ORIGINAL

Giorgio Conti
Luca Montini
Mariano Alberto Pennisi
Franco Cavaliere
Andrea Arcangeli
Maria Grazia Bocci
Rodolfo Proietti
Massimo Antonelli

A prospective, blinded evaluation of indexes proposed to predict weaning from mechanical ventilation

G. Conti (✉) · L. Montini · M. A. Pennisi ·
F. Cavaliere · A. Arcangeli · M. G. Bocci ·
R. Proietti · M. Antonelli
Istituto Anestesiologia e Rianimazione,
Policlinico Agostino Gemelli,
Università Cattolica del Sacro Cuore,
Largo Agostino Gemelli 8, 00168 Rome,
Italy
e-mail: g.conti@rm.unicatt.it
Tel.: +39-06-30154386/30154507
Fax: +39-06-3013450

Abstract *Objective:* To conduct a blinded evaluation of the predictors of weaning from mechanical ventilation. *Design:* A prospective clinical study. *Setting:* A 23-bed general intensive care unit. *Patients:* Ninety-three non-selected patients, ventilated for more than 48 h. *Methods:* The study had two steps: at first, patients' data were used to select the cut-off value for weaning predictors (the minimal false classification). The cut-off value for each index was prospectively assessed in a group of 52 patients. The predictive performance of these indexes was evaluated by calculating the area under the receiver operating characteristic curve. In the prospective-validation set we used Bayes' theorem to assess the probability of each test in predicting weaning. The physicians making decisions about the weaning process were always unaware of the predictive values. Weaning was considered successful if spontaneous breathing was sustained for more than 48 h after extubation. *Measurements and results:* During the first 2 min after discontinuation of mechanical ventilation the following tests were performed: vital capacity, tidal volume, airway occlusion pressure ($P_{0.1}$), minute ventilation, respiratory rate, maximal inspiratory pressure (MIP), respiratory frequency to tidal volume (f/V_T), $P_{0.1}$/MIP and $P_{0.1} \times f/V_T$. The areas under the curve showed that the tests had not the ability to distinguish between successful and unsuccessful weaning. *Conclusion:* Our results show that all the evaluated indexes are poor predictors of weaning outcome in a general intensive care unit population.

Keywords Mechanical ventilation · Weaning · Cut-off values · Receiver operating characteristic curve · Likelihood ratio · Weaning predictors

Introduction

Weaning from mechanical ventilation represents an important issue, because both an early and a delayed extubation can burden the patient's health, increasing the risk of infections and the length of hospital stay.

Although many patients show stable conditions just after disconnection from mechanical ventilation, spontaneous breathing can become gradually less effective in sustaining valid ventilation, sometimes requiring the re-institution of mechanical ventilation. This suggests the importance of identifying predictors of weaning from mechanical ventilation. Many studies have assessed the possibility to predict weaning in critically ill patients reliably [1, 2, 3, 4, 5, 6, 7, 8]. One of the major methodological limitations of many of these studies was the lack of blinding [9].

The aim of this study was to conduct a prospective, blinded evaluation of the most diffuse predictors of weaning in a non-selected sample of critically ill patients. We analyzed several indexes including: airway occlusion pressure ($P_{0.1}$), maximal inspiratory pressure (MIP), res-

piratory frequency to tidal volume (f/V_T) ratio, $P_{0.1}$ associated with MIP and f/V_T ratio, minute ventilation, respiratory rate, tidal volume and vital capacity. From a first group of patients (training set) we identified the threshold values for each index; then, we tested the predictive accuracy of these values in a prospective-validation set of patients.

Patients and methods

Ninety-three patients were evaluated; their clinical characteristics are reported in Table 1. Before the weaning trial, all the patients were receiving pressure support ventilation 10–15 cmH$_2$O and PEEP 3–5 cmH$_2$O. All patients were intubated with orotracheal tubes, 7.5–8.5 mm internal diameter. The ventilator used was the Servo Ventilator 300 (Siemens, Sweden).

Discontinuation from mechanical ventilation was attempted when the primary physician judged that the patient was ready to be weaned, according to the following criteria: (1) the cause for starting mechanical ventilation had resolved or clearly improved; (b) body temperature was below 38.5°C; (c) hemoglobin was equal to or higher than 8 g/dl; (d) no intravenous sedatives had been given for at least 24 h before the weaning trial; (e) there were no clinical signs of left ventricular failure/no cardiac rhythm or conduction disturbances [10]. These are the standard clinical criteria commonly adopted in our intensive care unit when deciding whether a patient is ready to be weaned. When all these criteria were present, the ability of the patient to sustain spontaneous breathing was evaluated with a 2-h T-piece trial.

During the first 2 min after discontinuation of mechanical ventilation the following tests were performed, six of these were single variables: vital capacity (ml/kg); tidal volume (ml/kg); airway occlusion pressure (cmH$_2$O); minute ventilation (l); respiratory rate (breaths/min) and maximal inspiratory pressure (cmH$_2$O), while three were derived variables: f/V_T (breaths/min per l); $P_{0.1}$/MIP and $P_{0.1} \times f/V_T$ (cmH$_2$O/breaths per min per l). Our measurement techniques have been extensively described in a previous study [3]. The data obtained were not available for the attending physician, who was unaware of the results of the weaning tests and, therefore, independently took the decision to continue the T-piece trial or reinstitute the ventilatory support.

During the 2-h period of spontaneous breathing, tolerance was continuously evaluated by the attending physician. The trial was stopped if at least one of the following intolerance criteria was present: respiratory rate above 35 (breaths/min); PaO$_2$ below 65 mmHg with FIO$_2$ less than 0.6; pH 7.34 or less; heart rate equal to or above 130 beats/min or increased by 20% or more, or if arrhythmias appeared; systolic blood pressure without inotropes below 80 mmHg or above 200 mmHg; ineffective cough; uncoordinated thoracoabdominal movement; activation of the accessory muscles; agitation or depressed mental status [11].

If the patient had poor clinical tolerance, ventilation was restarted; if, however, the patient remained stable at the end of the 2 h, the endotracheal tube was removed. A weaning trial was considered a failure when the patient did not tolerate the spontaneous breathing trial and required reconnection to mechanical ventilation. Weaning was considered successful if spontaneous breathing was sustained for more than 48 h after extubation. Finally, extubation was considered a failure if the patient required reintubation within 48 h.

The study had two different parts: during the first one, data were used to select the cut-off value for weaning predictors. The selected values were those that resulted in the fewest false classifications. During the second part, the threshold value for each index was assessed prospectively in an additional group of patients. All the patients in the first 4 months of the study served as the training set and those in the following 4 months as the prospective-validation set.

A true positive (TP) result was defined as when a test predicted successful weaning and weaning actually occurred; a false positive (FP) result was defined as when a test predicted successful weaning but weaning failed; a false negative (FN) result was defined as when a test predicted weaning failure but it was indeed successful; a true negative (TN) result was defined as when a test predicted weaning failure and the patient really failed the weaning trial [5].

Receiver operating characteristic curve analysis was performed with MedCalc software version 6.10.001 (2001 Frank Schoonjans, Belgium) [12]. This analysis provides a powerful means of assessing a test's ability to discriminate between two groups of patients with the advantage that the analysis does not depend on the threshold value selected. The value selected as the threshold value was the one that had the highest accuracy (minimal false negative and false positive results).

Standard formulas were used to calculate the sensitivity TP/(TP+FN), specificity TN/(TN+FP), accuracy (TP+TN)/(TP+TN+FP+FN), likelihood ratio of positive test (ρ_+) = sensitivity/(1-specificity) and likelihood ratio of negative test (ρ_-) = (1-sensitivity)/specificity. Positive and negative likelihood ratios are independent of the prevalence of the disease. The cut-off values were then assessed in an additional group of 51 patients and the predictive performance of each index was evaluated by calculating the area under the receiver operating characteristic curve (AUC).

According to an arbitrary guideline [13], one could distinguish between non-informative (AUC=0.5), less accurate (0.5<AUC≤0.7), moderately accurate (0.7<AUC≤0.9), highly accurate (0.9<AUC<1) and perfect test (AUC=1) [14]. If the 95% confidence interval for the area does not include the 0.5 value, there is evidence that the test has an ability to distinguish between the two groups [12].

In the prospective-validation set, the prevalence of weaning success and weaning failure were calculated. We also calculated the likelihood ratios = (ρ_+)/(ρ_-) for each index in the prospective-validation set [15]. Likelihood ratios between 0.5 and 2.0 indicate that a weaning parameter is associated with only small changes in the post-test probability of success or failure. Likelihood ratios from 2 to 5 and from 0.3 to 0.5 correlate with small but potentially important changes in probability, while ratios of 5–10 or 0.1–0.3 correlate with more clinically important changes in probability. Ratios of higher than 10 or lower than 0.1 correlate with very large changes in probability [16, 17, 18, 19].

Finally, in the prospective-validation set, according to Sassoon [5], we used Bayes' theorem to assess the performance of each test in predicting weaning outcome as a function of prevalence of weaning success and failure in our population. Bayes' theorem allows the calculation of the probability of success or failure of weaning after the performance of a test (post-test probability). The formulae used to calculate post-test probability are shown in Table 5 [20].

The results are reported as means ± standard deviation. Comparison between proportions was made using the chi-square test (with Yates' correction for continuity); comparison between means was made using the F-test: a probability of less than 0.05 was considered significant.

Results

We included 93 patients: their main characteristics are described in Table 1. Initially, we intended to distinguish between extubation failure and failure weaning patients. However, we could not make such a distinction because

Table 1 Clinical characteristics and cause of the acute respiratory failure in the entire study population (*ALI* acute lung injury, *ARDS* acute respiratory distress syndrome, *CHF* cardiac heart failure, *COPD* chronic obstructive pulmonary disease, *FW* weaning failure, *MV* duration of mechanical ventilation before weaning trial, *PEEP* positive end-expiratory pressure, *PSV* pressure support ventilation, *SAPS II* Simplified Acute Physiologic Score II, *SW* successful weaning). Note that one patient with extubation failure was not included in the analysis

Clinical characteristics	PSV	PEEP	All patients (*n*=92)	SW (*n*=59)	FW (*n*=33)
Age (years)	–	–	66±16	66±17	68±14
Weight (kg)	–	–	72±15	72±10	74±21
Height (cm)	–	–	167±8	168±7	166±8
SAPS II on admission	–	–	39±13	39±13	39±13
Duration of MV (days)	–	–	9±10	8±9	9±10
Male sex, *n* (%)	–	–	58 (63)	37 (62.7)	21 (63.6)
Cause of acute respiratory failure					
COPD, *n* (%)	15±0	5±1	25 (27)	17 (29)	8 (24)
CHF/cardiac arrest, *n* (%)	15±0	5±0	6 (6.5)	4 (7)	2 (6)
ALI/ARDS, *n* (%)	15±0	5±1	20 (22)	11 (19)	9 (27)
Neurological disorder, *n* (%)	15±0	5±0	14 (15)	6 (10)	8 (24)
Postoperative acute respiratory failure, *n* (%)	15±1	5±1	25 (27)	21 (36)	4 (12)[a]
Multiple trauma, *n* (%)	15±0	4±1	2 (2)	0 (0)	2 (6)

[a] *p*<0.05

Table 2 Clinical characteristics, cause of acute respiratory failure and weaning outcome in the training set and in the prospective-validation set (see Table 1 for explanation of abbreviations). Note that one patient with extubation failure was not included in the prospective-validation set. *p*<0.05 was considered significant

Clinical characteristic	Training set (*n*=41)	Prospective-validation set (*n*=51)	*p* value
Age (years)	69±15	65±16	0.22[a]
Weight (kg)	72±12	73±16	0.74[a]
Height (cm)	168±7	167±8	0.53[a]
SAPS II on admission	42±12	37±14	0.07[a]
Duration of MV (days)	8±9	9±11	0.64[a]
Male sex, *n* (%)	28 (68)	30 (58)	0.43[b]
Cause of acute respiratory failure			
COPD, *n* (%)	9 (22)	16 (31)	0.46[b]
CHF/cardiac arrest, *n* (%)	2 (5)	4 (8)	0.87[b]
ALI/ARDS, *n* (%)	9 (22)	11 (21.5)	0.84[b]
Neurological disorder, *n* (%)	9 (22)	5 (10)	0.19[b]
Postoperative acute respiratory failure, *n* (%)	11 (27)	14 (27)	0.81[b]
Multiple traumas, *n* (%)	1 (2)	1 (2)	0.45[b]
Weaning outcome			
SW, *n* (%)	22 (54)	37 (72)	0.11[b]
FW, *n* (%)	19 (46)	14 (28)	0.11[b]

[a] *t*-test; [b] chi-square test (with Yates' correction for continuity)

only one patient in the study required intubation within 48 h after extubation. The prevalence of extubation failure in this study was only 0.011 (1/93).

We did not observe significant differences between the groups "successful weaning" and "weaning failure" regarding their clinical characteristics, diagnosis, sex, weight, height, duration of mechanical ventilation before trial weaning and Simplified Acute Physiologic Score II, neither in the training set nor in the prospective-validation set (Table 2). No statistical difference between the two groups was observed concerning the values of heart rate, systolic blood pressure, pH and PaO_2 (Table 3).

Ninety patients had an inspiratory support level between 13 and 18 cmH_2O. In order to homogenize this level, we applied by protocol an inspiratory pressure support of 15 cmH_2O. Two patients were already ventilated with a pressure support level of 10 cmH_2O that was not modified.

In the training set, the threshold values that discriminated between successful weaning and weaning failure are shown in Table 4. In this group the prevalence of weaning success was 0.54 and that of weaning failure was 0.46. The accuracy was 0.71 for maximal inspiratory pressure, slightly higher than the accuracy of vital capacity and tidal volume. These results were in accordance with the values of the likelihood ratio of positive test and likelihood ratio of negative test.

In the prospective-validation set the prevalence of weaning success was 0.72 (37/51) while that of weaning failure was 0.28 (14/51). In the entire study population the prevalence of weaning success was 0.64 (59/92) and the prevalence of weaning failure was 0.36 (33/92). In the prospective-validation set the likelihood ratio values ranged between 0.69 and 1.87: therefore all the indexes were associated with small changes in the post-test probability of success or failure [9, 15] (Table 5). These results were in accordance with the values of the probability

Table 3 Index and clinical variables measured during the first 2 min after discontinuation of ventilator support (*f/V*$_T$ rapid shallow breathing, *FW* weaning failure, *HR* heart rate, *MIP* maximal inspiratory pressure, *P*$_{0.1}$ airway occlusion pressure, *RR* respiratory rate, *SBP* systolic blood pressure, *SW* successful weaning, *VC* vital capacity, *V*$_{MIN}$ minute ventilation, *V*$_T$ tidal volume). Note that one patient with extubation failure was not included in the prospective-validation set. No significant differences were observed between SW and FW in the training set and in the prospective-validation set

Index	Training set		Prospective-validation set	
	SW (*n*=22)	FW (*n*=19)	SW (*n*=37)	FW (*n*=14)
VC (ml/kg)	13±5	10±5	10±5	9±4
V$_T$ (ml/kg)	7±2	6±2	6±2	6±3
f/V$_T$ (breaths/min per l)	60±44	79±34	72±41	73±34
P$_{0.1}$ (cmH$_2$O)	3±1	3±1	2±1	2±1
P$_{0.1}$/MIP	0.14±0.06	0.16±0.07	0.10±0.04	0.11±0.03
V$_{MIN}$ (l)	10±2.6	8.6±2.1	9.9±3	9.4±2.2
RR (breaths/min)	25±14	27±6	26±8	26±8
MIP (cmH$_2$O)	21±1	17±6	25±7	23±6
P$_{0.1}$ × f/V$_T$ (cmH$_2$O/ breaths per min per l)	174±144	220±182	170±109	176±96
Clinical variables				
HR (beats/min)	91±18	97±13	91±18	95±21
SBP (mmHg)	142±23	136±26	137±22	139±33
pH	7.45±0.05	7.46±0.05	7.45±0.05	7.44±0.06
PaCO$_2$ (mmHg)	40±9	44±9	40±11	43±8
PaO$_2$/FIO$_2$	274±82	284±96	250±82	223±59

Table 4 Threshold values had the best discrimination between patients successfully weaned and those who failed in the training set; > and ≤ indicate whether the values above the threshold or those below it were predictive of successful weaning (*AUC* area under receiver operating characteristic curve, *CI* confidence interval, *FN* false negative, *FP* false positive, *f/V*$_T$ rapid shallow breathing, *MIP* maximal inspiratory pressure, *ρ+* likelihood ratio of positive test, *ρ-* likelihood ratio of negative test, *P*$_{0.1}$ airway occlusion pressure, *RR* respiratory rate, *SE* standard error, *VC* vital capacity, *V*$_{MIN}$ minute ventilation, *V*$_T$ tidal volume). Larger values of *ρ+* and smaller values of *ρ−* indicate greater diagnostic ability. *ρ−* and *ρ+* are independent of prevalence of weaning outcome

Index	Value	Sensitivity	Specificity	ρ+	ρ−	Diagnostic accuracy	FP+FN	AUC±SE	95% CI
VC (ml/kg)	>11	0.59	0.79	2.8	0.52	0.68	4+9	0.66±0.08	0.5–0.8
V$_T$ (ml/kg)	>5	0.82	0.53	1.74	0.34	0.68	9+4	0.67±0.08	0.5–0.8
f/V$_T$ (breaths/min per l)	≤100	0.91	0.16	1.1	0.56	0.56	16+2	0.66±0.08	0.5–0.8
P$_{0.1}$ (cmH$_2$O)	≤4	0.95	0.11	1.1	0.45	0.56	17+1	0.65±0.09	0.49–0.79
P$_{0.1}$/MIP	≤0.15	0.55	0.58	1.31	0.78	0.56	8+10	0.49±0.09	0.34–0.66
V$_{MIN}$ (l)	≤12	0.77	0.11	0.86	2.1	0.46	17+5	0.48±0.09	0.32–0.64
RR (breaths/min)	≤35	0.95	0.11	1.1	0.45	0.56	17+1	0.58±0.09	0.41–0.73
MIP (cmH$_2$O)	<−16	0.95	0.42	1.6	0.12	0.71	11+1	0.67±0.08	0.52–0.82
P$_{0.1}$ × f/V$_T$ (cmH$_2$O/ breaths per min per l)	≤300	0.86	0.21	1.1	0.67	0.56	15+3	0.58±0.09	0.41–0.73

Table 5 Area under receiver operating characteristic curve (*AUC*), likelihood ratio (*LR*) and probability to predict weaning outcome according to the prevalence in the prospective-validation set (see Table 4 for explanation of abbreviations). Note that one patient with extubation failure was not included in the prospective-validation set. Formulae for estimation of post-test probabilities (Bayes' theorem): *P(W+)* prevalence of weaning success (pre-test probability); *P(NW-)* prevalence of weaning failure (pre-test probability); *P(T+|W+)* true positive rate (sensitivity); *P(T+|NW-)* false positive rate; *P(T-|NW-)* true negative rate (specificity); *P(T-|W+)* false negative rate. Probability for weaning success if test is positive: $P(W+|T+) = \frac{P(T+|W+) \times P(W+)}{P(T+|W+) \times P(W+) + P(T+|NW-) \times P(NW-)}$ Probability for weaning success if test is negative: $P(W+|T-) = \frac{P(T-|W+) \times P(W+)}{P(T-|W+) \times P(W+) + P(T-|NW-) \times P(NW-)}$

| Index	AUC±SE	95% CI	Sensitivity	Specificity	ρ+	ρ−	LR	P(W+|T+)	P(W+|T-)
VC (ml/kg)	0.71±0.07	0.57–0.83	0.43	0.64	1.19	0.89	1.3	0.76	0.69
V$_T$ (ml/kg)	0.64±0.07	0.49–0.77	0.54	0.57	1.25	0.81	1.54	0.76	0.67
f/V$_T$ (breaths/min per l)	0.70±0.07	0.56–0.82	0.81	0.14	0.94	1.36	0.69	0.71	0.78
P$_{0.1}$ (cmH$_2$O)	0.47±0.09	0.33–0.61	0.94	0.07	1.01	0.86	1.17	0.72	0.64
P$_{0.1}$/MIP	0.71±0.08	0.57–0.83	0.92	0.14	1.07	0.57	1.87	0.73	0.59
V$_{MIN}$ (l)	0.54±0.08	0.39–0.68	0.86	0.14	1	1	1	0.72	0.69
RR (breaths/min)	0.52±0.09	0.37–0.66	0.94	0.07	1.01	0.86	1.17	0.72	0.64
MIP (cmH$_2$O)	0.57±0.09	0.43–0.71	0.92	0.07	0.99	1.14	0.87	0.72	0.75
P$_{0.1}$ × f/V$_T$ (cmH$_2$O/ breaths per min per l)	0.67±0.08	0.53–0.80	0.89	0.07	0.96	1.57	0.61	0.72	0.80

f/Vt
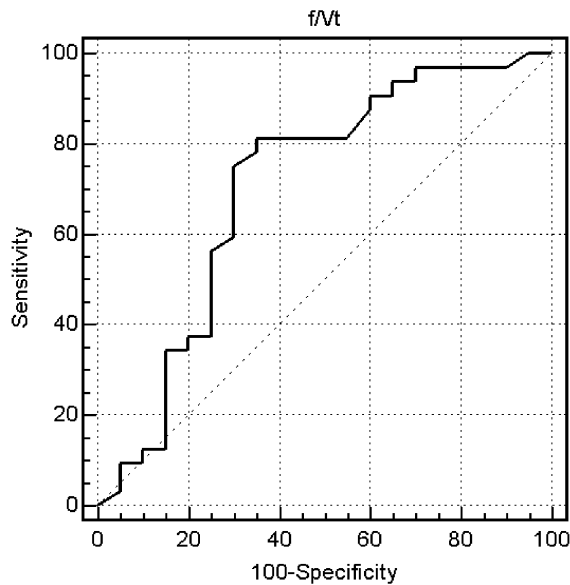
Fig. 1 Receiver operating characteristic curve for respiratory frequency to tidal volume (f/V_T) in the prospective-validation set. Area under the curve ± standard error and 95% confidence interval are given in Table 5

calculated by Bayes' theorem and according to our prevalence of success or failure of weaning.

The AUCs in the prospective-validation set are shown in Table 5: showing that all the evaluated tests appeared to be poor predictors of weaning outcome, as suggested by the 95% confidence interval estimate. In detail, the integrative indexes did not reveal a high ability to distinguish between successful weaning and weaning failure, because the AUC values for maximal inspiratory pressure and $P_{0.1}$ were not significantly different from the area for $P_{0.1}$/MIP ($p=0.72$ and $p=0.07$, respectively). Also the AUC for $P_{0.1} \times f/V_T$ was not different from the areas for f/V_T (Fig. 1) and $P_{0.1}$ ($p=0.52$ and $p=0.16$, respectively) [12].

Discussion

The purpose of weaning indexes is to provide easy discrimination between those patients who can be successfully weaned from mechanical ventilation and those who are unable to be weaned. Many factors can influence the weaning outcome: the functional parameters used as indexes of weaning, the criteria used to define failure or success, the moment at which the patients are studied, different clinical practice from unit to unit and the different populations.

This study included a non-selected population of a general intensive care unit and reflected the activity of our every day clinical practice. Specific care was adopted to avoid the limitation represented by the lack of blinding, a bias frequently observed in previous studies [10]. Our results clearly show that all the evaluated indexes are poor predictors of weaning outcome and are partially different from those previously reported [21, 22, 23, 24].

In the prospective-validation set, likelihood ratios were between 0.61 and 1.87 for all the indexes evaluated. These values indicate that weaning parameters were associated with only small, clinically unimportant changes in the post-test probability of success or failure [9]. Applying the Bayes' theorem in the prospective-validation set, we also found that, given the prevalence of the weaning outcome in this group, all indexes were of little use in discriminating between those patients who could be successfully weaned and those in whom the weaning trial would have failed.

It is also important to emphasize that, in our study, the prevalence of weaning outcome ('a priori' probability) was not only determined by the patient population but also by other factors, including the physician's clinical judgement and the standard protocol used in our intensive care unit.

According to the method proposed by Yang and Tobin [1], we determined the cut-off values by using the receiver operating characteristic curve analysis and selected as the threshold value the one that resulted in the highest accuracy (minimal false negative and false positive results) that is independent of specific cut-off values. This approach assumes that the outcomes related to false positive and false negative are equivalent and do not account for the pre-test probability. This concept is theoretically linked to the receiver operating characteristic curve through the optimality criterion: $S = [(1-P)/P] \times CR$, where P denotes the prevalence in the target population and CR (cost ratio) $= [(C_{FP}-C_{TN})/(C_{FN}-C_{TP})]$ represents the utilities associated with the four possible test outcomes, respectively, and S is the slope of the receiver operating characteristic curve at the optimal operating point [25, 26].

A weakness of this approach is that it requires the users to quantify the consequences of each possible test outcome. The slope approach requires a smoothed function (e.g. binomial distribution), which introduces additional uncertainties. Therefore, to plot the true positive rate (sensitivity) as a function of the false positive rate (100-specificity) for different cut-off points, provides a more practical solution to the problem.

In our patients, none of the indexes investigated appeared to be a good test of screening, as they were all characterized by a high sensitivity and a low specificity. We observed the highest sensitivity and specificity for vital capacity and tidal volume. Minute ventilation showed a high sensitivity and a low specificity because, compared to vital capacity and tidal volume, it had more true positives and fewer true negatives. Vital capacity, tidal volume and minute ventilation showed a high proportion of false positives and false negatives (12+5, 6+17 and 5+21, respectively). The poor predictive value

of such indexes was further supported by the respective values of the AUC and likelihood ratios. We further supported the finding of a poor predictive value for these tests by applying the Bayes' theorem, based on the prevalence of the weaning outcome in this group.

Generally, a low predictive value for a test is observed when the study population is heterogeneous with respect to clinical characteristics and diagnosis (as it was in our study). A low predictive value can also depend on the way measures are taken and the method used to determine the cut-off value for sensitivity and specificity estimates. In our study, by using the receiver operating characteristic curve analysis, we chose the cut-off value which was associated with the smallest number of false positives and negatives.

Another reason with which to explain the low predictive value of these indexes was the use of clinical criteria indicating the need to restart mechanical ventilation during the T-piece trial: these criteria could make respiratory rate and derived parameters ($P_{0.1} \times f/V_T$ and f/V_T) less useful for establishing the proportion of false negatives because, after the first 2 min, a respiratory rate of more than 35 was a sufficient criterion for the attending physician to stop the weaning trial. We considered it unethical to keep a patient in a T-piece and proceed to extubation when clear clinical signs of intolerance were present. Obviously, it was impossible to blind the respiratory rate to the attending physician, because the respiratory rate was used as the clinical criterion for confirming the ability of the patient to sustain spontaneous breathing.

The low discriminative ability of a test may also depend on the method used to take measurements. For example, in our study maximal inspiratory pressure was measured after expiration to functional residual capacity and not to residual volume [27]. The maximal inspiratory pressure mean value for the entire study population was 22 ± 7 (SW: 24 ± 7 and FW: 20 ± 6 cmH$_2$O, respectively). Such a relatively low value may be explained by the severity and old age of our case mix, which included many patients with COPD, ALI/ARDS and neurological disorders. Moreover, the group of patients with postoperative respiratory failure included only patients who underwent emergency surgery. P_{01} also was not a good test for screening, because it showed a high sensitivity (0.94) and a low specificity (0.07): these data were confirmed by using the Bayes' theorem.

According to a recent review [28], a distinction between weaning failure (inability to tolerate spontaneous breathing without ventilatory support) and extubation failure (inability to tolerate removal of the translaryngeal tube) has been increasingly recognized. This analysis was not made in our study because only one patient required intubation within 24 h of extubation, after 2 h of spontaneous breathing.

Two arguments can explain our low reintubation rate. First, a spontaneous breathing trial (T-piece) can yield a low reintubation rate [29]. Recent studies [30] have shown that almost 76% of ventilated patients can be extubated after a 2-h spontaneous breathing trial. Moreover, we included clinical signs indicating respiratory muscle capacity and load imbalance, such as uncoordinated thoracoabdominal movements and activation of the accessory muscles of respiration as criteria for spontaneous breathing intolerance. The use of these criteria, along with the traditional criteria for monitoring a spontaneous breathing trial [11], made it easier to identify those patients who presented early signs of increased muscle load. In this way, probably, the prevalence of extubation failure was underestimated in favor of the prevalence of weaning failure.

Finally, the low ability of the evaluated tests to discriminate successful weaning and weaning failure can also be explained by the fact that they represent only a static measure, collected at a specific moment, whereas weaning is a dynamic process during which the physiologic variable measured is continuously influenced by the patient's clinical condition.

On account of our results and those from other recent studies [29, 31], we suggest that weaning should be based on clinical evaluation and strict protocols, and that the use of predictive tests can poorly corroborate clinical judgment.

In conclusion, even when the methodological limitation represented by the lack of blinding of the physicians making decisions about the weaning process is avoided, none of the predictors of weaning studied is powerful enough to predict success: the systematic use of these weaning "predictors" is thus of little use clinically.

References

1. Yang KL, Tobin MJ (1991) A prospective study of indexes predicting the outcome of trials of mechanical ventilation. N Engl J Med 324:1445–1450

2. Sassoon CSH, Te TT, Mahutte CK, Light RW (1987) Airway occlusion pressure: an important indicator for successful weaning in patients with chronic obstructive pulmonary disease. Am Rev Respir Dis 135:107–113

3. Conti G, De Blasi R, Pelaia P, Benito S, Rocco M, Antonelli M, Bufi M, Mattia C, Gasparetto A (1992) Early prediction of successful weaning during pressure support ventilation in chronic obstructive pulmonary disease patients. Crit Care Med 20:366–371

4. Zeggwagh AA, Abouqal R, Madani N, Zekraoui A, Kerkeb O (1999) Weaning from mechanical ventilation: a model for extubation. Intensive Care Med 25:1077–1083

5. Sassoon CSH, Mahutte CK (1993) Airway occlusion pressure and breathing pattern as predictors of weaning outcome. Am Rev Respir Dis 148:860–866

6. Epstein SK (1995) Etiology of extubation failure and the predictive value of the rapid shallow breathing index. Am J Respir Crit Care Med 152:545–549

7. Esteban A, Alìa I, Ibanez J, Benito S, Tobin MJ (1994) Models of mechanical ventilation and weaning. Chest 106:1188–1193

8. Esteban A, Alìa I, Tobin MJ, Gil A, Gordo F, Vallverdu I, Blanch L, Bonet A, Vazquez A, de Pablo R, Torres A, De La Cal MA, Macias S (1999) Effect of spontaneous breathing trial duration on outcome of attempts to discontinue mechanical ventilation. Am J Respir Crit Care Med 159:512–518

9. A collective task force facilitated by the American College of Chest Physicians; the American Association for Respiratory Care and the American College of Critical Care Medicine (2001) Evidence-based guidelines for weaning and discontinuing ventilatory support. Chest 120:375S–395S

10. Brochard L, Rauss A, Benito S, Conti G, Mancebo J, Rekik N, Gasparetto A, Lemaire F (1994) Comparison of three methods of gradual withdrawal from mechanical ventilation. Am J Respir Crit Care Med 150:896–903

11. Esteban A, Alìa I, Gordo F, Fernandez R, Solsona JF, Vallverdu I, Macias S, Allegue JM, Blanco J, Carriedo D, Palazon E, Carrizosa F, Tomas R, Suarez J, Goldwasser RS (1997) Extubation outcome after spontaneous breathing trials with T-tube or pressure support ventilation. Am J Respir Crit Care Med 156:459–465

12. ROC curve analysis in MedCalc (2001) MedCalc software version 6.10.001, Belgium

13. Swets JA (1988) Measuring the accuracy of diagnostic systems. Science 240:1285–1293

14. Greiner M, Pfeiffer D, Smith RD (2000) Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Prev Vet Med 45:23–41

15. Cook D, Meade M, Guyatt G (1999) Evidence report on criteria for weaning from mechanical ventilation. Rockville, MD: Agency for Health Care Policy and Research, pp 1–35

16. Biggerstaff BJ (2000) Comparing diagnostic tests: a simple graphic using likelihood ratios. Stat Med 19:649–663

17. Jaeschke R, Guyatt GH, Sackett DL (1994) Users' guides to the medical literature: how to use an article about a diagnostic test; are the results of the study valid? JAMA 271:389–391

18. Jaeschke R, Guyatt GH, Sackett DL (1994) Users' guides to the medical literature: what are the results and will they help me in caring for my patients? JAMA 271:703–707

19. Irwing L, Tosteson ANA, Gatsonis C, Lau J, Colditz G, Chalmers TC, Mosteller F (1994) Guidelines for meta-analyses evaluating diagnostic test. Ann Intern Med 120:667–676

20. Pagano M, Gauvreau K (2000) Principles of Biostatistic. In: Duxbury (ed) Thomson learning 2nd edn, pp 136–140

21. Meade M, Guyatt G, Cook D, Griffith L, Sinuff T, Kergl C, Mancebo J, Esteban A, Epstein S (2001) Predicting success in weaning from mechanical ventilation. Chest 120:400S–424S

22. Del Rosario N, Sassoon CS, Chetty KG (1997) Breathing pattern during acute respiratory failure and recovery. Eur Respir J 10:2560–2565

23. Yang KL (1993) Inspiratory pressure/maximal inspiratory pressure ratio: a predictive index of weaning outcome. Intensive Care Med 19:204–208

24. Vallverdù I, Calaf N, Subirana M, Net A, Benito S, Mancebo J (1998) Clinical characteristics, respiratory functional parameters and outcome of a two-hour T-piece trial in patients weaning from mechanical ventilation. Am J Respir Crit Care Med 58:1855–1862

25. Metz CE (1978) Basic principles of ROC analysis. Semin Nucl Med 8:283–298

26. Smith RD (1995) Evaluation of diagnostic test. In: Veterinary clinical epidemiology. A problem-oriented approach. CRC Press, Boca Raton, pp 31–43

27. Rochester DF (1988) Test of respiratory muscle function. Clin Chest 9:249–261

28. Epstein SK (2002) Decision to extubate. Intensive Care Med 28:535–546

29. Ely EW, Baker AM, Dunagan DP, Burke HL, Smith AC, Kelly PT, Johnson MM, Browder RW, Bowton DL, Haponik EF (1996) Effect on the duration of mechanical ventilation of identifying patients capable of breathing spontaneously. N Engl J Med 335:1864–1869

30. Esteban A, Frutos F, Tobin MJ, Alia I, Solsona JF, Valverdù I, Fernandez R, De La Cal MA, Benito S, Tomas R, Carriedo D, Macias S, Blanco J for the Spanish Lung Failure Collaborative Group (1995) A comparison of four methods of weaning patients from mechanical ventilation. N Engl J Med 332:345–350

31. Ely EW, Meade MO, Haponik EF, Kollef MH, Cook DJ, Guyatt GH, Stoller JK (2001) Mechanical ventilator weaning protocols driven by nonphysician health-care professionals. Evidence-based clinical practice guidelines. Chest 120:454S–463S