Inter-Observer Agreement of Spontaneous Breathing Trial Outcome

Juan B Figueroa-Casas MD, Afshin Broukhim MD, Adrian Vargas RRT, Linda Milam RRT, and Ricardo Montoya RRT

BACKGROUND: Spontaneous breathing trials (SBTs) are a very important test in the weaning process. The trial involves evaluation of multiple objective and subjective variables. These characteristics could lead to variability in interpreting outcomes with important clinical implications. We aimed to measure the inter-observer agreement between respiratory therapists when analyzing SBT outcomes. METHODS: In the context of a respiratory therapist-driven weaning protocol, 2 respiratory therapists independently interpreted the subjective variables (use of accessory muscles, agitation, and diaphoresis) and the overall outcome of SBTs (success vs failure) performed in adult subjects mechanically ventilated for any duration. Raw agreements between respiratory therapists and kappa statistics were calculated. RESULTS: One-hundred fifty-one SBTs were interpreted. The overall trial outcome raw agreement was 93.3% (95% CI 88.2–96.3) and kappa 0.63 (95% CI 0.47–0.79). Raw agreement for subjective variables ranged between 92.1% (agitation) and 99.3% (diaphoresis). The group with disagreements in overall trial outcome had higher breathing frequency, breathing-frequency-to-tidal-volume ratio, and systolic blood pressure prior to the trial. CONCLUSIONS: Within a respiratory therapist-driven weaning protocol, we found a near 90% inter-observer agreement in the interpretation of SBT outcomes. Our findings illustrate the complexity of interpreting fluctuating subjective and objective variables and their integration into one result: SBT success versus failure. Refining the definitions of variables and their limits for failure along with education might reduce this variability. *Key words: ventilator weaning; mechanical ventilators; inter-observer variability; airway extubation.* [Respir Care 2014;59(9):1324–1328. © 2014 Daedalus Enterprises]

Introduction

A spontaneous breathing trial (SBT) is a recommended[1] and widely performed[2] test in the process of weaning from mechanical ventilation. The outcome of this test is used clinically to decide on liberation from mechanical ventilation and in research as a reference standard to evaluate potential predictors of weaning outcome.

The SBT is a test with a dichotomous outcome consisting of or failure. In clinical and research applications, this test is composed of a combination of outcomes from individual variables that are evaluated during the trial. These variables commonly include objective physiologic variables, such as breathing frequency and arterial oxygen saturation, and other variables that are subjective for detection and quantification, such as patient comfort.[3] These characteristics of the SBT make its outcome prone to variable interpretation by different observers. This variability has not been investigated extensively, and it may have important influence on clinical decisions regarding weaning and on

SEE THE RELATED EDITORIAL ON PAGE 1458

Dr Figueroa-Casas is affiliated with the Division of Pulmonary and Critical Care Medicine, and Dr Broukhim is affiliated with the Department of Anesthesia, Paul L Foster School of Medicine, Texas Tech University Health Sciences Center, El Paso, Texas. Mr Vargas, Ms Milam, and Mr Montoya are affiliated with the Department of Respiratory Care, University Medical Center of El Paso, El Paso, Texas.

The authors have disclosed no conflicts of interest.

Correspondence: Juan B Figueroa-Casas MD, Division of Pulmonary and Critical Care Medicine, Paul L Foster School of Medicine, Texas Tech University Health Sciences Center, 4800 Alberta Avenue, Suite 136, El Paso, TX 79905. E-mail: juan.figueroa@ttuhsc.edu.

DOI: 10.4187/respcare.03022

the reliability of the reference standard against which the accuracy of other tests are measured. The primary aim of this study was to measure the inter-observer agreement of the SBT outcome adjudication and components among respiratory therapists.

QUICK LOOK

Current knowledge

Spontaneous breathing trials (SBTs) have been found to be the best method for determining the timing for discontinuation of mechanical ventilation. The determination of a successful SBT requires evaluation of multiple objective and subjective variables.

What this paper contributes to our knowledge

Within a respiratory therapist-driven weaning protocol, the inter-observer agreement in the interpretation of SBT outcomes was ~90%. Interpreting fluctuating subjective and objective variables and their integration into a binary result (success versus failure) are complex tasks open to individual interpretation. Improved definitions for defining failure can reduce the variability in interpretation.

Methods

The study consisted of independent interpretations of SBT outcomes by 2 respiratory therapists and subject data collection. This was a prospective observational study performed in a 30-bed ICU of a university hospital (University Medical Center of El Paso, El Paso, Texas) that admits adult medical, surgical, and trauma patients. The ICU has a single ventilator weaning protocol that is applied daily to all ventilated patients. Consent from patients was waived by the local institutional review board due to the nature of the study and the collection of existing clinical data. Twenty-five respiratory therapists verbally consented to participating in the study with their interpretations and were aware of the purpose of the study.

Subjects

Subjects eligible for inclusion were 18 y of age or older, were intubated and on mechanical ventilation for any duration, and had an SBT ordered by the treating ICU team. Subjects with tracheostomy were excluded. Data collection included demographics, primary diagnosis requiring mechanical ventilation, duration of ventilation, Acute Physiology and Chronic Health Evaluation (APACHE) II score,[4] P_{aO_2}/F_{IO_2} the morning of the SBT, the ratio of breathing frequency to tidal volume (f/V_T) measured prior to the SBT, and the physiologic variables measured before the start and at end of the SBT.

Weaning Protocol

The protocol described here is modified from those of published studies[5,6] and was in place during the entire period of the study. All patients undergoing mechanical ventilation were screened daily by a respiratory therapist for the following criteria: $F_{IO_2} \leq 0.5$, PEEP ≤ 5 cm H_2O, minute ventilation < 15 L/min, $P_{aO_2}/F_{IO_2} \geq 150$, $f/V_T \leq 105$ breaths/min/L (while on CPAP of 5 cm H_2O without pressure support for 1 min, a modification from the original method[7]), presence of cough and gag reflex, absence of sedative or vasopressor infusions (except dopamine at ≤ 5 μg/kg/min), and Glasgow coma scale ≥ 10 (for trauma patients only). Upon meeting all screening criteria, ICU treating physicians were notified and decided upon ordering a SBT, but this could also be ordered when physicians considered it appropriate even if not all screening criteria were met. All SBTs were performed at a CPAP of 5 cm H_2O without pressure support and were supervised by respiratory therapists for a maximum duration of 30 min. Trial failure was defined as meeting any one of the following criteria for the variables monitored before the start of and during the trial: breathing frequency > 35 breaths/min (obtained from the ventilator display), pulse oximetry $< 90\%$, change in heart rate $> 20\%$, increase in systolic blood pressure $> 25\%$, and presence of agitation, diaphoresis, or use of accessory muscles of respiration. Patients tolerating the trial successfully were communicated to the ICU treating physicians to decide on extubation.

SBT Outcome Interpretation

The respiratory therapist caring for the included subject (primary) conducted and evaluated the SBT following the usual protocol. This therapist was responsible for termination of the trial and recording the subjective and objective variable results before and at the end of the trial in the ventilator flow sheet. A second respiratory therapist (secondary) working in the ICU but not directly involved in the care of the subject was asked to simultaneously observe the SBT. The secondary therapist had access to all medical records and the variables monitored before and during the trial except for the records of the trial itself made by the primary therapist. At the end of the SBT, both therapists were asked to independently indicate whether the subject had shown each of the following signs during the SBT: excessive use of accessory muscles, agitation, and diaphoresis. In addition, based on their overall assessment of those signs and the objective variables evaluated before and during the SBT, they were asked whether the

Table 1. Characteristics of Subjects on Day of SBT

Characteristic	Value
Age (y)	54 (18)
Male (%)	60
Cause requiring mechanical ventilation (%)	
Head and/or multiple trauma	28
Postoperative	25
Neurologic emergency	13
Nonpulmonary severe sepsis	11
Primary lung disease	7
Other	16
Duration of ventilation on day of SBT (d)	2 (1–3)
APACHE II score	14 (5)
P_{aO_2}/F_{IO_2}	257 (217–323)
f/V_T (breaths/min/L)	56 (39–83)
$f/V_T > 105$ (%)	13
SBT pass by primary respiratory therapist (%)	90

Data are expressed as mean (SD) or median (interquartile range) unless specified.
APACHE = Acute Physiology and Chronic Health Evaluation
f/V_T = breathing-frequency-to-tidal-volume ratio
SBT = spontaneous breathing trial

subject had passed the trial. These yes/no responses were recorded by each therapist confidentially in separate specific forms.

Statistical Analysis

Subjects' characteristics were summarized with proportions, mean ± SD, or median and interquartile range according to the type of variable and distribution of the data. Inter-observer agreement was analyzed by calculation of raw agreement and kappa statistics with their respective 95% CI values. Differences between agreement and disagreement groups were analyzed by *z* test for proportions and by the Mann-Whiney test or independent-sample *t* test for continuous variables. All analyses of differences were 2-tailed, and $P < .05$ was considered to indicate statistical significance.

Results

One-hundred fifty-one SBTs performed in different subjects were included. Subjects' characteristics on the day of the evaluated SBT are shown in Table 1. Immediately after the SBT, 135 of 136 subjects that passed the SBT (according to the primary therapist) were extubated, whereas none of the 15 subjects that failed the SBT were extubated. Eventually, 149 subjects were extubated. Six percent of these were re-intubated, and an additional 2% received only noninvasive ventilation within 48 h. One subject died, and one had care withdrawn without ever being extubated. One-hundred thirty of 131 subjects that passed the SBT with agreement between both therapists were extubated immediately after passing the trial. Of these 130 subjects, 7% required re-intubation and/or noninvasive ventilation within 48 h.

Table 2. Agreement Between Respiratory Therapists for Individual Signs and for Overall SBT Outcome

Variable	Raw Agreement (95% CI)	Kappa Coefficient (95% CI)
Use of accessory muscles	97.4 (93.4–98.9)	0.74 (0.58–0.89)
Agitation	92.1 (86.6–95.4)	0.29 (0.14–0.44)
Diaphoresis	99.3 (96.3–99.9)	0.80 (0.64–0.95)
Overall SBT outcome	93.3 (88.2–96.3)	0.63 (0.47–0.79)

SBT = spontaneous breathing trial

Both therapists agreed on an overall SBT success in 131 trials and on an overall SBT failure in 10 trials. Of the 10 failure-agreement trials, 2 trials met only objective failure criteria, and 4 trials met only subjective failure criteria as determined by both therapists, and 4 trials met both objective and subjective criteria as determined by both therapists. In 6 of the 131 success-agreement trials, one of the respiratory therapists indicated that agitation was present but that the overall trial was passed.

There was disagreement on overall SBT success versus failure in 10 trials. Among these disagreement trials, the primary and secondary therapists interpreted SBT failure in 5 trials each. In 5 of the 10 disagreement trials, there were also disagreements regarding the presence of one or more of the subjective signs. The measures of agreement between respiratory therapists for the presence of each individual subjective sign and for the overall SBT outcome are shown in Table 2. A comparison of characteristics between SBTs with overall outcome agreement versus disagreement is shown in Table 3. In this comparison, the disagreement group had higher breathing frequency, f/V_T, and systolic blood pressure before the start of the SBT.

Discussion

In the context of a clinically applied respiratory therapist-driven weaning protocol, this study found a near 10% rate of disagreement between respiratory therapists in the interpretation of the outcome of SBTs. Although the analysis by kappa statistics (95% CI 0.47–0.79) suggests a moderate-to-substantial inter-observer agreement for the test,[8] this degree of variability may have important implications.

At least some of the criteria for SBT failure used in our protocol are widely used in clinical practice and have been applied in important clinical studies of weaning[5,9,10] as well as in defining the reference standard for identification

Table 3. Differences Between Groups With Agreement Versus Disagreement on Overall SBT Outcome

Variable	Agreement ($n = 141$)	Disagreement ($n = 10$)	P
Age (y)	57 (18)	54 (18)	.59
Male (%)	62	60	.90
Duration of ventilation (d)	2 (1–3)	3.5 (1–8)	.18
APACHE II score	13 (9–17)	15 (14–18)	.07
P_{aO_2}/F_{IO_2}	258 (212–323)	256 (247–271)	.90
f/V_T (breaths/min/L)	54 (37–75)	118 (88–127)	.001
$f/V_T > 105$ (%)	9	60	.001
Breathing frequency (breaths/min)	18 (14–23)	30 (25–32)	< .001
Heart rate (beats/min)	91 (19)	101 (13)	.09
Systolic blood pressure (mm Hg)	131 (22)	148 (20)	.01
S_{pO_2} (%)	99 (97–100)	97 (96–99)	.14

All physiologic parameters are pre-SBT start. Data expressed as mean (SD) or median (interquartile range) unless specified.
APACHE = Acute Physiology and Chronic Health Evaluation
f/V_T = breathing-frequency-to-tidal-volume ratio
SBT = spontaneous breathing trial

of determinants or predictors of weaning outcome.[7,11-13] The SBT involves clinical observation and therefore will never reach perfect inter-observer agreement. Given the importance of this test, however, it is pertinent to recognize and quantify this limitation of the test and to identify potentially correctable factors to minimize its variability.

To characterize the sources of disagreement, we tested the overall success/failure outcome of the trial but also the subjective variables assessed in our protocol, suspected sources of disagreement. We found that approximately half of the overall outcome disagreements were related to the detection of the subjective variables, whereas the others were related to the interpretation of the objective variables. In addition to disagreements about the presence of subjective variables, we found few inconsistencies in therapists detecting these signs of failure, specifically agitation, but interpreting the overall trial as passed. When questioned about these interpretations, the therapists reported that although the sign had developed during the trial, its severity was insufficient to independently warrant trial failure. Regarding the apparent disagreements in interpretation of objective variables, a review of these trials showed acceptable results of objective variables at the end of the trial, except for tachypnea, near our limit criteria for failure. We can only speculate that fluctuations of the breathing frequency around the failure limit were interpreted differently by different observers. The interpretation disagreements noted allow us to hypothesize that some measures could minimize inter-observer variability for some of the components and the overall adjudication of SBT outcome. For subjective variables, such as agitation and use of accessory muscles, reproducible graded scores with absolute or relative change thresholds to define failure may be considered. For monitored objective variables, such as breathing frequency, the use of relative changes rather than absolute limits may be helpful, particularly when the level is close to an absolute limit.

Limitations related to study subject selection need to be considered before generalizing our findings. First, the group duration of ventilation prior to SBT in our sample was relatively short, which might be associated with a lesser challenge in the weaning process. Second, the selection criteria for entry into the SBT may influence the magnitude of the inter-observer agreement studied. In support of this concept, our comparison between SBTs that led to overall outcome agreement versus disagreement found that the latter was associated with higher systolic blood pressure, breathing frequency, and f/V_T before the start of the trial. Patients who enter the SBT with variable levels closer to the failure limits probably have a limited cardiorespiratory reserve and are logically more likely to be interpreted as failing by at least one observer. We consider our selection criteria for SBT entry to be relatively restrictive since 87% of our subjects had a pre-SBT f/V_T of < 105,[7] and only 3% had a P_{aO_2}/F_{IO_2} of < 150. From these considerations regarding patient selection, we suspect that our measurement of inter-observer agreement for SBT outcome represents a high-end estimate. Third, being a single-center study and having only 2 observers per trial are additional limitations to the precision of our agreement estimations. However, regardless of the magnitude of the disagreement, its identified sources are likely applicable to most weaning scenarios.

To our knowledge, only one prior study[14] had measured the inter-observer agreement of the SBT overall outcome and reported a level slightly lower than ours: raw agreement of 86% (95% CI 79–91) and kappa 0.57 for pairs of respiratory therapists. The reported agreement for pairs of physicians was not significantly different. In that study, f/V_T was not included and the SBTs were performed with various modalities, including T-tube and pressure support.

Conclusions

Within a respiratory therapist-driven weaning protocol, we found a near 90% inter-observer agreement in the interpretation of SBT outcomes. These findings illustrate the complexity of interpreting fluctuating subjective and objective variables and their integration into one result: SBT success versus failure. In the setting of respiratory therapist-conducted SBTs, further clarity regarding the definition and training in the detection of each failure criterion might reduce the variability in interpreting SBT outcome.

REFERENCES

1. MacIntyre N. Discontinuing mechanical ventilatory support. Chest 2007;132(3):1049-1056.
2. Esteban A, Frutos-Vivar F, Muriel A, Ferguson ND, Peñuelas O, Abraira V, et al. Evolution of mortality over time in patients receiving mechanical ventilation. Am J Respir Crit Care Med 2013;188(2): 220-230.
3. MacIntyre NR. Evidence-based ventilator weaning and discontinuation. Respir Care 2004;49(7):830-836.
4. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med 1985; 13(10):818-829.
5. Ely EW, Baker AM, Dunagan DP, Burke HL, Smith AC, Kelly PT, et al. Effect on the duration of mechanical ventilation of identifying patients capable of breathing spontaneously. N Engl J Med 1996; 335(25):1864-1869.
6. Namen AM, Ely EW, Tatter SB, Case LD, Lucia MA, Smith A, et al. Predictors of successful extubation in neurosurgical patients. Am J Respir Crit Care Med 2001;163(3):658-664.
7. Yang KL, Tobin MJ. A prospective study of indexes predicting the outcome of trials of weaning from mechanical ventilation. N Engl J Med 1991;324(21):1445-1450.
8. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159-174.
9. Esteban A, Alía I, Gordo F, Fernández R, Solsona JF, Vallverdú I, et al. Extubation outcome after spontaneous breathing trials with T-tube or pressure support ventilation. Am J Respir Crit Care Med 1997;156(2):459-465.
10. Esteban A, Alía I, Tobin MJ, Gil A, Gordo F, Vallverdú I, et al. Effect of spontaneous breathing trial duration on outcome of attempts to discontinue mechanical ventilation. Am J Respir Crit Care Med 1999;159(2):512-518.
11. Jubran A, Grant BJ, Laghi F, Parthasarathy S, Tobin MJ. Weaning prediction: esophageal pressure monitoring complements readiness testing. Am J Respir Crit Care Med 2005;171(11):1252-1259.
12. Vassilakopoulos T, Zakynthinos S, Roussos C. The tension-time index and the frequency/tidal volume ratio are the major pathophysiologic determinants of weaning failure and success. Am J Respir Crit Care Med 1998;158(2):378-385.
13. Wysocki M, Cracco C, Teixeira A, Mercat A, Diehl JL, Lefort Y, et al. Reduced breathing variability as a predictor of unsuccessful patient separation from mechanical ventilation. Crit Care Med 2006; 34(8):2076-2083.
14. Cappati KR, Tonella RM, Damascena AS, Pereira CA, Caruso P. Interobserver agreement rate of the spontaneous breathing trial. J Crit Care 2013;28(1):62-68.