

Reappraisal of Ventilator-Free Days in Critical Care Research

✉ Nadir Yehya^{1,2}, Michael O. Harhay^{3,4}, Martha A. Q. Curley^{2,5,6}, David A. Schoenfeld^{7,8,9}, and Ron W. Reeder¹⁰

¹Department of Anesthesiology and Critical Care Medicine, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; ²Department of Anesthesiology and Critical Care Medicine, ³Department of Biostatistics, Epidemiology and Informatics, and ⁴Palliative and Advanced Illness Research Center, Perelman School of Medicine, and ⁵Department of Family and Community Health, School of Nursing, University of Pennsylvania, Philadelphia, Pennsylvania; ⁶Children's Hospital of Philadelphia Research Institute, Philadelphia, Pennsylvania; ⁷Biostatistics Center, Massachusetts General Hospital, Boston, Massachusetts; ⁸Department of Medicine, Harvard Medical School, Boston, Massachusetts; ⁹Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, Massachusetts; and ¹⁰Department of Pediatrics, University of Utah, Salt Lake City, Utah

ORCID IDs: 0000-0002-2996-217X (N.Y.); 0000-0002-0553-674X (M.O.H.); 0000-0001-5228-6694 (M.A.Q.C.).

Abstract

Ventilator-free days (VFDs) are a commonly reported composite outcome measure in acute respiratory distress syndrome trials. VFDs combine survival and duration of ventilation in a manner that summarizes the “net effect” of an intervention on these two outcomes. However, this combining of outcome measures makes VFDs difficult to understand and analyze, which contributes to imprecise interpretations. We discuss the strengths and limitations of VFDs and other “failure-free day” composites, and we provide a framework for when and how to use these outcome measures. We also provide a comprehensive discussion of the different analytic methods for analyzing and interpreting VFDs, including Student's *t* tests and rank-sum tests, as well as competing risk regressions treating extubation as the primary

outcome and death as the competing risk. Using simulations, we illustrate how the statistical test with optimal power depends on the relative contributions of mortality and ventilator duration on the composite effect size. Finally, we recommend a simple analysis and reporting framework using the competing risk approach, which provides clear information on the effect size of an intervention, a statistical test and measure of confidence with the ability to adjust for baseline factors and allow interim monitoring for trials. We emphasize that any approach to analyzing a composite outcome, including other “failure-free day” constructs, should also be accompanied by an examination of the components.

Keywords: acute respiratory distress syndrome; ARDS; ventilator-free days; VFDs; competing risk regression

Ventilator-free days (VFDs) at 28 days are one of several organ failure-free outcome measures used in critical care (Table 1) to quantify the efficacy of therapies and interventions (1, 2). The failure-free day concept was developed to summarize the effect of an intervention on morbidity in the presence of the competing event of death (3). Frequently used in adult and

pediatric acute respiratory distress syndrome (ARDS) trials, VFDs are one of the most common iterations of failure-free days (4–6). VFDs are typically defined as follows (1):

- VFDs = 0 if subject dies within 28 days of mechanical ventilation.
- VFDs = 28 - *x* if successfully liberated from ventilation *x* days after initiation.

- VFDs = 0 if the subject is mechanically ventilated for >28 days.

The 28-day time frame was initially chosen because most subjects with ARDS will have died or been extubated by Day 28 (1), although VFDs at 60 days have been reported in at least one ARDS trial (7). Despite their widespread use, critical assessment of the measurement and

(Received in original form October 30, 2018; accepted in final form April 29, 2019)

✉ This article is open access and distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). For commercial usage and reprints, please contact Diane Gern (dgern@thoracic.org).

Supported by NIH grants K23-HL136688 (N.Y.), K99-HL141678 (M.O.H.), and U01-HL123009 (D.A.S.).

Author Contributions: N.Y. and R.W.R. conceived of the study. N.Y., D.A.S., and R.W.R. performed data collation and analysis. M.O.H. and M.A.Q.C. provided additional critical intellectual content for the work. All authors contributed to writing the manuscript. N.Y. is the guarantor of the manuscript.

Correspondence and requests for reprints should be addressed to Nadir Yehya, M.D., Department of Anesthesiology and Critical Care Medicine, Children's Hospital of Philadelphia, Wood Building 6040A, 3401 Civic Center Boulevard, Philadelphia, PA 19104. E-mail: yehyan@email.chop.edu.

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

Am J Respir Crit Care Med Vol 200, Iss 7, pp 828–836, Oct 1, 2019

Copyright © 2019 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.201810-2050CP on April 29, 2019

Internet address: www.atsjournals.org

Table 1. Iterations of Failure-Free Days in Emergency and Critical Care Trials

Outcome	Time Frame (d)	Primary or Secondary Outcome	References
Ventilator-free days	28	Primary	4, 6, 42, 43
	60	Primary	44
	180	Secondary	7
Vasopressor-free days	28	Secondary	45, 46, 47
Kidney failure-free days	28	Primary	48
Renal replacement therapy-free days	28	Secondary	46, 47, 49
Delirium-/coma-free days	14	Primary	50
Organ failure-free days	14	Secondary	51
	28	Secondary	4, 6, 45, 47
	28	Primary	22
ICU-free days	7	Secondary	52
	28	Primary	7
	180	Secondary	53
Hospital-free days	28	Primary	53

analysis of VFDs and other failure-free days is limited (8, 9). The goals of this review are to discuss the advantages and limitations of VFDs and to provide recommendations based on an assessment of available strategies to analyze and report the outcome of a study when duration of ventilation is determined to be a meaningful outcome measure and death is a competing event to be either accounted for or simultaneously assessed as a clinically meaningful outcome measure.

Rationale for VFDs

VFDs were considered for ARDS trials as early as the 1994 American-European Consensus Conference (10) and indeed offer several attractions. First, and most significantly, VFDs penalize nonsurvivors, thus making this a more defensible endpoint for trials than comparing total ventilator duration or ventilator duration in survivors (11). Second, as a continuous measure, VFDs provide greater statistical power to detect a treatment effect than the binary outcome measure of mortality (1). Third, ARDS is heterogeneous, with substantial mortality risk conferred by underlying comorbidities and concurrent nonpulmonary organ failures; only a minority of deaths are actually caused by ARDS itself (12, 13). Thus, it is unlikely that interventions predominantly targeting lung injury would have substantial effects on mortality, whereas such interventions should plausibly shorten ventilator duration (1). This rationale has become more salient as control group mortality has

decreased by half in ARDS trials in the 25 years since it was proposed in 1994 (14). Lower mortality has made the sample sizes required to detect meaningful reductions in mortality impractical (15). Fourth, VFDs implicitly assume that interventions that improve respiratory physiology will plausibly both shorten ventilator duration and improve mortality, thereby increasing efficiency as an outcome measure. Finally, shortened ventilator duration is clinically and economically meaningful. Patients will have less exposure to the risks of mechanical ventilation, including discomfort, sedation, delirium, neuromuscular weakness, and ventilator-associated infections, whereas payers see benefits to shorter ICU stays and fewer complications. Furthermore, prolonged ventilator duration is associated with both ICU and postdischarge morbidity (16) and mortality in adults (17–19), justifying the ventilator duration component of VFDs as a patient-centered outcome measure.

Limitations of Using Failure-Free Days

Although VFDs and other failure-free days offer inferential benefits, they also create several limitations. The main criticism is that a single composite risk estimate does not adequately distinguish between component risks (20). This is especially salient when one of the components (mortality) is considerably more important than the other component (ventilation duration) (Figure 1). Because subjects are assigned equal weight for death as for being

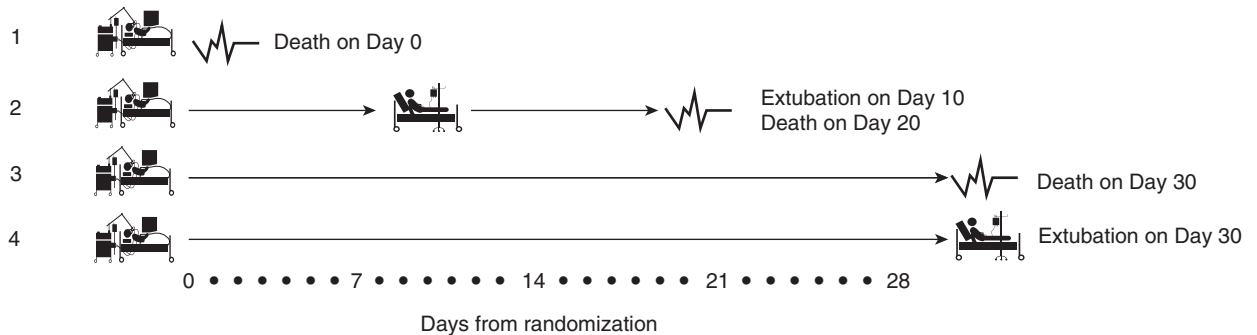
ventilated for 28 days or longer (21, 22), VFDs can give an ambiguous impression that an intervention reduces both death and ventilator duration, whereas the less important component of ventilator duration commonly drives the effect. This has been observed in several trials. For example, in the fluid management arm of the FACTT (Fluid and Catheter Treatment Trial) study, conservative fluid management resulted in 2.57 additional VFDs, driven almost entirely by shorter ventilator duration (22). In the LaSRS (Late Steroid Rescue Study) trial, methylprednisolone resulted in identical 60-day mortality (primary endpoint) but more VFDs at 28 days. The 28-day time frame masked that the greater VFDs were due to faster extubation in the methylprednisolone group (implying benefit), but this did not account for subsequent reintubation (threefold higher in the methylprednisolone group) and subsequent mortality.

This ambiguity associated with composite outcome measures is consistent with a recent assessment of the cardiology literature (23, 24), which suggested that composite outcomes have a large gradient of importance and that less important endpoints occur more frequently and thus drive effect estimates. For example, a trial of irbesartan versus amlodipine or placebo showed benefit of irbesartan over amlodipine for the composite outcome of doubling of creatinine, onset of end-stage renal disease, or death of any cause (25). However, doubling of creatinine was the most common component of the composite to occur, accounting for more than 50% of all component events. Furthermore, mortality was also nominally higher with irbesartan, calling into question the clinical utility of conclusions based on assessment of a single composite outcome measure.

A related issue with VFDs, as with all composite outcome measures, is that the gains in statistical efficiency using VFDs are realized only if the intervention affects each component in the same direction (1), an assumption that is not always explicitly assessed before using VFDs. This was the problem encountered during the LaSRS trial, wherein methylprednisolone shortened ventilator duration up to 28 days but also contributed to reintubations and later mortality, resulting in a misleading interpretation of VFDs that inappropriately

A Ventilator-free days = 0

Scenario

**B** Ventilator-free days = 14

Scenario

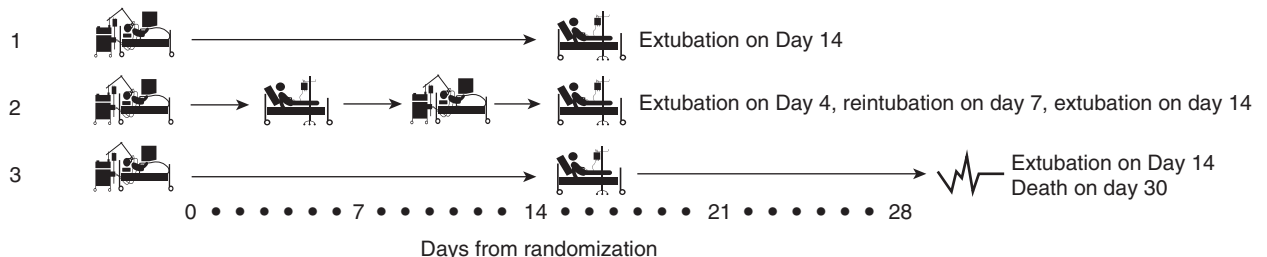


Figure 1. Different clinical trajectories of subjects assigned to (A) 0 or (B) 14 ventilator-free days. A criticism of ventilator-free days is that the “net effect” being reported does not adequately discriminate between these distinct patient outcomes.

cast methylprednisolone as advantageous. Mortality and ventilator duration being oppositely affected by an intervention negates the efficiency and utility of VFDs, rendering a trial uninterpretable.

Thus, a risk of using failure-free day composites is to make an intervention appear more efficacious than it actually is. This problem is compounded by poor reporting of the components of VFDs, leaving readers uncertain about which component is driving the effect. The term “VFDs” itself is confusing because the reference to “days” seems to imply differences in ventilator duration rather than mortality. Another concern with how VFDs are used is the lack of a standardized definition. In an analysis of 55 papers reporting VFDs, only 34 (62%) detailed how VFDs were calculated, with 13 different definitions used (9). Sources of heterogeneity included varying start time (intubation or randomization), accounting for reintubations (interval extubations before Day 28), death before Day 28 if the patient was extubated, death after Day 28 if

the subject was extubated before Day 28, and noninvasive ventilation. This variation in definitions maligns the comparison of results across different studies and precludes accurate estimates in meta-analyses.

Is There a Place for Failure-Free Days?

Given the limitations inherent to VFDs and other failure-free days as outcome measures, two main criteria should be met before using this composite as a primary outcome measure. First, the simpler and more clinically meaningful endpoint of mortality should be impractical. In many situations, mortality is too infrequent to be useful. In pediatric ARDS, for instance, mortality is consistently less than 20% (26, 27). This does not necessarily imply that VFDs should be used; it just means that VFDs should be considered as one of several potential approaches, each with its own strengths and limitations. Strategies for prognostic enrichment could mitigate some

of the concerns regarding low baseline mortality rates. Recent trials of neuromuscular blockade (28) and prone positioning (29) in adult ARDS restricted enrollment to subjects with Pa_{O_2}/Fi_{O_2} less than 150, with double the baseline mortality rates of other contemporary ARDS trials enrolling subjects with Pa_{O_2}/Fi_{O_2} less than or equal to 300 (4, 6), thereby allowing both trials to use mortality as the primary endpoint.

Second, it must be clinically and biologically plausible that the tested intervention affects both ventilator duration (or organ failure or ICU stay) in the same direction as mortality. VFDs cannot be chosen solely because mortality is impractical or infrequent; the efficiency of this composite outcome is best realized when the two components (mortality and ventilator duration) are both improved simultaneously by the tested treatment, even if only nominally. This is significant not only for the statistical performance of VFDs; it is also critical for acceptance of trial results using this outcome. Readers and

practitioners should feel confident that VFDs appropriately capture the combined effect of an experimental intervention on both mortality and ventilator duration. For some tested interventions in ARDS, such as conservative fluid balance (22), it is plausible that the mechanism by which the intervention shortens mechanical ventilation also benefits mortality, given the associations between fluid overload and mortality. However, this is not always the case, nor is it always obvious prospectively. For instance, much of the preclinical rationale for inhaled β -agonists in ARDS relied on physiologic surrogates, such as improved pulmonary mechanics (30, 31) and faster clearance of extravascular lung water (32). Because impaired extravascular lung water clearance was implicated in prolonging ARDS and thus mechanical ventilation, it was reasonable to test inhaled β -agonists to shorten ventilator duration in the ALTA (Albuterol for the Treatment of Acute Lung Injury) trial (4). However, the off-target effects of β -agonists, specifically tachycardias and

arrhythmias, may have resulted in the higher mortality seen with albuterol in ALTA. This would have been difficult to appreciate in preclinical models, most of which did not assess mortality. However, the increased rate of arrhythmias with intravenous β -agonists in a previous pilot trial (33) could have cautioned that the off-target effects of albuterol may have impacted outcome, specifically mortality, requiring a reassessment of whether VFDs were the appropriate outcome measure in ALTA.

The requirement that a tested intervention simultaneously improve both mortality and ventilator duration is more complex than it appears. Consider a thought experiment of only two types of patients: those who will die and those who will survive absent an intervention. In the presence of an intervention, it is plausible that subjects who would have died will now survive, albeit at the expense of longer ventilation among those who would have survived. This scenario is commonplace (e.g., extracorporeal support), and in situations

such as this, mortality should be chosen as the appropriate endpoint. VFDs would be defensible if, in addition to reducing mortality among those who would have died, the intervention also reduces ventilator duration among those who would have survived.

Analytic Methods

Upon deciding to use VFDs, researchers must then decide how to appropriately analyze this outcome measure. VFDs are a problematic variable to analyze, with a skewed distribution and inflation of both 0s and 28s (34). By definition, “0 VFDs” can refer to a nonsurvivor who died within 24 hours of randomization or to a survivor ventilated for greater than or equal to 28 days, which complicates the interpretation of the efficacy of an intervention.

Traditional Methods of Analyzing VFDs

Traditional strategies for analyzing VFDs include Student’s *t* test (and analysis of

Table 2. Power Calculations for Different Statistical Tests in Which Primary Outcome of Interest Is Ventilator Duration Censored at 28 Days

Effects	Mortality [†]	Ventilator Days among Survivors [‡] (Mean)	Power*					
			Fine and Gray Regression	Gray’s Test	Log-Rank Test [§]	Rank-Sum Test	Student’s <i>t</i> Test	Fisher’s Exact Test [¶]
Mortality only								
Treatment	15%	7	76%	75%	76%	55%	71%	85%
Control	25%	7						
Strong mortality and weak duration			94%	94%	94%	89%	94%	85%
Treatment	15%	6						
Control	25%	7						
Moderate mortality and duration			79%	79%	79%	84%	81%	33%
Treatment	15%	5						
Control	20%	6.5						
Weak mortality and strong duration			85%	84%	85%	97%	90%	5%
Treatment	15%	5						
Control	16%	8						
Duration only			79%	77%	79%	95%	86%	4%
Treatment	15%	5						
Control	15%	8						
Conflicting			5%	5%	5%	14%	5%	33%
Treatment	15%	6.5						
Control	20%	5						

The highest power for any scenario is in bold.

*Results are each based on 3,000 simulated trials with 300 subjects in each of two treatment groups, a two-sided alternative hypothesis, and a type I error rate of $\alpha = 0.05$.

[†]Mortality is simulated according to a Bernoulli distribution.

[‡]Duration of ventilation among survivors is simulated according to an exponential distribution.

[§]Deaths were set as higher than any duration for log-rank test.

^{||}Owing to computational limits, the normal approximation with continuity correction was used for the Wilcoxon rank-sum test.

[¶]For Fisher’s exact test, the outcome is mortality; duration of ventilation is ignored. It is provided here for comparison with the other tests.

variance) and the Wilcoxon rank-sum test. Student's *t* tests permit the comparison of means, and when extended to analysis of variance, can accommodate the inclusion of baseline variables. One criticism of Student's *t* tests is that VFDs are treated as intervals; that is, the difference between 0 and 1 VFD is the same as the difference between 1 and 2 VFDs, even though a VFD of 0 is assigned for death and other values indicate 28-day survival. The importance of 2 VFDs versus 1 may not be important, but the difference between death and life is paramount. A similar criticism could be made for every approach to testing that we consider. Although it does not invalidate the approach, it does make interpreting differences in average VFDs problematic, regardless of sample size. Additional issues that can impact the results of a Student's *t* test are the inherent skew and zero inflation of VFDs. This can adversely impact power, even with large sample sizes (2).

Rank-sum tests are also used for comparing VFDs (1). One advantage is that power is less impacted by a skewed distribution of VFDs. Disadvantages of the rank-sum test are that it does not provide a

measure of the magnitude of treatment effect and does not readily lend itself to interim monitoring using the α -spending approach. Although the rank-sum test can be stratified by categorical variables, it cannot adjust for continuous variables. Effect size is often reported as a difference in medians, which is problematic. There is a common misconception that the rank-sum test is a test of the equality of medians. Although commonly reported as such, the rank-sum test is not based on differences of medians. Rather, it tests whether outcomes in one group tend to be better or worse than outcomes in the comparator, which can occur even if medians are identical. Finally, the mortality component (i.e., VFDs = 0) is critically important but will have little effect on the median, making the reporting of median VFDs of dubious value.

VFDs Are a Competing Risk Problem

VFDs can also be evaluated with a time-to-event analysis censored at 28 days, with the event of interest as extubation and mortality a competing event. Fine and Gray

competing risk regression is used to assess VFDs in this framework (35) (see online supplement). Competing risk regression addresses the situation when more than one mutually exclusive endpoint is possible: in this case, successful extubation or death. This analysis provides a subdistribution hazard ratio (SHR), the magnitude of which is affected by both the time to extubation and the probability of death. SHR assesses the association between an intervention and extubation, accounting for the existence of the alternative outcome of death. Advantages of this approach are that SHR measures an appropriate effect size and that the regression readily accommodates covariates for improved precision of treatment effects (36–38). The primary disadvantage is that competing risk regression, such as Cox regression, relies on the proportional hazards assumption.

Relative Power of Statistical Tests

Relative power of different tests in 3,000 simulations of a two-arm trial with $n = 300$ per arm was computed (Table 2). We varied whether the effect was driven by

Table 3. Recommendations for Defining, Analyzing, and Reporting Ventilator-Free Days

Recommendation	Rationale	Comments
Define VFDs explicitly	Facilitates comparison between and across interventions, trials, and meta-analyses	Day 0 (day of randomization) Time frame (28 d) Successful extubation (extubation >48 h without reintubation in a 28-d survivor) Interval reintubations (count from last successful extubation) Death before 28 d (VFD = 0 to penalize nonsurvival, regardless of intubation status) Death after 28 d (censor after 28 d; use 28-d ventilation and survival status for calculating VFDs) Noninvasive support (do not count) Tracheostomy (treat as all invasive ventilation)
Use competing risk regression to analyze VFDs	Valid, comprehensible estimate of the combined effect; allows adjustment for baseline variables; allows for interim monitoring	The power calculation is for the "net effect" size (SHR) in which duration of ventilation is the primary outcome and death is the competing event Wilcoxon rank-sum test has higher power if effect is primarily driven by ventilator duration
Report effect size, confidence interval, and <i>P</i> value of the composite; graph the cumulative incidence function	Reporting of the primary "net effect" size and confidence of estimate	Competing risk allows this (SHR); Wilcoxon rank-sum test does not
Report effect size, confidence interval, and <i>P</i> value of each component individually (mortality and ventilator duration)	Transparent reporting of whether one or both components is driving the effect	Adjust for the same baseline factors as in the primary analysis of the composite Report cumulative incidence function for the components of interest

Definition of abbreviations: SHR = subdistribution hazard ratio; VFD = ventilator-free days.

mortality, ventilator duration, both, or in opposite directions (see additional details in the online supplement). Fine and Gray competing risk regression had higher power than the rank-sum test when there was a dominant mortality signal, whereas the rank-sum test had slightly higher power when ventilator duration drove the effect. None of the tests were powerful when component effects moved in conflicting directions. Gray's test and the log-rank test, which also treat VFDs as a time-to-event analysis (see online supplement), have comparable power to Fine and Gray competing risk regression.

The Student's *t* test had reasonable power when VFDs at 28 days were evaluated (Table 2). However, when VFDs at 60 or 90 days were considered, the Student's *t* test performed poorly, especially when the effect was driven by duration of ventilation (Tables E1 and E2 in the online supplement).

A Framework for Improvement

We have four broad recommendations for reporting failure-free days (Table 3): 1) define the outcome measure (e.g., VFDs) explicitly, 2) use competing risk regression for analysis, 3) report the main effect of the

composite, and 4) report the components. First, VFDs need to be explicitly defined. Building on prior studies (9), we suggest the standards outlined below for VFDs.

Start Time

For clinical trials, we suggest using day of randomization as Day 0 and reporting duration of intubation before randomization as a baseline variable (9). For observational studies, day of ARDS onset or day of intubation is acceptable. Day of intubation may also be used as Day 0 in trials with very short time frames between intubation and randomization.

Time Frame

We suggest using 28 days. Durations longer than 28 days are rarely justified for VFDs or most organ failure-free days, because the majority of extubations, organ failure resolutions, or deaths occur within 28 days. Use of VFDs at 60 or more days further increases the skew and can adversely impact power.

Successful Extubation

We suggest using extubation more than 48 hours as success, as is done in most adult trials (4, 21, 39–41). Pediatric trials have used both >24 hours (42) and >48 hours (43) without reintubation as cutoffs. Per our

recommendations for how to handle 28-day nonsurvivors below, "successful extubation" also implies survival to at least 28 days.

Interval Extubation

We suggest counting from the day of final successful extubation if there were repeat intubation episodes in the first 28 days, as has been reported in adult trials (6). This prioritizes the clinical relevance of VFDs because patients are not credited for interval extubations. Not crediting modest interval extubations is also consistent with the recent WIND (Weaning according to a New Definition) study (34).

Noninvasive Support and Tracheostomies

We recommend *not* counting noninvasive support, and we suggest that tracheostomies should be treated as other invasive ventilation (i.e., >48 h off of positive pressure constitutes success).

Value for Extubated Decedent

We recommend assigning all 28-day nonsurvivors 0 VFDs, regardless of their intubation status, and censoring observations after 28 days. Assigning all nonsurvivors 0 VFDs appropriately penalizes mortality. Censoring after 28 days (thus ignoring deaths >28 d)

Table 4. Proposed Alternative Reporting of Trial Data in Which Primary Outcome of Interest Is Ventilator Duration

	ARMA	ALVEOLI	FACTT	ALTA	OMEGA
Intervention	Low V _T	Higher PEEP	Conservative	Albuterol	Supplements
Control	High V _T	Lower PEEP	Liberal	Placebo	Formula
Extubated alive (composite)					
SHR	1.30	0.91	1.30	0.79	0.73
95% CI	1.09 to 1.54	0.75 to 1.11	1.12 to 1.51	0.60 to 1.02	0.56 to 0.97
<i>P</i> value	0.003	0.356	<0.001	0.072	0.027
28-d mortality					
Intervention	24%	22%	26%	20%	22%
Control	34%	22%	29%	14%	13%
RR 28-d mortality	0.70	1.04	0.86	1.47	1.70
95% CI	0.57 to 0.87	0.76 to 1.42	0.69 to 1.08	0.87 to 2.51	0.99 to 2.91
<i>P</i> value	0.001	0.810	0.186	0.159	0.049
Mean ± SD ventilator days in 28-d survivors					
Intervention	8.9 ± 7.1	8.3 ± 6.0	8.2 ± 5.8	6.8 ± 6.0	7.3 ± 5.2
Control	8.6 ± 7.8	8.7 ± 6.2	10.3 ± 6.4	7.1 ± 5.9	6.9 ± 5.1
Δ Mean ventilator days in survivors	0.26	−0.32	−2.02	−0.24	0.46
95% CI	−0.96 to 1.47	−1.57 to 0.93	−2.95 to −1.10	−1.88 to 1.39	−1.01 to 1.93
Student's <i>t</i> test <i>P</i> value	0.680	0.619	<0.001	0.770	0.539

Definition of abbreviations: ALTA=Albuterol for the Treatment of Acute Lung Injury; ALVEOLI=Assessment of Low Tidal Volume and Elevated End-expiratory Volume to Obviate Lung Injury; ARMA=Respiratory Management in ARDS; CI=confidence interval; FACTT=Fluid and Catheter Treatment Trial; OMEGA=Omega Nutritional Supplement Trial; PEEP=positive end-expiratory pressure; RR=risk ratio; SHR=subdistribution hazard ratio.

reflects that the primary aim of VFDs is to capture the effect of an intervention on the combined 28-day ventilator duration and mortality rather than mortality at any time point. Thus, the subject's status at Day 28 should determine their VFDs.

Second, the analytic technique should be explicitly described and appropriate to the requirements of the study. We recommend competing risk regression for most analyses because it provides good power, allows for adjustment of covariates or stratification variables, allows interim monitoring, and provides a meaningful overall effect size (SHR).

Finally, we recommend reporting both the main effect and the components of the composite outcome measure. Detailed reporting of the components is often omitted from analyses. The main effect would be reported as either the adjusted or unadjusted SHR with 95% confidence intervals and a *P* value. To illustrate our recommendations for reporting and analyzing VFDs, we present a reanalysis of five ARDSNet (Acute Respiratory Distress Syndrome Clinical Network) trials in which we define VFDs as recommended above

and use competing risk regression (Table 4): ARMA (Respiratory Management in ARDS) (21), ALVEOLI (Assessment of Low Tidal Volume and Elevated End-expiratory Volume to Obviate Lung Injury) (39), the fluid management arm of FACTT (22), ALTA (4), and OMEGA (Omega Nutritional Supplement Trial) (6). These trials were chosen because VFDs were either a primary or coprimary outcome or were the major positive finding in the study. As a complement to reporting results in this manner, plotting the cumulative incidence function also provides insights into the time dependency of the effect size and absolute risk estimates at specific time points (Figure 2). If the two curves cross, it indicates that the probability of successful extubation (accounting for mortality) was initially higher in one group but later higher in the other. In all scenarios, the reported SHR would be interpreted as the average SHR.

These analyses illustrate the conceptual and practical advantages of this approach. First, this allows an estimation of effect size (SHR) with confidence intervals. Second, the SHR can be presented as an effect estimate

similar to the hazard ratio, a commonly used approach to communicate efficacy data. In ALTA and FACTT, for instance, the intervention group had a 30% higher rate of intact extubation (SHR, 1.30), whereas in OMEGA, the intervention group had a 27% lower rate (SHR, 0.73). We argue that this interpretation is more meaningful and less prone to misinterpretation than a report of additional (or fewer) VFDs, which, because VFDs include the term "days," is prone to misinterpretation. Third, relative contributions of mortality and ventilator duration are explicit in this framework. ARMA and FACTT have identical SHRs, but the benefit of low V_T in ARMA is driven entirely by mortality, whereas the efficacy of conservative fluids in FACTT is primarily due to shortening ventilator days. Finally, competing risk regression accommodates interim monitoring that incorporates partial outcome data on subjects who are alive and intubated (i.e., censored at time of interim analysis). Furthermore, the test statistic has independent increments, which simplifies calculation of early stopping rules.

Disadvantages of this approach are that 1) statistical power may be lower when the

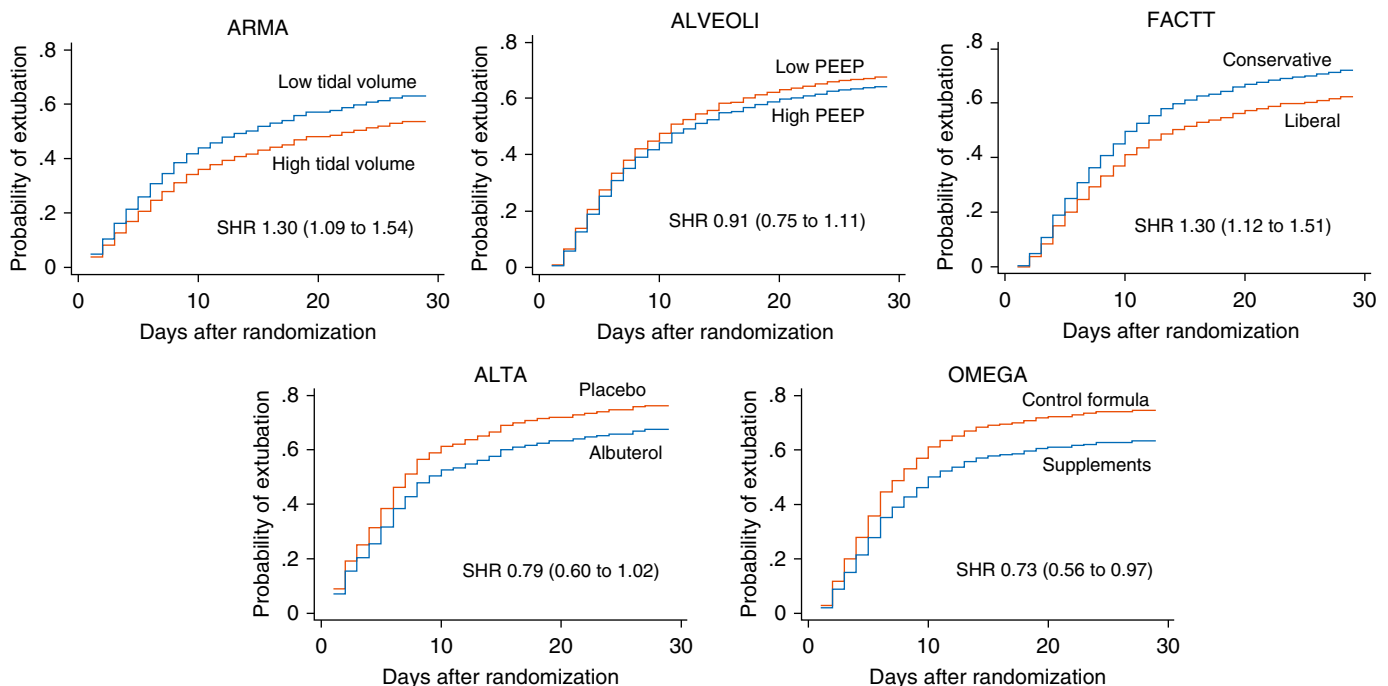


Figure 2. Cumulative incidence functions for the primary event (extubation) in five ARDSNet (Acute Respiratory Distress Syndrome Clinical Network) trials. The subdistribution hazard ratio (SHR) and 95% confidence intervals are provided. Intervention (blue) and control (red) arms are displayed, and SHR greater than 1 is interpreted as greater hazard of intact extubation. ARMA (Respiratory Management in ARDS) and FACTT (Fluid and Catheter Treatment Trial) demonstrated a benefit of the intervention related to the probability of extubation, whereas OMEGA (Omega Nutritional Supplement Trial) demonstrated harm. ALTA = Albuterol for the Treatment of Acute Lung Injury; ALVEOLI = Assessment of Low Tidal Volume and Elevated End-expiratory Volume to Obviate Lung Injury; PEEP = positive end-expiratory pressure.

effect is driven primarily by ventilator duration and 2) valid inference depends on underlying assumptions. The key assumption for competing risk regression is the proportionality of hazards. However, if this assumption is not met, the SHR can be interpreted as an average SHR. Nonproportionality would also be visually evident in a cumulative incidence function plot.

When to Use the Wilcoxon Rank-Sum Test

When the effect of an intervention on VFDs is primarily through shortened duration of ventilation rather than mortality, the Wilcoxon rank-sum test has higher power than competing risk regression (Table 2). If this is suspected *a priori*, rank-sum tests may be preferred in these scenarios, particularly if covariate adjustment or interim monitoring is not needed. Specific situations, such as stratified randomization by center or severity-of-illness score, are particularly difficult to adjust for at the analysis stage using the rank-sum test.

Conclusions

It can reasonably be questioned whether VFDs and similar composite outcome measures should continue to be used, given their limitations, and the components should be required to be reported. However, for certain populations (e.g., pediatric patients), there is little choice, because interventions are postulated to improve some physiology or organ dysfunction, and yet death occurs commonly enough that it must be accounted for in analysis. We propose that when mortality alone is not a practical endpoint, and if an intervention is hypothesized to improve both mortality and ventilator duration on the basis of strong biological plausibility, then VFDs are an efficient and useful method to assess the *combined effect* of an intervention on ventilator duration and mortality. In this paper, we outline situations in which VFDs may be considered, and in those situations, we provide guidance for definition, analysis, and reporting of VFDs. These

recommendations are broadly applicable to other failure-free day outcomes, although less common metrics (e.g., delirium-free or hospital-free days) require more investigation to assess the optimal time frame (14 vs. 28 vs. 60 d) to ensure that most subjects will experience the outcome event within the time frame and to minimize skew. For most organ failure- and vasopressor-free days, a 28-day time frame is likely most appropriate. Finally, we believe that quantifying treatment effect in critical care in a competing risk framework offers several inferential and statistical benefits. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: This article was prepared using ALTA, ALVEOLI, ARMA, FACTT, and OMEGA research materials obtained from the NHLBI Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC) and does not necessarily reflect the opinions or views of either the parent studies or the NHLBI.

References

- Schoenfeld DA, Bernard GR; ARDS Network. Statistical evaluation of ventilator-free days as an efficacy measure in clinical trials of treatments for acute respiratory distress syndrome. *Crit Care Med* 2002;30:1772-1777.
- Schoenfeld DA, Hayden D, Oldmixon C, Ringwood N, Thompson BT. Statistical design and analysis issues for the ARDS Clinical Trials Network: the coordinating center perspective. *Clin Invest* 2012;2: 275-289.
- Bernard GR, Wheeler AP, Arons MM, Morris PE, Paz HL, Russell JA, et al.; The Antioxidant in ARDS Study Group. A trial of antioxidants N-acetylcysteine and procysteine in ARDS. *Chest* 1997;112: 164-172.
- Matthay MA, Brower RG, Carson S, Douglas IS, Eisner M, Hite D, et al.; National Heart, Lung, and Blood Institute Acute Respiratory Distress Syndrome (ARDS) Clinical Trials Network. Randomized, placebo-controlled clinical trial of an aerosolized β_2 -agonist for treatment of acute lung injury. *Am J Respir Crit Care Med* 2011;184:561-568.
- Rice TW, Wheeler AP, Thompson BT, Steingrub J, Hite RD, Moss M, et al.; National Heart, Lung, and Blood Institute Acute Respiratory Distress Syndrome (ARDS) Clinical Trials Network. Initial trophic vs full enteral feeding in patients with acute lung injury: the EDEN randomized trial. *JAMA* 2012;307:795-803.
- Rice TW, Wheeler AP, Thompson BT, deBoisblanc BP, Steingrub J, Rock P; NIH NHLBI Acute Respiratory Distress Syndrome Network of Investigators. Enteral omega-3 fatty acid, gamma-linolenic acid, and antioxidant supplementation in acute lung injury. *JAMA* 2011;306: 1574-1581.
- Steinberg KP, Hudson LD, Goodman RB, Hough CL, Lanke PN, Hyzy R, et al.; National Heart, Lung, and Blood Institute Acute Respiratory Distress Syndrome (ARDS) Clinical Trials Network. Efficacy and safety of corticosteroids for persistent acute respiratory distress syndrome. *N Engl J Med* 2006;354:1671-1684.
- Bodet-Contentin L, Frasca D, Tavernier E, Feuillet F, Foucher Y, Giraudeau B. Ventilator-free day outcomes can be misleading. *Crit Care Med* 2018;46:425-429.
- Contentin L, Ehrmann S, Giraudeau B. Heterogeneity in the definition of mechanical ventilation duration and ventilator-free days. *Am J Respir Crit Care Med* 2014;189:998-1002.
- Bernard GR, Artigas A, Brigham KL, Carlet J, Falke K, Hudson L, et al. The American-European Consensus Conference on ARDS: definitions, mechanisms, relevant outcomes, and clinical trial coordination. *Am J Respir Crit Care Med* 1994;149: 818-824.
- Harhay MO, Ratcliffe SJ, Small DS, Suttner LH, Crowther MJ, Halpern SD. Measuring and analyzing length of stay in critical care trials. *Med Care* 2019;57:e53-e59.
- Stapleton RD, Wang BM, Hudson LD, Rubenfeld GD, Caldwell ES, Steinberg KP. Causes and timing of death in patients with ARDS. *Chest* 2005;128:525-532.
- Dowell JC, Parvathaneni K, Thomas NJ, Khemani RG, Yehya N. Epidemiology of cause of death in pediatric acute respiratory distress syndrome. *Crit Care Med* 2018;46:1811-1819.
- Spragg RG, Bernard GR, Checkley W, Curtis JR, Gajic O, Guyatt G, et al. Beyond mortality: future clinical research in acute lung injury. *Am J Respir Crit Care Med* 2010;181:1121-1127.
- Harhay MO, Wagner J, Ratcliffe SJ, Bronheim RS, Gopal A, Green S, et al. Outcomes and statistical power in adult critical care randomized trials. *Am J Respir Crit Care Med* 2014;189: 1469-1478.
- Herridge MS, Chu LM, Matte A, Tomlinson G, Chan L, Thomas C, et al.; RECOVER Program Investigators (Phase 1: towards RECOVER); Canadian Critical Care Trials Group. The RECOVER program: disability risk groups and 1-year outcome after 7 or more days of mechanical ventilation. *Am J Respir Crit Care Med* 2016;194: 831-844.
- Combes A, Costa MA, Trouillet JL, Baudot J, Mokhtari M, Gibert C, et al. Morbidity, mortality, and quality-of-life outcomes of patients requiring ≥ 14 days of mechanical ventilation. *Crit Care Med* 2003;31:1373-1381.
- Unroe M, Kahn JM, Carson SS, Govert JA, Martinu T, Sathy SJ, et al. One-year trajectories of care and resource utilization for recipients of prolonged mechanical ventilation: a cohort study. *Ann Intern Med* 2010;153:167-175.

19. Hill AD, Fowler RA, Burns KE, Rose L, Pinto RL, Scales DC. Long-term outcomes and health care utilization after prolonged mechanical ventilation. *Ann Am Thorac Soc* 2017;14:355–362.
20. Varadhan R, Weiss CO, Segal JB, Wu AW, Scharfstein D, Boyd C. Evaluating health outcomes in the presence of competing risks: a review of statistical methods and clinical applications. *Med Care* 2010;48(Suppl):S96–S105.
21. Brower RG, Matthay MA, Morris A, Schoenfeld D, Thompson BT, Wheeler A; Acute Respiratory Distress Syndrome Network. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. *N Engl J Med* 2000;342:1301–1308.
22. Wiedemann HP, Wheeler AP, Bernard GR, Thompson BT, Hayden D, deBoisblanc B, et al.; National Heart, Lung, and Blood Institute Acute Respiratory Distress Syndrome (ARDS) Clinical Trials Network. Comparison of two fluid-management strategies in acute lung injury. *N Engl J Med* 2006;354:2564–2575.
23. Manja V, AlBashir S, Guyatt G. Criteria for use of composite end points for competing risks—a systematic survey of the literature with recommendations. *J Clin Epidemiol* 2017;82:4–11.
24. Ferreira-González I, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, Bryant DM, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *BMJ* 2007;334:786.
25. Lewis EJ, Hunsicker LG, Clarke WR, Berl T, Pohl MA, Lewis JB, et al.; Collaborative Study Group. Renoprotective effect of the angiotensin-receptor antagonist irbesartan in patients with nephropathy due to type 2 diabetes. *N Engl J Med* 2001;345:851–860.
26. Yehya N, Keim G, Thomas NJ. Subtypes of pediatric acute respiratory distress syndrome have different predictors of mortality. *Intensive Care Med* 2018;44:1230–1239.
27. Khemani RG, Smith L, Lopez-Fernandez YM, Kwok J, Morzov R, Klein MJ, et al.; Pediatric Acute Respiratory Distress Syndrome Incidence and Epidemiology (PARDIE) Investigators; Pediatric Acute Lung Injury and Sepsis Investigators (PALISI) Network. Paediatric Acute Respiratory Distress Syndrome Incidence and Epidemiology (PARDIE): an international, observational study. *Lancet Respir Med* 2019;7:115–128.
28. Papazian L, Forel JM, Gacouin A, Penot-Ragon C, Perrin G, Loundou A, et al.; ACURASYS Study Investigators. Neuromuscular blockers in early acute respiratory distress syndrome. *N Engl J Med* 2010;363:1107–1116.
29. Guérin C, Reignier J, Richard JC, Beuret P, Gacouin A, Boulain T, et al.; PROSEVA Study Group. Prone positioning in severe acute respiratory distress syndrome. *N Engl J Med* 2013;368:2159–2168.
30. Khimenko PL, Barnard JW, Moore TM, Wilson PS, Ballard ST, Taylor AE. Vascular permeability and epithelial transport effects on lung edema formation in ischemia and reperfusion. *J Appl Physiol* (1985) 1994;77:1116–1121.
31. Palmieri TL, Enkhbaatar P, Bayliss R, Traber LD, Cox RA, Hawkins HK, et al. Continuous nebulized albuterol attenuates acute lung injury in an ovine model of combined burn and smoke inhalation. *Crit Care Med* 2006;34:1719–1724.
32. McAuley DF, Frank JA, Fang X, Matthay MA. Clinically relevant concentrations of β_2 -adrenergic agonists stimulate maximal cyclic adenosine monophosphate-dependent airspace fluid clearance and decrease pulmonary edema in experimental acid-induced lung injury. *Crit Care Med* 2004;32:1470–1476.
33. Perkins GD, McAuley DF, Thickett DR, Gao F. The Beta-Agonist Lung Injury Trial (BALTI): a randomized placebo-controlled clinical trial. *Am J Respir Crit Care Med* 2006;173:281–287.
34. Béduneau G, Pham T, Schortgen F, Piquilloud L, Zogheib E, Jonas M, et al.; WIND (Weaning according to a New Definition) Study Group and the REVA (Réseau Européen de Recherche en Ventilation Artificielle) Network. Epidemiology of weaning outcome according to a new definition: the WIND study. *Am J Respir Crit Care Med* 2017;195:772–783.
35. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc* 1999;94:496–509.
36. Hauck WW, Anderson S, Marcus SM. Should we adjust for covariates in nonlinear regression analyses of randomized trials? *Control Clin Trials* 1998;19:249–256.
37. Raab GM, Day S, Sales J. How to select covariates to include in the analysis of a clinical trial. *Control Clin Trials* 2000;21:330–342.
38. Steingrimsson JA, Hanley DF, Rosenblum M. Improving precision by adjusting for prognostic baseline variables in randomized trials with binary outcomes, without regression model assumptions. *Contemp Clin Trials* 2017;54:18–24.
39. Brower RG, Lanken PN, MacIntyre N, Matthay MA, Morris A, Ancukiewicz M, et al.; National Heart, Lung, and Blood Institute ARDS Clinical Trials Network. Higher versus lower positive end-expiratory pressures in patients with the acute respiratory distress syndrome. *N Engl J Med* 2004;351:327–336.
40. Thille AW, Richard JC, Brochard L. The decision to extubate in the intensive care unit. *Am J Respir Crit Care Med* 2013;187:1294–1302.
41. Ruan SY, Teng NC, Wu HD, Tsai SL, Wang CY, Wu CP, et al. Durability of weaning success for liberation from invasive mechanical ventilation: an analysis of a nationwide database. *Am J Respir Crit Care Med* 2017;196:792–795.
42. Willson DF, Thomas NJ, Markovitz BP, Bauman LA, DiCarlo JV, Pon S, et al.; Pediatric Acute Lung Injury and Sepsis Investigators. Effect of exogenous surfactant (calfactant) in pediatric acute lung injury: a randomized controlled trial. *JAMA* 2005;293:470–476.
43. Curley MA, Hibberd PL, Fineman LD, Wypij D, Shih MC, Thompson JE, et al. Effect of prone positioning on clinical outcomes in children with acute lung injury: a randomized controlled trial. *JAMA* 2005;294:229–237.
44. Bein T, Weber-Carstens S, Goldmann A, Müller T, Staudinger T, Brederlau J, et al. Lower tidal volume strategy (≈ 3 ml/kg) combined with extracorporeal CO_2 removal versus 'conventional' protective ventilation (6 ml/kg) in severe ARDS: the prospective randomized Xtravent-study. *Intensive Care Med* 2013;39:847–856.
45. Annane D, Renault A, Brun-Buisson C, Megarbane B, Quenot JP, Siami S, et al.; CRICS-TRIGGERSEP Network. Hydrocortisone plus fludrocortisone for adults with septic shock. *N Engl J Med* 2018;378:809–818.
46. De Backer D, Biston P, Devriendt J, Madl C, Chochrad D, Aldecoa C, et al.; SOAP II Investigators. Comparison of dopamine and norepinephrine in the treatment of shock. *N Engl J Med* 2010;362:779–789.
47. Russell JA, Walley KR, Singer J, Gordon AC, Hébert PC, Cooper DJ, et al.; VASST Investigators. Vasopressin versus norepinephrine infusion in patients with septic shock. *N Engl J Med* 2008;358:877–887.
48. Gordon AC, Mason AJ, Thirunavukkarasu N, Perkins GD, Cecconi M, Cepkova M, et al.; VANISH Investigators. Effect of early vasopressin vs norepinephrine on kidney failure in patients with septic shock: the VANISH randomized clinical trial. *JAMA* 2016;316:509–518.
49. Semler MW, Self WH, Wanderer JP, Ehrenfeld JM, Wang L, Byrne DW, et al.; SMART Investigators and the Pragmatic Critical Care Research Group. Balanced crystalloids versus saline in critically ill adults. *N Engl J Med* 2018;378:829–839.
50. Page VJ, Casarin A, Ely EW, Zhao XB, McDowell C, Murphy L, et al. Evaluation of early administration of simvastatin in the prevention and treatment of delirium in critically ill patients undergoing mechanical ventilation (MoDUS): a randomised, double-blind, placebo-controlled trial. *Lancet Respir Med* 2017;5:727–737.
51. Dulhunty JM, Roberts JA, Davis JS, Webb SA, Bellomo R, Gomersall C, et al.; BLING II Investigators for the ANZICS Clinical Trials Group. A multicenter randomized trial of continuous versus intermittent β -lactam infusion in severe sepsis. *Am J Respir Crit Care Med* 2015;192:1298–1305.
52. Agus MS, Wypij D, Hirshberg EL, Srinivasan V, Faustino EV, Luckett PM, et al.; HALF-PINT Study Investigators and the PALISI Network. Tight glycemic control in critically ill children. *N Engl J Med* 2017;376:729–741.
53. Self WH, Semler MW, Wanderer JP, Wang L, Byrne DW, Collins SP, et al.; SALT-ED Investigators. Balanced crystalloids versus saline in noncritically ill adults. *N Engl J Med* 2018;378:819–828.