2021) Prediction of extubation outcome in critically ill patients: a systematic review and meta-analysis. Crit Care 25:391
2. Jaber S, Quintard H, Cinotti R et al (2018) Risk factors and outcomes for airway failure versus non-airway failure in the intensive care unit: a multicenter observational study of 1514 extubation procedures. Crit Care 22:236
3. Pham T, Heunks L, Bellani G et al (2023) Weaning from mechanical ventilation in intensive care units across 50 countries (WEAN-SAFE): a multicentre, prospective, observational cohort study. Lancet Respir Med 11:465–476
4. Thille AW, Boissier F, Ghezala HB et al (2015) Risk factors for and prediction by caregivers of extubation failure in ICU patients: a prospective study. Crit Care Med 43:613–620
5. Ferrer M, Valencia M, Nicolas JM et al (2006) Early noninvasive ventilation averts extubation failure inpatients at risk. Am J Respir Crit Care Med 173:164–170
6. Nava S, Gregoretti C, Fanfulla F et al (2005) Noninvasive ventilation to prevent respiratory failure after extubation in high-risk patients. Crit Care Med 33:2465–2470
7. Hernandez G, Vaquero C, Ortiz R et al (2022) Benefit with noninvasive ventilation in subgroups of patients at high-risk for reintubation: a post-hoc analysis. J Intensive Care 10:43. https://doi.org/10.1186/s40560-022-00635-2
8. De Jong A, Capdevila M, Aarab Y et al (2024) Incidence, risk factors, and long term outcomes for extubation failure in ICU patients with obesity. Lancet Respir Med. https://doi.org/10.1016/j.chest.2024.07.171
9. De Jong A, Bignon A, Stephan F et al (2023) Effect of non-invasive ventilation after extubation in critically ill patients with obesity in France: a multicenter, unblinded, pragmatic randomised clinical trial. Lancet Respir Med 11(6):530–539
10. Epstein SK, Ciubotaru RL (1998) Independent effects of etiology of failure and time to reintubation on outcome for patients failing extubation. Am J Respir Crit Care Med 158:489–493
11. Hernandez G, Rodriguez P, Soto J et al (2024) Effect of aggressive vs conservative screening and confirmatory test on time to extubation among patients at low or intermediate risk: a randomized clinical trial. Intensive Care Med 50:258–267
12. Hernandez G, Paredes I, Moran F et al (2022) Effect of postextubation noninvasive ventilation with active humidification vs high-flow nasal cannula on reintubation in patients at very high risk for extubation failure: a randomized trial. Intensive Care Med 48(12):1751–1759
13. Thille AW, Muller G, Gacouin A et al (2019) Effect of postextubation high-flow nasal oxygen with noninvasive ventilation vs high-flow nasal oxygen alone on reintubation among patients at high risk of extubation failure: a randomized clinical trial. JAMA 322(15):1465–1475
14. Hernandez G, Vaquero C, Gonzalez P et al (2016) Effect of postextubation high-flow nasal cannula vs conventional oxygen therapy on reintubation in low risk patients: a randomized clinical trial. JAMA 315(13):1354–1361
15. Hernandez G, Vaquero C, Colinas L et al (2016) Effect of postextubation high-flow nasal cannula vs noninvasive ventilation on reintubation and postextubation respiratory failure in high-risk patients: a randomized clinical trial. JAMA 316(15):1565–1574
16. Boles JM, Bion J, Connors A et al (2007) Weaning from mechanical ventilation. Eur Respir J 29(5):1033–1056
17. Beduneau G, Pham T, Schortgen F et al (2017) Epidemiology of weaning outcome according to a new definition. The WIND study. Am J Respir Crit Care Med 195(6):772–783
18. Kress JP, Hall JB (2014) ICU-acquired weakness and recovery from critical illness. N Engl J Med 370:1626–1635
19. Hernandez G, Dianti J, Paredes I et al (2024) Humidified noninvasive ventilation vs high-flow therapy to prevent reintubation in obese patients. Am J Respir Crit Care Med. https://doi.org/10.1164/rccm.202403-0523OC
20. Cohen JF, Korevaar DA, Altman DG et al (2016) STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open 6:e012799
21. Sackett DL (1992) A primer on the precision and accuracy of the clinical examination. JAMA 267:2638–2644
22. Steyerberg EW, Calster BV, Pencina MJ (2011) Performance measures for prediction models and markers: evaluation of predictions and classifications. Rev Esp Cardiol 64(9):788–794
23. Sklar MC, Burns K, Rittayamai N et al (2017) Effort to breathe with various spontaneous breathing trial techniques: a physiological meta-analysis. Am J Respir Crit Care Med 195(11):1477–1485
24. Tobin MJ (2012) Extubation and the myth of "minimal ventilator settings." Am J Respir Crit Care Med 185:349–350
25. Subira C, Hernandez G, Vazquez A et al (2019) Effect of pressure support vs T-piece ventilation strategies during spontaneous breathing trials on successful extubation among patients receiving mechanical ventilation. A randomized trial. JAMA 321(22):2175–2182
26. Ippolito M, Sardo S, Tripodi VF et al (2024) Association between spontaneous breathing trial methods and reintubation in adult critically ill patients. A systematic review and network meta-analysis of randomized controlled trials. Chest 166:1020–1034. https://doi.org/10.1016/j.chest.2024.06.3773
27. Boque MC, Gualis B, Sandiumenge A et al (2004) Endotracheal tube intraluminal diameter narrowing after mechanical ventilation: use of acoustic reflectometry. Intensive Care Med 30(12):2204–2209
28. Fernandez MM, Gonzalez-Castro A, Magret M et al (2017) Reconnection to mechanical ventilation for 1 h after a successful spontaneous breathing trial reduces reintubation in critically ill patients: a multicenter randomized controlled trial. Intensive Care Med 43:1660–1667
29. Hernandez G, Fernandez R, Luzon E et al (2007) The early phase of the minute ventilation recovery curve predicts extubation failure better than the minute ventilation recovery time. Chest 131:1315–1322
30. Fleuren LM, Dam TA, Tonutti M et al (2021) Predictors for extubation failure in COVID-19 patients using a machine learning approach. Crit Care 25:448
31. Schmidt GA, Girard TD, Kress JP et al (2017) Official executive summary of an American Thoracic Society/American College of Chest Physicians clinical practice guideline: liberation from mechanical ventilation in critically ill adults. Am J Respir Crit Care Med 195(1):115–119
32. Westreich D, Greenland S (2013) The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. Am J Epidemiol 177(4):292–298
33. Thille AW, Gacouin A, Coudroy R et al (2022) Spontaneous-breathing trials with pressure-support ventilation or a T-piece. N Engl J Med 387:1843–1854
34. Thille AW, Coudroy R, Nay MA et al (2021) Beneficial effects of noninvasive ventilation after extubation in obese or overweight patients: a post-hoc analysis of a randomized clinical trial. Am J Respir Crit Care Med. https://doi.org/10.1164/rccm.202106-1452OC