Intensive Care Med (2006) 32:2002–2012
DOI 10.1007/s00134-006-0439-4

ORIGINAL

Martin J. Tobin
Amal Jubran

Variable performance of weaning-predictor tests: role of Bayes' theorem and spectrum and test-referral bias

Received: 3 March 2006
Accepted: 6 October 2006
Published online: 8 November 2006
© Springer-Verlag 2006

Electronic supplementary material
Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00134-006-0439-4 and is accessible for authorized users.

This article is discussed in the editorial available at: http://dx.doi.org/10.1007/s00134-006-0440-y

This work was supported by a Merit Review grant from the Veterans Administration Research Service and by the National Institutes of Health (RO1 NR008782)

M. J. Tobin (✉) · A. Jubran
Loyola University of Chicago, Division of Pulmonary and Critical Care Medicine, Edward Hines Jr. Veterans Affairs Hospital, and Stritch School of Medicine,
5th Avenue and Roosevelt Road, Hines 60141, IL, USA
e-mail: mtobin2@lumc.edu
Tel.: +1-708-2022705
Fax: +1-708-2027907

Abstract *Objective:* We examined whether variation in reported reliability of the frequency-to-tidal volume ratio (f/V_T) in predicting weaning success is explained by spectrum and test-referral bias, as reflected by variation in pretest probability of success. *Design:* Two authors extracted data from all studies on reliability of f/V_T as a weaning predictor. *Results:* Prevalence of successful weaning in studies of f/V_T revealed significant heterogeneity; mean success rate was 0.75. The heterogeneity and high success rate reflects occurrence of spectrum bias, suggested by the lower value of f/V_T in subsequent studies than in the original report (77.4 vs. 89.1) and test-referral bias, suggested by lower specificity of f/V_T in subsequent studies than in the original report (0.52 vs. 0.64). When data from studies in the ACCP Task Force's meta-analysis of studies on f/V_T were entered into a Bayesian model with pretest probability (prevalence of success) as the operating point, observed posttest probabilities were closely correlated with values predicted by the original report on f/V_T: positive-predictive value $r = 0.86$ and negative-predictive value $r = 0.82$. Average sensitivity, the most precise measure of screening-test reliability, was 0.87 ± 0.14 and average specificity 0.52 ± 0.26. *Conclusions:* Much of the heterogeneity in performance of f/V_T can be explained by variation in pretest probability of successful outcome, which may be secondary to spectrum and test-referral bias. The average sensitivity of 0.87 indicates that f/V_T is a reliable screening test for successful weaning.

Keywords Mechanical ventilation · Weaning · Monitoring · Clinical decision making · Diagnostic testing · Breathing pattern

Introduction

The hazards of mechanical ventilation make it imperative to disconnect patients from the ventilator at the earliest feasible time [1, 2, 3, 4, 5, 6]. Studies, however, indicate that clinicians are slow to recognize a patient's ability to tolerate ventilator weaning [7, 8]. Psychology research has shown that delays in decision making result from over reliance on heuristics and insufficient attention to prior probability [9]. Minimizing delay in diagnosis is the primary reason that screening tests are performed [10, 11, 12]. To attain maximal benefit screening tests should be performed when the prior (pretest) probability is very low (ideally $< 20\%$) [10, 11, 12]. The tests used to screen for readiness to tolerate ventilator discontinuation are weaning-predictor tests [13].

Recently an Evidence-Based Medicine Task Force of the American College of Chest Physicians (ACCP) [14, 15] evaluated the usefulness of weaning-predictor tests using a meta-analysis. The ACCP Task Force focused predominantly on the weaning-predictor test that has been most frequently studied (> 25 studies): the ratio of

frequency-to-tidal volume (f/V_T), a measure of rapid shallow breathing [13, 16]. The Task Force calculated pooled likelihood ratios for f/V_T and judged the summed values to signify that f/V_T is not a reliable predictor of weaning outcome. The Task Force concluded that physicians should bypass measurement of all weaning-predictor tests and begin the weaning process with a trial of spontaneous breathing.

When assessing the reliability of weaning-predictor tests, it is critically important to recognize that weaning procedures constitute a form of diagnostic testing. Consequently evaluation of their reliability must comply with the canons developed for evaluating diagnostic tests [10, 11, 12]. In the assessment of published reports of weaning-predictor tests, the element most often ignored is the enormous influence of pretest probability on the test results. In their textbook on medical decision analysis Sox and colleagues [12] state, "Perhaps the most important idea in this book is the following: The interpretation of a test result depends on the pretest probability of disease." The importance of this point is heightened whenever research is carried out on a diagnostic test that has already been accepted by clinicians and incorporated into their everyday practice [17].

The implications of pretest probability are greater for weaning than for many clinical situations because weaning involves a sequence of three diagnostic tests: measurement of predictors, followed by a weaning trial, followed by an extubation trial. The undertaking of three diagnostic tests in a sequential manner poses an enormous risk for the occurrence of test-referral bias [11, 12]. Test-referral bias arises when a test under evaluation (weaning-predictor test) influences which patients undergo either of the two subsequent tests. If tolerance of extubation is used as the gold standard for evaluating the reliability of the weaning-predictor test, the requirement to pass a weaning trial (e.g., T-tube trial) before extubation necessarily excludes all patients who fail a weaning trial. The study population is thereby skewed towards less severely ill patients, an effect termed spectrum bias [18]. This step not only alters pretest probability. It also alters both the sensitivity and specificity of weaning-predictor test [11, 12].

Failure to take into account the effects of spectrum and test-referral bias on pretest probability leads to fundamental misinterpretation of the reliability of weaning-predictor tests. We hypothesized that much of the variation among studies that have evaluated the reliability of f/V_T in predicting weaning outcome is explained by spectrum and test-referral bias, as reflected by variation in pretest probability of successful outcome. We further hypothesized that once variation in pretest probability among subsequent studies of f/V_T is taken into account, these studies confirm the sensitivity and specificity reported in the original 1991 study on f/V_T.

Methods

All articles included in the meta-analysis of the ACCP Task Force were retrieved [13, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39]. In five of these articles [18, 23, 26, 31, 32] the authors did not report data on pretest probability, sensitivity, and specificity. Beyond articles included in the ACCP Task Force's meta-analysis we retrieved additional articles [40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50] via Medline search of studies published up to May 2005 and by search of personal files. The studies evaluated here are listed in Table 1. The two present authors examined the full text of all articles. The following data were abstracted from each: number of patients studied, definition of study endpoint (toleration of weaning trial, extubation trial, or combination), threshold value of f/V_T, sensitivity, specificity, positive-predictive value, negative-predictive value, prevalence of successful outcome, and whether primary clinicians were blinded to the data. Investigators varied in the efforts they took to blind physicians to f/V_T measurements; most made no explicit attempt.

When an "average" statistic is computed from a meta-analysis, erroneous interpretations can arise if there is significant heterogeneity among the included studies [51, 52]. When heterogeneity is significant, it is recommended to search for a factor that may be acting as an effect modifier [51]. We accordingly investigated whether the heterogeneity in the pretest probability of successful outcome among 20 studies included in the ACCP Task Force's meta-analysis was significant by means of χ^2 analysis [53]. We subsequently show that the heterogeneity in pretest probability is significant, and that this may arise from test-referral bias and spectrum bias consequent to the sequential nature of diagnostic testing during weaning.

Bayes' theorem is an equation that describes the relationship between a physician's initial clinical gestalt of the probability of a particular condition (pretest probability) and the physician's revised probability after obtaining the result of a diagnostic test (posttest probability) [11, 54, 55]. It is used to estimate how much the uncertainty of weaning outcome changes from before measurement of a weaning-predictor test (pretest probability) to after obtaining the new information (conditional probability) [12]. In particular, Bayes' theorem is used to transform the information contained in sensitivity and specificity into a format that can be employed in diagnostic testing (calculation of posttest probability, in the format of positive- and negative-predictive value) [11].

To determine the influence of spectrum bias and test-referral bias on the reported posttest probabilities of f/V_T we used pretest probability as an indirect measure of these two biases. In everyday practice a clinician's pretest probability of a clinical outcome is his or her clinical gestalt. When applying a Bayesian framework to the evaluation of studies of diagnostic tests, prevalence of the outcome

Table 1 Accuracy of f/V_T in predicting weaning or extubation outcome. The listed studies are those that report data on the accuracy of f/V_T as a predictor of weaning outcome. Four studies (nos. 9, 16, 17, and 18) report data under two different conditions in their articles; both sets of data are presented. Pretest probability of success in a study is the fraction of patients with a successful outcome out of the total population (both success and failure patients) included in the study (*f/V*$_T$ frequency-to-tidal volume ratio, *PPV* positive-predictive value, *NPV* negative-predictive value, *WF* weaning failure, *EF* extubation failure, *PS* pressure support, *IMV* intermittent mandatory ventilation, *bpm* breaths per min, *MICU* medical ICU, *RICU* respiratory ICU, *Md ICU* multidisciplinary ICU, *SICU* surgical ICU, *PICU* pediatric ICU, *M-SICU* medical-surgical ICU, *CCU* cardiac care unit, *NS* nonstated location, *NLR* negatve likehood ratio, *PLR* positive likehood ratio)

Study no.	Authors	*n*	Outcome endpoint	Threshold	Sensitivity	Specificity	PPV	NPV	PLR	NLR	Pretest prob. of success	Data avail. to primary	Location
1	Yang and Tobin [13]	64	WF or EF	≤ 105	0.97	0.64	0.78	0.95	2.69	0.05	0.56	No	MICU
2	Gandia and Blanco [35]	40	WF or EF	<96	0.89	0.83	0.93	0.77	5.24	0.13	0.7	No	NS
3	Sassoon and Mahutte [20]	45	WF or EF	≤ 100	0.97	0.40	0.85	0.80	1.62	0.08	0.78	No	NS
4	Yang [24]	31	WF or EF	≤100	0.94	0.73	0.79	0.92	3.48	0.08	0.52	No	MICU
5	Mohsenifar et al. [38]	29	WF or EF	≤105(PS 7–8)	1.00	0.27	0.69	1.00	1.37	0.00	0.62	Not clear	RICU
6	Lee et al. [28]	52	EF only	≤105(PS?)	0.72	0.11	0.79	0.08	0.81	2.55	0.83	Yes	MICU
7	Capdevila et al. [34]	67	EF only	60	0.73	0.75	0.92	0.36	2.92	0.36	0.82	Yes	Md ICU
8	Epstein [25]	94	EF only	< 100	0.92	0.22	0.83	0.40	1.18	0.36	0.81	Yes	MICU
9a	Chatila et al. [21]	100	WF or EF	≤100	0.89	0.41	0.72	0.68	1.51	0.27	0.63	Yes	MICU, CCU
9b	Chatila et al. [21]	100	WF or EF	≤100	0.98	0.59	0.83	0.94	2.39	0.03	0.63	Yes	MICU, CCU
10	Dojat et al. [22]	38	WF or EF	< 100	0.94	0.81	0.80	0.94	4.95	0.07	0.45	No	MICU, SICU
11	Leitch et al. [27]	163	EF only	≤100(PS 7)	0.96	0.00	0.98	0.00	0.96	>2.55	0.982	Yes	M-SICU
12	Mergoni et al. [29]	75	WF or EF	< 105	0.65	0.58	0.60	0.63	1.54	0.61	0.49	Yes	M-SICU
13	Bouachour et al. [40]	15	WF only	≤105	1.00	0.40	0.77	1.00	1.67	0.00	0.67	Not clear	NS
14	Baumeister et al. [33]	47 Ped	EF only	≤11 bpm/ml per kg	0.79	0.78	0.94	0.47	3.59	0.27	0.81	No	PICU
15	Gologorskii et al. [31]	127	Not defined	Not stated	0.84	0.83	0.80	0.86	–	–	–	Not clear	SICU
16a	Jacob [36]	183	WF or EF	100	0.97	0.33	0.94	0.50	1.45	0.09	0.92	Yes	SICU
16b	Jacob [36]	183	WF or EF	100	0.96	0.31	0.94	0.40	1.39	0.13	0.92	Yes	SICU
17a	Kreiger [30]	49	WF	≤105	0.74	0.73	0.90	0.44	2.74	0.36	0.78	Yes	MICU
17b	Kreiger [30]	49	WF	≤130 at 3 h	0.93	0.89	0.97	0.80	8.45	0.08	0.78	Yes	MICU
18a	Rivera and Weissman [41]	40	WF only	65 (PS 5)	0.90	0.80	0.90	0.70	4.50	0.13	0.7	Yes	SICU
18b	Rivera and Weissman [41]	40	WF only	65 (PS +IMV)	1.00	0.82	0.84	1.00	5.56	0.00	0.7	Yes	SICU
19	Farias et al. [39]	84 Ped	WF	≤11 bpm/ml per kg	0.48	0.86	0.53	0.83	3.43	0.60	0.75	No	PICU
20	Vallverdu et al. [37]	217	WF or EF	≤100	0.90	0.36	0.66	0.73	1.41	0.28	0.58	Yes	M-SICU
21	Thiagarajan et al. [42]	227 Ped	EF only	≤8 bpm/ml per kg	0.74	0.74	0.97	0.22	2.85	0.35	0.89	No	PICU
22	Zeggwagh et al. [43]	101	EF only	<88	0.77	0.79	0.68	0.86	3.67	0.29	0.63	No	MICU
23	Maldonado et al. [44]	27	WF only	≤105	0.93	0.75	0.83	0.89	3.72	0.09	0.56	Yes	RICU
24	Uusaro et al. [45]	68	EF only	< 100 (PS 5)	0.96	0.18	0.78	0.60	1.17	0.22	0.75	Yes	M-SICU
25	Khamiees et al. [46]	100	EF only	≤105	0.84	0.17	0.82	0.19	1.01	0.94	0.82	Yes	MICU, CCU
26	Smina et al. [47]	115	EF only	< 100	0.90	0.42	0.92	0.36	1.55	0.24	0.89	Yes	MICU, CCU
27	Conti et al. [48]	51	WF or EF	≤100	0.81	0.14	0.71	0.22	0.94	1.36	0.73	No	MICU, SICU
28	Fernandez et al. [49]	57	WF; EF	< 50	0.35	0.56	0.81	0.14	0.80	1.16	0.84	Yes	MICU, SICU
29	Jiang et al. [50]	55	EF only	≤105	0.81	0.57	NR	NR	1.88	0.33	0.58	Yes	MICU

under investigation is used as a surrogate for the pretest probability [56, 57, 58, 59, 60, 61]. Accordingly, we calculated pretest probability as the prevalence of successful outcome divided by the sum of patients with a successful and unsuccessful outcome in a study. (The conclusions of our study would remain the same if the term "pretest probability" were deleted and replaced by "prevalence of successful outcome." We choose to frame the analysis in terms of "pretest probability" because it is a more intuitive expression when conducting a Bayesian analysis, and because the two terms (pretest probability and prevalence of the condition) are used interchangeably in writings on diagnostic testing [56, 57, 58, 59, 60, 61].)

To assess whether subsequent studies of f/V_T reproduce the sensitivity and specificity reported in the original study on f/V_T [13], we used Bayes' theorem. The framework for this portion of the data analysis was based on the true-positive rate (sensitivity 0.97) and false-positive rate (1–specificity 0.36) in the original report [13]. Using these data and the formulae below (based on Bayes' theorem [12]), we calculated the posttest probability of f/V_T (positive-predictive value and negative-predictive value) for 0.01-unit increments in pretest probability between 0.00 and 1.00:

$$\mathrm{PPV} = (\mathrm{PPS} \times \mathrm{TPR})/\{(\mathrm{PPS} \times \mathrm{TPR}) + [(1 - \mathrm{PPS}) \times \mathrm{FPR}]\} \quad (1)$$

and

$$\mathrm{NPV} = [(1 - \mathrm{PPS}) \times \mathrm{TNR}]/\{[(1 - \mathrm{PPS}) \times \mathrm{TNR}] + (\mathrm{PPS} \times \mathrm{FNR})\}, \quad (2)$$

where PPV = positive-predictive value, PPS = pretest probability of success, TPR = true-positive rate, FPR = false-positive rate, TNR = true-negative rate, and FNR = false-negative rate. The resulting values of positive- and negative-predictive value (which we refer to as the predicted values) were plotted against pretest probability. The upper and lower 95% confidence intervals were then calculated and superimposed on the plots [62].

We checked each study for internal consistency. We took the reported sensitivity, specificity, and pretest probability of success and entered them into the above formulae. All but two studies [39, 43] showed good internal consistency. Zeggwagh et al. [43] reported a positive-predictive value of 0.68 and a negative-predictive value of 0.86; we calculated respective values of 0.86 and 0.67. Farias et al. [39] reported a positive-predictive value of 0.53 and a negative-predictive value of 0.83; we calculated respective values of 0.91 and 0.36. Because of these inconsistencies we excluded these two studies from further data analysis. (The conclusions of our study would not change if these data [39, 43] were included.)

The values of negative-predictive value and positive-predictive value reported in the each study of f/V_T were entered on the above plots. We examined the influence of pretest probability of success on positive-predictive value and negative-predictive value of f/V_T because those relationships are explicated by Bayes' theorem. In contrast, an equivalent governing framework to encompass the relationships between pretest probability and sensitivity and specificity (and thus likelihood ratio) has not been developed, and it seems unlikely that one can be developed.

We used a weighted Pearson's correlation analysis (adjusting for the number of patients contained in a study) to compare the relationship between pretest probability (prevalence of success) and reported values of positive- and negative-predictive value. Secondly, we used a weighted Pearson's correlation analysis to compare the relationship between the predicted values of positive- and negative-predictive value and the actual values reported in each study. Thirdly, we undertook a Bland-Altman analysis to determine whether the reported values of positive- and negative-predictive value fall within the 95% confidence intervals of the values predicted by entering (reported) pretest probabilities into Eqs. 1 and 2.

Results

Pretest probability of successful outcome in the studies included in Table 1 varied from 0.45 to 0.98. For studies

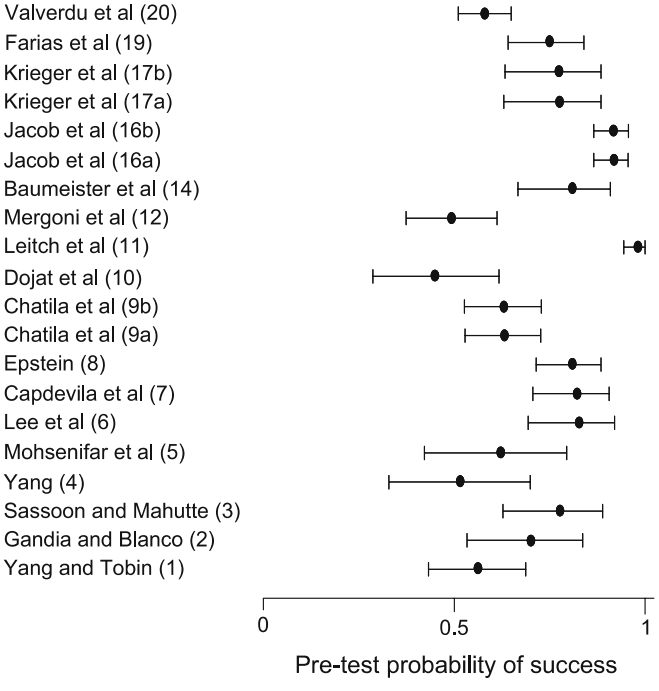

Fig. 1 Pretest probability of successful outcome for studies included in the ACCP Task-Force's meta-analysis. *Error bars* 95% confidence intervals. *Numbering of studies* corresponds to that in Table 1 and not to that in the references. The heterogeneity in pretest probability of success is statistically significant ($p < 0.00001$)

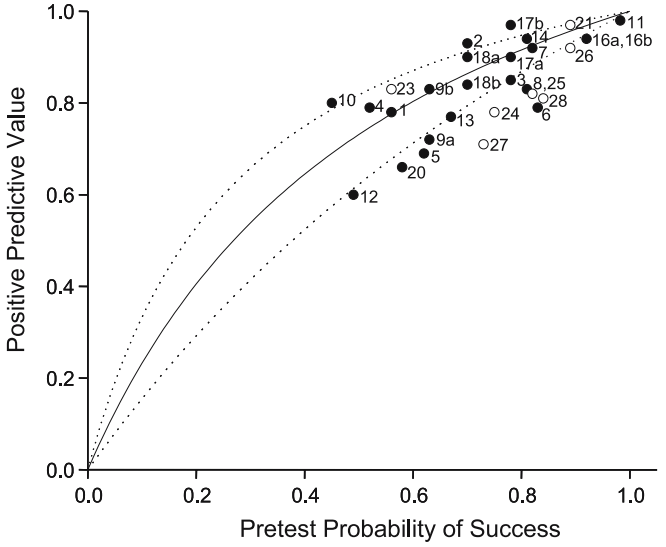

Fig. 2 Positive-predictive value (posttest probability of successful outcome) for f/V_T plotted against pretest probability of successful outcome. *Closed symbols* studies included in ACCP Task Force meta-analysis; *open symbols* additional studies (see Methods). The curve is based on the sensitivity, specificity originally reported by Yang and Tobin [13] and Bayes' formula for 0.01-unit increments in pretest probability between 0.00 and 1.00 [12]. The *lines* represent the upper and lower 95% confidence intervals for the predicted relationship of the positive predictive values against pretest probability. The observed positive-predictive value in a study is plotted against the pretest probability of weaning success (prevalence of successful outcome). *Numbering of studies* corresponds to that in Table 1 and not to that in references. Study nos. 5 [38], 6 [28], 11 [27], 18a [41], 18b [41], and 24 [45] include measurements of f/V_T obtained during pressure support; nos. 14 [33] and 21 [42] include measurements obtained in pediatric patients; nos. 7 [34], 18a [41], 18b [41], and 28 [49] used f/V_T threshold values less than 65

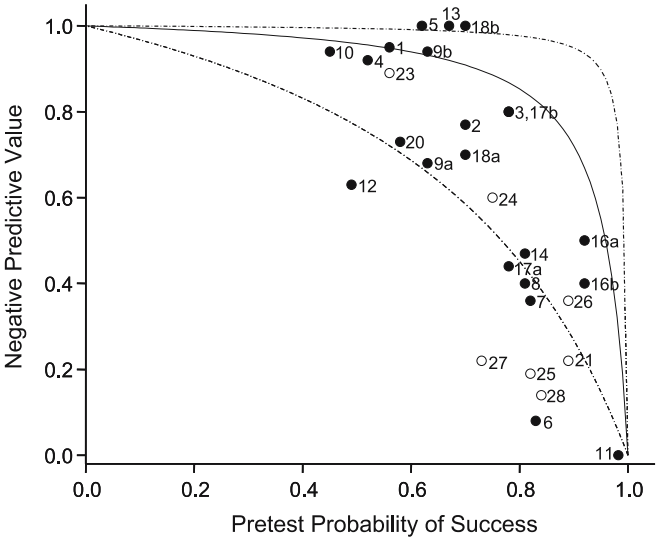

Fig. 3 Negative-predictive value (posttest probability of unsuccessful outcome) for f/V_T. *Closed symbols* studies included in ACCP Task Force meta-analysis; *open symbols* additional studies (see Methods) are indicated by. The curve, its 95% confidence intervals, and placement of a study on the plot are described in the legend to Fig. 2. The observed negative-predictive value in a study is plotted against the pretest probability of weaning success (prevalence of successful outcome). *Numbering of studies* corresponds to that in Table 1 and not to that in the references. (See legend to Fig. 2 for the numbering of studies that include measurements of f/V_T during pressure support, in pediatric patients, or operating at a threshold value below 65.) Study no. 11 [27] has a negative-predictive value of 0.00 and specificity of 0.00, which are predictable given its pretest probability of weaning success of 98.2%; the large number of subjects ($n = 163$) means that this study made a substantial contribution to the pooled likelihood ratio calculated in the meta-analysis of the ACCP Task Force

included in the ACCP Task Force's meta-analysis, pretest probability of successful outcome was 0.75 ± 0.15. The studies included in the ACCP Task Force's meta-analysis also demonstrated a significant degree of heterogeneity in the pretest probability of success ($\chi^2 = 227.1$, df 19, $p < 0.00001$; Fig. 1). Actual values of f/V_T were listed in 15 studies [22, 24, 29, 34, 35, 36, 37, 40, 41, 44, 46, 47, 49, 50]. The mean value was lower in these studies than in the original report [13], 77.4 ± 25.2 vs. 89.1, providing evidence for the occurrence of spectrum bias.

Reported specificity ranged from 0.00 to 0.89 (Table 1), with a mean of 0.52 ± 0.26 (excluding two studies with inconsistent data [39, 43]). The lower specificity in subsequent reports than in the original report on f/V_T [13], 0.64, provides evidence for the occurrence of test-referral bias. Reported sensitivity ranged from 0.35 to 1.00 (Table 1), with a mean of 0.87 ± 0.14. Test-referral bias is also expected to produce an increase in sensitivity over that originally reported. The sensitivity of 0.97 in the original report on f/V_T [13] approaches the ceiling of 1.00, not allowing much room to detect a further increase (allowing for usual biological noise created in any experiment). Seventeen subsequent studies [20, 21, 22, 24, 25, 27, 30, 36, 37, 38, 40, 41, 44, 45, 47] reveal sensitivity values for f/V_T at least 0.90, a finding consistent with test-referral bias.

Positive- and negative-predictive values of f/V_T and pretest probability of success were reported by 27 investigators; four groups [21, 30, 36, 41] evaluated reliability of f/V_T under two sets of conditions. The range in reported reliability was wide: negative-predictive values range from 0.00 to 1.00 and positive-predictive values range from 0.53 to 0.98 (Table 1). The reported positive-predictive for f/V_T was correlated with pretest probability of successful outcome ($r = 0.69$, $p < 0.0001$); likewise, the reported negative-predictive for f/V_T was correlated with pretest probability of successful outcome ($r = -0.75$, $p < 0.0001$).

Figures 2 and 3 show that most of the positive- and negative-predictive values in the studies fall close to or above the lower 95% confidence intervals of the values predicted by Bayes' theorem for pretest probability (using the sensitivity and specificity originally

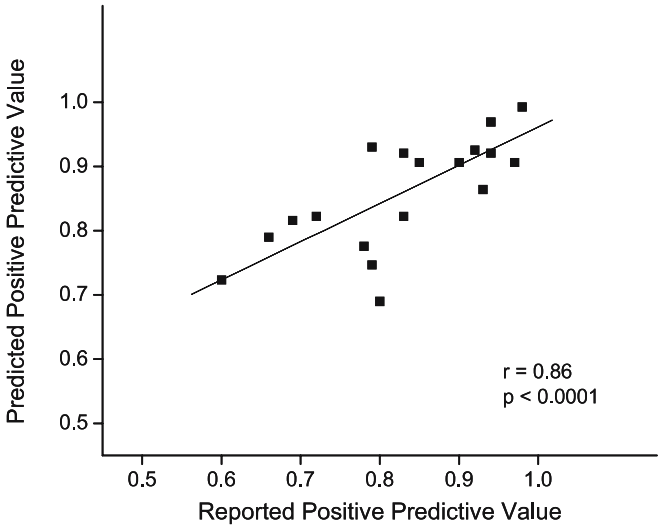

Fig. 4 The relationship between the reported values of positive-predictive value among the studies included in the ACCP Task Force's meta-analysis and the values predicted by observed pretest probability together with the sensitivity and specificity originally reported by Yang and Tobin [13]. The weighted Pearson's correlation is r=0.86 ($p < 0.0001$)

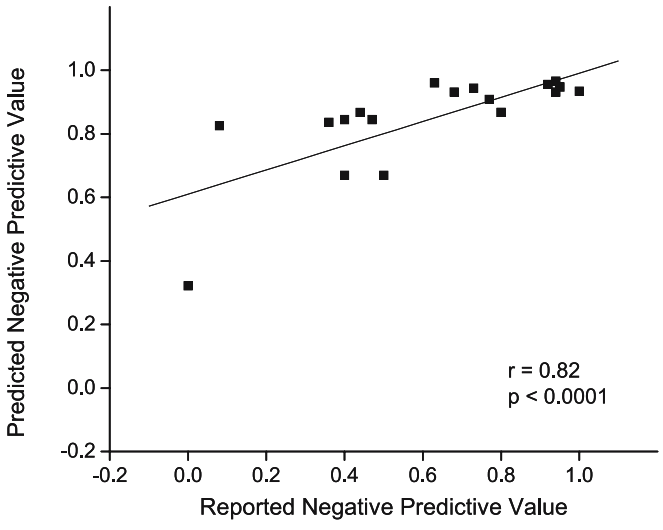

Fig. 5 The relationship between the reported values of negative-predictive value among the studies included in the ACCP Task Force's meta-analysis and the values predicted by observed pretest probability together with the sensitivity and specificity originally reported by Yang and Tobin [13]. The weighted Pearson's correlation is $r = 0.82$ ($p < 0.0001$)

reported by Yang and Tobin [13]). For the studies included in the ACCP Task Force's meta-analysis the correlation between reported and predicted positive-predictive value was $r = 0.86$ ($p < 0.0001$, Fig. 4); the correlation between reported and predicted negative-predictive values was $r = 0.82$ ($p < 0.0001$, Fig. 5). For the entire group of 29 studies the correlation between reported and predicted positive-predictive value was $r = 0.67$ ($p < 0.0001$), and that between reported and predicted negative-predictive value was $r = 0.66$ ($p < 0.0001$).

A Bland-Altman analysis was undertaken to determine the extent of agreement between reported values of positive- and negative-predictive value and the values predicted by (reported) pretest probability together with the sensitivity and specificity originally reported by Yang and Tobin [13]. For the studies included in the ACCP Task Force's meta-analysis, all of the reported positive-predictive values and all but two of the reported negative-predictive values fell within the 95% confidence interval of the values predicted. For the entire group of 29 studies, all of the reported positive- and negative-predictive values fell within the 95% confidence interval of the values predicted.

Discussion

The ACCP Task Force concluded that f/V_T is not a reliable predictor of weaning success based on their meta-analysis of likelihood ratios. For a meta-analysis to be statistically valid, however, it must be free of significant heterogeneity (or control for it) [51, 52]. Figure 1 reveals marked heterogeneity ($p < 0.00001$) in pretest probability of successful outcome among studies in the meta-analysis. This heterogeneity in pretest probability accounts for most of the variation in reported reliability of f/V_T. Once these data are entered into a Bayesian model with pretest probability as the operating point, the reported positive-predictive values are significantly correlated with the values predicted by the original report on f/V_T [13], $r = 0.86$ ($p < 0.0001$); likewise, reported negative-predictive values are correlated with the values predicted, $r = 0.82$ ($p < 0.0001$) (Figs. 4, 5). Moreover, the rate of successful outcome was 75% or higher in more than half the studies, reflecting the influence of spectrum bias and test-referral bias (Table 1). Thus the low values of likelihood ratios for f/V_T reported by the Task Force are largely explained by their failure to correct for the occurrence of spectrum bias and test-referral bias.

A more fundamental conceptual problem arises with the ACCP Task Force's evaluation strategy. Their meta-analysis is not focused on the goal that a weaning-predictor test is designed to meet: to detect the earliest time a patient might tolerate a weaning trial. That is, a weaning-predictor test serves solely as a screening test. As discussed below, the most precise tool for evaluating screening-test reliability is sensitivity [11]. In contrast, the Task Force based their entire evaluation on likelihood ratio. Likelihood ratio, however, is not precisely suited to screening-test evaluation because it includes test compo-

nents vital for screening (true-positive and false-negative rates) but also components not directly focused on screening (true-negative and false-positive rates); the latter cloud the contribution of the vital components [11].

Bayes' theorem and reliability of weaning predictors

Bayes' theorem uses new information (the conditional probability) to update old information (the pretest probability) [12]. Conditional probability refers to the probability that a particular event will occur (a patient will tolerate a weaning trial) given that some other condition has been met (obtaining a positive result on a weaning-predictor test) [11]. The updated result is termed the posttest probability, expressed as positive (or negative) predictive value. According to Bayes' theorem, three factors determine posttest probability: pretest probability, sensitivity, and specificity (of a weaning-predictor test) [11, 12].

Figures 2 and 3 convey the relationship between pretest probability and posttest probability of weaning success based on the theoretical framework of Bayes' theorem. The weighted Pearson's correlation analysis reveals that these two variables were closely related ($p < 0.0001$). For studies in the ACCP Task Force's meta-analysis pretest probability explained 74% of the variation in positive-predictive value and 62% of the variation in negative-predictive value of f/V_T. (The remaining variation in the relationship between pretest probability and posttest probability among the studies probably resulted from population differences, differences in instrumentation, measurements during pressure support, and random variation.)

The information presented in Figs. 2 and 3 represents the interaction between two conceptual models. The overall map is generated by means of Bayes' theorem; the specific contour (interrupted) lines enclosing predicted values of posttest probability for every possible pretest probability is generated by the sensitivity and specificity reported by Yang and Tobin [13]. Without these conceptual models, the wide scatter in reported posttest probability by different investigators suggests that f/V_T is an unreliable weaning-predictor test. When the two models are applied, however, the values of posttest probability reported by most investigators are largely those one would predict for each reported value of pretest probability ($p < 0.0001$). The importance of pretest probability was not taken into account by the Task Force when reaching their conclusion that f/V_T is an unreliable predictor of weaning outcome. Yet, according to Bayes' theorem, no factor has a greater influence on posttest probability than pretest probability [12].

The importance of pretest probability is further emphasized by the results of the second weighted Pearson's correlation analysis. After adjusting for variation in pretest probability (among studies in the ACCP Task Force meta-analysis), this analysis revealed a significant relationship between the reported and predicted posttest probability of f/V_T: the relationship for positive-predictive values was $r = 0.86$ ($p < 0.0001$) and that for negative-predictive values $r = 0.82$ ($p < 0.0001$; Figs. 4, 5). The relationship was further confirmed by the Bland-Altman analysis (of studies in the meta-analysis): all of the reported positive-predictive values and all but two of the reported negative-predictive values fell within the 95% confidence interval of the predicted values. For the entire study group all of the reported values fell within the 95% confidence limits of the predicted values. The apparent discrepancy between the number of points lying outside the 95% confidence intervals on Figs. 2 and 5 and the Bland-Altman analysis is related to different entities being quantified. The outer curves on Figs. 2 and 3 represent the upper and lower 95% confidence interval for predicted posttest probability at a particular pretest probability. The Bland-Altman analysis (the usual method for quantifying the agreement between a prediction against a reference standard) measures the difference between predicted and reported posttest probability as related to the mean of these two values.

Pretest probability: wide variation, and overall above 0.75

More than one-half the studies were conducted in populations in which the rate of successful outcome was 75% or higher (Table 1). Such a high pretest probability has a major influence on posttest probability [63, 64].

Consider a clinician who obtains a positive reading on a hypothetical weaning-predictor test that has sensitivity 0.90 and specificity 0.90. If pretest probability of weaning success is 0.40, according to Bayes' theorem posttest probability is 0.86. If pretest probability is 0.80, posttest probability is 0.97. The increase between pretest and posttest probability in the second instance (21%, 0.17/.80) is only a fraction of that in the first instance (115%, 0.46/0.40) despite the sensitivity and specificity being identical. Thus a high pretest probability markedly decreases the apparent reliability of a weaning-predictor test.

Spectrum and test-referral bias

When two or more diagnostic tests that are not conditionally independent are used in sequence, spectrum and test-referral bias become almost inevitable [11, 12, 17, 64, 65, 66]. Spectrum bias occurs when a new study population contains fewer (or more) sick patients than the population in which a diagnostic test was originally developed [11, 12, 18]. For example, researchers may obtain measurements of a test that was originally developed to predict the outcome of a weaning trial. The researchers

then decide to assess the reliability of that same test in predicting the outcome of a trial of extubation. By design, the researchers must exclude patients who failed the weaning trial. By excluding sicker patients, those failing a weaning trial, the researchers change the spectrum of disease severity in the new population compared with that in the original study population and thus increase pretest probability of success. Evidence for the occurrence of spectrum bias is provided by the lower (average) value of f/V_T in 15 studies (where data were reported) than in the original study of Yang and Tobin [13], 77.4 vs. 89.1.

A second form of bias, test-referral bias, occurs when the results of a test under evaluation are used to select patients for the gold-standard test [11, 12]. Consider a weaning-predictor test where its reliability is evaluated in terms of its ability to predict the successful toleration of extubation. If patients are required to pass a weaning trial before extubation, this study-entry requirement necessarily excludes all patients who fail. This step has three effects on the study population; firstly, fewer patients with negative results (of the weaning-predictor test) are included; secondly, relatively more patients with positive results are included; thirdly, pretest probability of success is increased [11, 12]. The first consequence produces a decrease in the specificity of the weaning-predictor test in this population compared with the population in which the test was originally developed. The second consequence increases the sensitivity of the test. (See S.F1 in the Electronic Supplementary Material, which provides a hypothetical example of how test-referral bias leads to changes in pretest probability, sensitivity, and specificity.)

Specificity of f/V_T in the original report was 0.64. Of subsequent studies free of major problems (excluding [31, 39, 43]), 18 report specificity values for f/V_T that are less than 0.64 (Table 1). Sensitivity of f/V_T in the original report was 0.97. Because sensitivity has a ceiling of 1.00, a value of 0.97 does not leave much room to detect a further increase in sensitivity (allowing for usual biological noise created in any experiment). Of subsequent studies free of major problems 17 report sensitivity values for f/V_T that are greater than 0.90. These lower specificities and high sensitivities provide evidence for the occurrence of test-referral bias.

Screening testing and confirmatory testing

The ACCP Task Force recommendation to bypass a weaning-predictor test and go directly to a weaning trial [14, 15] contravenes a cardinal precept of diagnostic testing: use of a screening test followed by a confirmatory test [10, 11, 12]. Diagnostic testing is commonly seen as a monolithic entity—a test is a test is a test. In reality, diagnostic testing is expected to fulfill two very different demands [10, 11, 12]. One is screening: to pick up cases of a condition at the earliest possible time. This demand requires a test with high sensitivity [10, 11, 12]. The second is confirmation of a condition for which there is already a strong suspicion. This demand requires a test with high specificity [10, 11, 12]. With rare exceptions a single diagnostic test does not satisfy both demands [10, 11, 12]. Thus before evaluating a test's performance, it is imperative to ask to which demand is it directed.

A weaning-predictor test is used to spot the earliest point in time that a patient might tolerate a weaning trial [13]. It serves solely as a screening test. On its own a positive predictor-test result is not used as justification for extubation [67]. Before that step a weaning trial (a confirmatory test) is undertaken. The ideal time to undertake a screening test is when the pretest probability of weaning success is 20% or less [10]. In contrast, weaning trials are commonly performed when the pretest probability of success is 75% or more. None of the 29 studies in Table 1 had a pretest probability under 45%. This finding is not surprising. Physicians know that a weaning trial takes as long as 30 min–2 h to perform, and staff must be available to closely monitor the patient. Thus physicians do not initiate a weaning trial unless they think the patient has a reasonably high likelihood of success.

The development of a reliable screening test hinges on avoiding false-negative results (a test predicting failure, but the patient actually succeeds) [10, 11]. Simultaneously the test needs to pick up every possible true-positive result—the mindset is to miss no patient who can breathe without the ventilator. To capture the maximum meaningful number of true-positive results, the threshold for defining a positive screening test may be set deliberately high [10, 11]. This necessarily increases the number of false-positive results, producing a proportional decrease in specificity.

Sensitivity captures exactly the components that define the reliability of a screening test since it contains only true-positive and false-negative rate. Likewise, specificity captures exactly the constituents of a reliable confirmatory test: avoidance of false-positive results (a test predicting success, but the patient actually fails) and maximizing true-negative rate [10, 11]. The studies listed in Table 1 reveal sensitivity values for f/V_T that are at least 0.90 [22, 24, 25, 27, 30, 37, 44, 45, 47] or at least 0.97 [13, 20, 21, 36, 38, 40, 41]. Thus f/V_T constitutes a reliable screening test. In contrast, the sensitivity of a weaning trial as a diagnostic test has never been tested.

Limitations

The studies shown in Figs. 2 and 3 include every study that provided the necessary information on f/V_T. We recognize that a case could be made to exclude data from cer-

tain studies, for example, those conducted in infants [33, 39, 42], those that included measurements of f/V_T during pressure-support ventilation [27, 28, 38, 45], or those in which pretest probability exceeded 88% [27, 36, 42, 47]. The reasons to exclude a study are necessarily arbitrary in nature. Because no study, other than the two studies with inconsistent data [39, 43], was excluded, the relationships between reported posttest probability of f/V_T with both pretest probability and predicted posttest probability may be underestimates.

Our data analysis is framed in terms of pretest probability, although that value was not reported directly by authors of the primary studies. We took prevalence of successful outcome as a surrogate for pretest probability because these two terms are used interchangeably in the literature on diagnostic testing [56, 57, 58, 59, 60, 61]. The primary aim of the present study was to determine whether spectrum bias and test-referral bias explain some of the reported variation in f/V_T reliability as a screening test (for weaning success). Evidence for spectrum and test-referral bias is provided by the lower values of f/V_T and specificity, respectively, in subsequent reports than in the original study. The conclusion would remain the same were we to eliminate all mention of pretest probability, and express our findings in terms of prevalence.

Conclusion

Based on a meta-analysis of likelihood ratios, an ACCP Task Force concluded that f/V_T is not a reliable predictor of weaning success. The included studies, however, exhibited significant heterogeneity ($p < 0.00001$), a factor that nullifies a meta-analysis. The heterogeneity in pretest probability (prevalence of successful outcome) most likely resulted from spectrum and test-referral bias. When data from 29 studies were entered into a Bayesian model with pretest probability as the operating point, the observed posttest probabilities were closely correlated with the values predicted by the original study on f/V_T ($p < 0.0001$).

A separate problem was the Task Force's failure to focus on the goal of a weaning-predictor test: to screen for weanability. Likelihood ratio is not precisely suited to assessing screening-test reliability (because it includes constituents not directly relevant), whereas sensitivity solely captures the vital components. The average reported sensitivity of f/V_T was 0.87. Thus contrary to the conclusion reached by the ACCP Task Force, the facts included in the aggregated studies show that f/V_T is a reliable predictor of weaning success.

Acknowledgements. We thank Dr. S. Banks and Dr. B. Grant for advice and assistance on statistical analyses.

References

1. Tobin MJ (2001) Advances in mechanical ventilation. N Engl J Med 344:1986–1996
2. Tobin MJ, Jubran A (2006) Weaning from mechanical ventilation. In: Tobin MJ (ed) Principles and practice of mechanical ventilation. McGraw-Hill, New York, pp 1185–1220
3. Fagon JY, Chastre J, Domart Y, Trouillet JL, Pierre J, Darne C, Gibert C (1989) Nosocomial pneumonia in patients receiving continuous mechanical ventilation. Prospective analysis of 52 episodes with use of a protected specimen brush and quantitative culture techniques. Am Rev Respir Dis 139:877–884
4. Zakynthinos S, Routsi C, Vassilakopoulos T, Kaltsas P, Zakynthinos E, Kazi D, Roussos C (2005) Differential cardiovascular responses during weaning failure: effects on tissue oxygenation and lactate. Intensive Care Med 31:1634–1642
5. Richard C, Teboul JL (2005) Weaning failure from cardiovascular origin. Intensive Care Med 31:1605–1607
6. Bien MY, Hseu SS, Yien HW, Kuo BI, Lin YT, Wang JH, Kou YR (2004) Breathing pattern variability: a weaning predictor in postoperative patients recovering from systemic inflammatory response syndrome. Intensive Care Med 30:241–247
7. 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 ventilatory support during weaning from mechanical ventilation. Am J Respir Crit Care Med 150:896–903
8. Esteban A, Frutos F, Tobin MJ, Alia I, Solsona JF, Valverdu I, Fernandez R, de La Cal MA, Benito S, Tomas R, (1995) A comparison of four methods of weaning patients from mechanical ventilation. Spanish Lung Failure Collaborative Group. N Engl J Med 332:345–350
9. Tversky A, Kahneman D (1974) Judgment under uncertainty: Heuristics and biases. Science 185:1124–1131
10. Griner PF, Mayewski RJ, Mushlin AI, Greenland P (1981) Selection and interpretation of diagnostic tests and procedures. Principles and applications. Ann Intern Med 94:557–592
11. Feinstein AR (1985) Clinical epidemiology: the architecture of clinical research. Saunders, Philadelphia
12. Sox HC Jr, Clatt MA, Higgins MC, Marton KI (1988) Medical decision making. Butterworths, Boston
13. Yang KL, Tobin MJ (1991) A prospective study of indexes predicting the outcome of trials of weaning from mechanical ventilation. N Engl J Med 324:1445–1450
14. 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
15. MacIntyre NR, Cook DJ, Ely EW Jr, Epstein SK, Fink JB, Heffner JE, Hess D, Hubmayer RD, Scheinhorn DJ (2001) Evidence-based guidelines for weaning and discontinuing ventilatory support: 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. Chest 120:375S-395S

16. Tobin MJ, Perez W, Guenther SM, Semmes BJ, Mador MJ, Allen SJ, Lodato RF, Dantzker DR (1986) The pattern of breathing during successful and unsuccessful trials of weaning from mechanical ventilation. Am Rev Respir Dis 134:1111–1118
17. Rozanski A, Diamond GA, Berman D, Forrester JS, Morris D, Swan HJ (1983) The declining specificity of exercise radionuclide ventriculography. N Engl J Med 309:518–522
18. Ransohoff DF, Feinstein AR (1978) Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 299:926–930
19. Del Rosario N, Sassoon CS, Chetty KG, Gruer SE, Mahutte CK (1997) Breathing pattern during acute respiratory failure and recovery. Eur Respir J 10:2560–2565
20. Sassoon CS, Mahutte CK (1993) Airway occlusion pressure and breathing pattern as predictors of weaning outcome. Am Rev Respir Dis 148:860–866
21. Chatila W, Jacob B, Guaglionone D, Manthous CA (1996) The unassisted respiratory rate-tidal volume ratio accurately predicts weaning outcome. Am J Med 101:61–67
22. Dojat M, Harf A, Touchard D, Laforest M, Lemaire F, Brochard L (1996) Evaluation of a knowledge-based system providing ventilatory management and decision for extubation. Am J Respir Crit Care Med 153:997–1004
23. Afessa B, Hogans L, Murphy R (1999) Predicting 3-day and 7-day outcomes of weaning from mechanical ventilation. Chest 116:456–461
24. Yang KL (1993) Inspiratory pressure/maximal inspiratory pressure ratio: a predictive index of weaning outcome. Intensive Care Med 19:204–208
25. 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
26. Epstein SK, Ciubotaru RL (1996) Influence of gender and endotracheal tube size on preextubation breathing pattern. Am J Respir Crit Care Med 154:1647–1652
27. Leitch EA, Moran JL, Grealy B (1996) Weaning and extubation in the intensive care unit. Clinical or index-driven approach? Intensive Care Med 22:752–759
28. Lee KH, Hui KP, Chan TB, Tan WC, Lim TK (1994) Rapid shallow breathing (frequency-tidal volume ratio) did not predict extubation outcome. Chest 105:540–543
29. Mergoni M, Costa A, Primavera S, Salvadori A, Saccani A, Zuccoli P (1996) Valutazione di alcuni nuovi parametri predittivi dell'esito dello svezzamento dalla ventilazione meccanica. Minerva Anestesiol 62:153–164
30. Krieger BP, Isber J, Breitenbucher A, Throop, Ershowsky P (1997) Serial measurements of the rapid-shallow-breathing index as a predictor of weaning outcome in elderly medical patients. Chest 112:1029–1034
31. Gologorskii VA, Gelfand BR, Stamov VI, Lapshina IIu, Nistratov SL (1997) Cessation of prolonged artificial ventilation of the lungs and transition to spontaneous respiration of surgical patients (in Russian). Anesteziol Reanimatol 1:4–10
32. Khan N, Brown A, Venkataraman ST (1996) Predictors of extubation success and failure in mechanically ventilated infants and children. Crit Care Med 24:1568–1579
33. Baumeister BL, el Khatib M, Smith PG, Blumer JL (1997) Evaluation of predictors of weaning from mechanical ventilation in pediatric patients. Pediatr Pulmonol 24:344–352
34. Capdevila XJ, Perrigault PF, Perey PJ, Roustan JP, d'Athis F (1995) Occlusion pressure and its ratio to maximum inspiratory pressure are useful predictors for successful extubation following T-piece weaning trial. Chest 108:482–489
35. Gandia F, Blanco J (1992) Evaluation of indexes predicting the outcome of ventilator weaning and value of adding supplemental inspiratory load. Intensive Care Med 18:327–333
36. Jacob B, Chatila W, Manthous CA (1997) The unassisted respiratory rate/tidal volume ratio accurately predicts weaning outcome in postoperative patients. Crit Care Med 25:253–257
37. Vallverdu 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 158:1855–1862
38. Mohsenifar Z, Hay A, Hay J, Lewis MI, Koerner SK (1993) Gastric intramural pH as a predictor of success or failure in weaning patients from mechanical ventilation. Ann Intern Med 119:794–798
39. Farias JA, Alia I, Esteban A, Golubicki AN, Olazarri FA (1998) Weaning from mechanical ventilation in pediatric intensive care patients. Intensive Care Med 24:1070–1075
40. Bouachour G, Guiraud MP, Gouello JP, Roy PM, Alquier P (1996) Gastric intramucosal pH: an indicator of weaning outcome from mechanical ventilation in COPD patients. Eur Respir J 9:1868–1873
41. Rivera L, Weissman C (1997) Dynamic ventilatory characteristics during weaning in postoperative critically ill patients. Anesth Analg 84:1250–1255
42. Thiagarajan RR, Bratton SL, Martin LD, Brogan TV, Taylor D (1999) Predictors of successful extubation in children. Am J Respir Crit Care Med 160:1562–1566
43. 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
44. Maldonado A, Bauer TT, Ferrer M, Hernandez C, Arancibia F, Rodriguez-Roisin R, Torres A (2000) Capnometric recirculation gas tonometry and weaning from mechanical ventilation. Am J Respir Crit Care Med 161:171–176
45. Uusaro A, Chittock DR, Russell JA, Walley KR (2000) Stress test and gastric-arterial PCO2 measurement improve prediction of successful extubation. Crit Care Med 28:2313–2319
46. Khamiees M, Raju P, DeGirolamo A, Amoateng-Adjepong Y, Manthous CA (2001) Predictors of extubation outcome in patients who have successfully completed a spontaneous breathing trial. Chest 120:1262–1270
47. Smina M, Salam A, Khamiees M, Gada P, Amoateng-Adjepong Y, Manthous CA (2003) Cough peak flows and extubation outcomes. Chest 124:262–268
48. Conti G, Montini L, Pennisi MA, Cavaliere F, Arcangeli A, Bocci MG, Proietti R, Antonelli M (2004) A prospective, blinded evaluation of indexes proposed to predict weaning from mechanical ventilation. Intensive Care Med 30:830–836
49. Fernandez R, Raurich JM, Mut T, Blanco J, Santos A, Villagra A (2004) Extubation failure: diagnostic value of occlusion pressure (P0.1) and P0.1-derived parameters. Intensive Care Med 30:234–240
50. Jiang JR, Tsai TH, Jerng JS, Yu CJ, Wu HD, Yang PC (2004) Ultrasonographic evaluation of liver/spleen movements and extubation outcome. Chest 126:179–185
51. Brand R, Kragt H (1992) Importance of trends in the interpretation of an overall odds ratio in the meta-analysis of clinical trials. Stat Med 11:2077–2082

52. Schmid CH, Lau J, McIntosh MW, Cappelleri JC (1998) An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. Stat Med 17:1923–1942
53. Breslow NE, Day NE (1980) Statistical methods in cancer research: the analysis of case-control studies. IARC, Lyon
54. Barnard GA (1958) Studies in the history of probability and stastistics. IX. Thomas Bayes essay towards solving a problem in the doctrine of chances. Biometrika 45:294–315
55. Howson C, Urbach P (2006) Scientific reasoning: the Bayesian approach, 3rd edn. Open Court, Chicago
56. Elstein AS (1999) Heuristics and biases: selected errors in clinical reasoning. Acad Med 74:791–794
57. Eisenberg MJ, Schiller NB (1991) Bayes' theorem and the echocardiographic diagnosis of cardiac tamponade. Am J Cardiol 68:1242–1244
58. Weiner DA, Ryan TJ, McCabe CH, Kennedy JW, Schloss M, Tristani F, Chaitman BR, Fisher LD (1979) Exercise stress testing. Correlations among history of angina, ST-segment response and prevalence of coronary-artery disease in the Coronary Artery Surgery Study (CASS). N Engl J Med 301:230–235
59. Kassirer JP (1989) Diagnostic reasoning. Ann Intern Med 110:893–900
60. Feinstein AR (2002) Misguided efforts and future challenges for research on "diagnostic tests". J Epidemiol Community Health 56:330–332
61. Eisenberg MJ (1995) Accuracy and predictive values in clinical decision-making. Cleve Clin J Med 62:311–316
62. Anonymous (1992) CenterSoft statistical programs for medical decision making research 2x2, version 1.0. University of Birmingham, Birmingham
63. Sox HC Jr (1986) Probability theory in the use of diagnostic tests. An introduction to critical study of the literature. Ann Intern Med 104:60–66
64. Sox HC (1996) The evaluation of diagnostic tests: principles, problems, and new developments. Annu Rev Med 47:463–471
65. Diamond GA (1986) Reverend Bayes' silent majority. An alternative factor affecting sensitivity and specificity of exercise electrocardiography. Am J Cardiol 57:1175–1180
66. Lachs MS, Nachamkin I, Edelstein PH, Goldman J, Feinstein AR, Schwartz JS (1992) Spectrum bias in the evaluation of diagnostic tests: lessons from the rapid dipstick test for urinary tract infection. Ann Intern Med 117:135–140
67. Tobin MJ, Laghi F (2006) Extubation. In: Tobin MJ (ed) Principles and practices of mechanical ventilation. McGraw-Hill, New York, pp 1221–1237