Alive and Ventilator Free: A Hierarchical, Composite Outcome for Clinical Trials in the Acute Respiratory Distress Syndrome*

Victor Novack, MD, PhD[1,2]; Jeremy R. Beitler, MD, MPH[3]; Maayan Yitshak-Sade, PhD[1,4]; B. Taylor Thompson, MD[5]; David A. Schoenfeld, PhD[6]; Gordon Rubenfeld, MD[7]; Daniel Talmor, MD, MPH[2]; Samuel M. Brown, MD, MS[8]

*See also p. 257.

[1]Clinical Research Center, Soroka University Medical Center, Beer Sheva, Israel.

[2]Department of Anesthesia, Critical Care and Pain Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA.

[3]Center for Acute Respiratory Failure and Division of Pulmonary, Allergy, and Critical Care Medicine, Columbia University College of Physicians and Surgeons, New York, NY.

[4]Exposure, Epidemiology, and Risk Program, Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA.

[5]Division of Pulmonary and Critical Care Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA.

[6]Biostatistics Center, Massachusetts General Hospital and Harvard Medical School, Boston, MA.

[7]Sunnybrook Hospital Toronto, Toronto, ON, Canada.

[8]Division of Pulmonary and Critical Care Medicine, Intermountain Medical Center and University of Utah School of Medicine, Salt Lake City, UT.

Drs. Novack, Beitler, and Brown were involved in drafting the work. Drs. Yitshak-Sade, Thompson, Schoenfeld, Rubenfeld, and Talmor were involved in critical revision for important intellectual content. Drs. Novack, Beitler, Yitshak-Sade, and Brown were involved in statistical analysis. All authors were involved in substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work. All authors were involved in final approval of the version to be published. All authors were involved in agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (http://journals.lww.com/ccmjournal).

Supported, in part, by grant from the National Institutes of Health.

Dr. Novack received funding from Cardiomed Consultants LLC. Drs. Beitler's and Schoenfeld's institution received funding from the National Institutes of Health (NIH). Drs. Beitler, Thompson, Schoenfeld, and Brown received support for article research from the NIH. Dr. Thompson's institution received funding from the National Heart, Lung, and Blood Institute (NHLBI) and Department of Defense; reports consulting for Bayer, Boehringer Ingelheim, and GlaxoSmithKline; and authorship for UpToDate, all outside the submitted work. Drs. Talmor (1UM1HL108724) and Beitler (K23HL133489) received funding from NHLBI. The remaining authors have disclosed that they do not have any potential conflicts of interest.

For information regarding this article, E-mail: Samuel.Brown@imail.org

DOI: 10.1097/CCM.0000000000004104

Objectives: Survival from acute respiratory distress syndrome is improving, and outcomes beyond mortality may be important for testing new treatments. The "ventilator-free days" score, is an established composite that equates ventilation on day 28 to death. A hierarchical outcome treating death as a worse than prolonged ventilation would enhance face validity, but performance characteristics and reporting of such an outcome are unknown. We therefore evaluated the performance of a novel hierarchical composite endpoint, the Alive and Ventilator Free score.

Design: Using data from four Acute Respiratory Distress Syndrome Network clinical trials, we compared Alive and Ventilator Free to the ventilator-free days score. Alive and Ventilator Free compares each patient with every other patient in a win-lose-tie for each comparison. Duration of mechanical ventilation is only compared if both patients survived. We evaluated power of Alive and Ventilator Free versus ventilator-free days score under various circumstances.

Setting: ICUs within the Acute Respiratory Distress Syndrome Network.

Patients: Individuals enrolled in four Acute Respiratory Distress Syndrome Network trials.

Interventions: None for this analysis.

Measurements and Main Results: Within the four trials ($n = 2,410$ patients), Alive and Ventilator Free and ventilator-free days score had similar power, with Alive and Ventilator Free slightly more powerful when a mortality difference was present, and ventilator-free days score slightly more powerful with a difference in duration of mechanical ventilation. Alive and Ventilator Free less often found in favor of treatments that increased mortality and increased days free of ventilation among survivors.

Conclusions: A hierarchical composite endpoint, Alive and Ventilator Free, preserves statistical power while improving face validity. Alive and Ventilator Free is less prone to favor a treatment with discordant effects on survival and days free of ventilation. This general approach can support complex outcome hierarchies with multiple constituent outcomes. Approaches to interpretation of differences in Alive and Ventilator Free are also presented. (*Crit Care Med* 2020; 48:158–166)

Key Words: acute respiratory distress syndrome; composite outcomes; trial endpoints

Acute respiratory distress syndrome (ARDS) is an often fatal, highly morbid condition for which specific treatments are actively being sought (1). An important traditional outcome for clinical trials of therapies for ARDS has been hospital mortality at 28 or 60 days (2, 3), although some recent trials have used, for example, 90 days or longer (4). Unfortunately, using early mortality, very few interventions have proved beneficial, an observation that may reflect type II statistical error, that is, a "false negative" (3). Especially in early-phase trials, a mortality outcome requires larger than feasible sample sizes and may ignore important treatment effects on morbidity. Furthermore, assuming similar mortality, improvement in other outcomes may be adequate to endorse the effectiveness of a therapy, particularly when the nonmortality outcomes are patient-centered. Testing endpoints separately is highly inefficient, requiring extremely large sample sizes, which compounds the risk of rejecting efficacious therapies. By contrast, a composite outcome that incorporates mortality and important morbidity to model a "better" outcome for patients would overcome this limitation.

The most widely used composite outcome in the ARDS literature is commonly called "ventilator-free days" (VFDs). Described by Schoenfeld and Bernard (5), this composite endpoint combines mortality with number of days after successful liberation from mechanical ventilation among survivors, truncated at 28 days. Although often erroneously reported as such, the units are not actually days. Crucially, the VFD composite endpoint treats a patient dead on any day before day 28 as identical with a patient alive but dependent on the ventilator on day 28. Equating 28 days of mechanical ventilation with death does not reflect patient, family, clinician, or societal values and beliefs (6). Nor does the standard reporting of the outcome facilitate interpretation, since it is a merger of the probabilities of death or ventilation on day 28 with days free of ventilation among patients alive and free of the ventilator on day 28. To avoid common misinterpretation of the units of this composite endpoint, we refer to the "VFD score" throughout the remainder of the text.

An alternative approach using a topic-specific hierarchical composite outcome (7) has been employed in acquired immune deficiency syndrome (8) and cardiovascular clinical trials (9) to avoid the problem of making more and less severe constituent outcomes equivalent (10). We hypothesized that a similar approach (statistically equivalent to modifications of the parallel "worst-rank ordinal" approach [11, 12]) could be useful in ARDS trials and critical care more broadly (and indeed have implemented it in the Study of Mechanical Ventilation Directed by Transpulmonary Pressures (EP-Vent2) trial of esophageal-manometry-guided ventilator management [13, 14] with similar application in another trial [15]).

In this work, we therefore evaluate performance characteristics of a novel hierarchical composite endpoint for ARDS trials using data from four published multicenter trials by the National Heart, Lung, and Blood Institute (NHLBI) ARDS Network (ARDSnet). This hierarchical endpoint, the Alive and Ventilator Free (AVF) score, incorporates death and days after successful liberation from mechanical ventilation at 28 days in such a manner that death constitutes a worse outcome than prolonged mechanical ventilation. This technique can be modified or expanded to accommodate multiple hierarchically ranked outcomes to reflect outcomes considered more and less important by patients, family members, clinicians, researchers, and regulators. In this article, we consider three related questions: 1) a general approach to hierarchical composite endpoint construction, 2) a specific hierarchical composite endpoint, and 3) a useful and interpretable effect measure for reporting rank-based endpoints.

MATERIALS AND METHODS

Patients

We studied patients enrolled in four multicenter clinical trials who met American European Consensus Conference definitions of ARDS (then termed Acute Lung Injury and ARDS) (16). For simplicity, following the subsequent Berlin consensus definition (1), we refer to patients in these trials as having had ARDS.

Using the National Institutes of Health/NHLBI Biological Specimen and Data Repository Information Coordinating Center data repository, we accessed the deidentified datasets for the NHLBI ARDSnet And Respiratory Management in Acute Lung Injury and Acute Respiratory Distress Syndrome (ARMA) (17), Assessment of Low tidal Volume and Elevated end-expiratory volume to Obviate Lung Injury (ALVEOLI) (18), and Fluids and Catheter Treatment Trial (FACTT) (19, 20) trials. The four datasets (FACTT included two factorialized interventions) are described in **Table 1**. The ARMA trial was performed among 861 patients with ARDS and compared higher versus lower tidal volumes. The ALVEOLI trial was performed among 549 patients with ARDS who were randomly assigned to lower versus higher positive end-expiratory pressures (PEEPs). The FACTT trial (1,000 patients) was a 2×2 design comparing a liberal versus conservative fluid management strategy (FACTT-Fluid) and, in a factorial randomization, treatment guided by either a pulmonary artery catheter (PAC) or a central venous catheter (CVC) (FACTT-Catheter). ARDSnet methods for defining liberation from mechanical ventilation are presented in the **Online Data Supplement** (Supplemental Digital Content 1, http://links.lww.com/CCM/F63).

The study protocol was reviewed and approved by the institutional review board of Beth Israel Deaconess Medical Center in Boston, MA.

Construction of the Hierarchical Composite Endpoint, Alive and Ventilator Free

With minor modifications, we followed the method of Finkelstein and Schoenfeld (8) by which each patient is compared

TABLE 1. National Institutes of Health/National Heart, Lung, and Blood Institute Acute Respiratory Distress Syndrome Network Trials Studied

Trial	Treatment	Control	*n* (Treatment: Control)	Primary Endpoint
And Respiratory Management in Acute Lung Injury and Acute Respiratory Distress Syndrome (ARMA)	Low tidal volume ventilation	Traditional tidal volume ventilation	432: 429	Death before discharge home breathing without assistance, to 180 d; ventilator-free days to day 28
Assessment of Low tidal Volume and Elevated end-expiratory volume to Obviate Lung Injury (ALVEOLI)	High PEEP	Low PEEP	276: 273	Death before discharge home breathing without assistance, to 60 d
Fluids and Catheter Treatment Trial (FACTT)-Fluid[a]	Liberal fluid strategy	Conservative fluid strategy	494: 501	60-d mortality prior to discharge home
Fluids and Catheter Treatment Trial (FACTT)-Catheter[a]	Pulmonary artery catheter	Central venous catheter	486: 509	60-d mortality prior to discharge home

PEEP = positive end-expiratory pressure.

[a]These two trials were a 2 × 2 factorial design in the same patient population. Only 995 patients had evaluable ventilator-free day (VFD) scores within the Biological Specimen and Data Repository Information Coordinating Center datasets for these two trials; the reported numbers reflect those with evaluable VFD.

with every other patient in the trial. For each patient-to-patient comparison, a win, loss, or tie is defined in a hierarchical manner. The comparisons are first performed on the basis of the most important outcome (typically death for critical care trials), and only if neither patient has experienced that outcome will the win-lose-tie comparison be based on a less important outcome (typically a measure of morbidity), for example, duration of mechanical ventilation. This technique can accommodate multiple secondary outcomes, arranged hierarchically and tailored to the population and treatment of study.

With this framework, we developed the novel hierarchical outcome AVF. This outcome incorporates vital status and time since successful liberation from mechanical ventilation through day 28. Following standard practice, we defined time since successful liberation as 28–*n*, where *n* is the number of days between the first and last day of mechanical ventilation; patients who are still ventilator-dependent on day 28 are assigned a value of 0 days since successful liberation. To compute AVF, each subject is compared with every other subject in both trial arms and assigned a score (win = +1; lose = –1; tie = 0) for each pairwise comparison, based on which fared better (**Table 2**). If one subject survives and the other does not, scores of +1 and –1, respectively, are assigned for that pairwise comparison. If both subjects in the pairwise comparison survive, their scores are determined by time since successful liberation from mechanical ventilation: the subject with more time since successful liberation from the ventilator is assigned a score of +1, and the subject with less time since liberation is assigned a score of –1. If both subjects either die at any time during the 28-day period or have equal duration of mechanical ventilation, both are assigned a score of 0 for that pairwise comparison. Then, the points from all pairwise comparisons are summed to obtain a cumulative score for each subject. These cumulative scores are ranked and compared between treatment versus control groups using the Mann-Whitney *U* and Wilcoxon rank-sum tests. (For efficiency of calculation, identical statistical comparisons can be obtained using an ordinal outcome in which worse outcomes are ranked lower than better

TABLE 2. Method for Calculating the Hierarchical Endpoint

Index Subject Died	Comparison Subject Died	Days Free of Ventilator for Index Subject vs Comparison Subject	Points for Index Subject	Points for Comparison Subject
Yes	Yes	NA	0 (tie)	0 (tie)
No	Yes	NA	+1 (win)	−1 (lose)
Yes	No	NA	−1 (lose)	+1 (win)
No	No	More	+1 (win)	−1 (lose)
No	No	Less	−1 (lose)	+1 (win)
No	No	Same	0 (tie)	0 (tie)

NA = not applicable.

The points are summed up to obtain a cumulative score for each subject. Every patient is compared with every other patient in both study arms. The scores are compared between study arms by a Mann-Whitney *U* test.

outcomes.) Fundamentally, the proposed comparisons seek to answer the clinically relevant question: with which treatment strategy would a patient be likeliest to have a better outcome?

Reporting of the Hierarchical Composite Endpoint AVF

Best practice for reporting any composite endpoint always should include separate reporting of each constituent endpoint and, ideally, a measure of the difference between groups for the composite endpoint itself. For AVF, we therefore recommend reporting four aspects of the endpoint:

1) the main effect estimate, the probability of a superior outcome (θ) with 95% CI
2) p value obtained via the Mann-Whitney U and Wilcoxon rank-sum tests
3) mortality, by treatment group
4) time since successful liberation from mechanical ventilation through day 28 among survivors only, by treatment group

The probability of superior outcome (θ), also known as the "probabilistic index" or "common language effect size statistic," is defined as the estimated probability that an individual randomly selected from the study population will have a superior outcome if assigned to a given treatment arm. Details of its calculation are in the Online Data Supplement (Supplemental Digital Content 1, http://links.lww.com/CCM/F63).

Traditional Composite Outcome: The VFD Score

In its standard form (5), the VFD score incorporates 28-day mortality and number of days after successful liberation from mechanical ventilation through day 28 into a single composite score. Two types of patients are assigned 0 VFDs under this schema: patients who die on or before day 28, and patients who are alive and mechanically ventilated at day 28. Similar to the ventilator-free component of AVF, survivors who are no longer ventilated on day 28 are assigned VFD equal to 28–n, where n is the number of days between the first and last day of mechanical ventilation. We describe VFD score calculations for each trial in the Online Data Supplement (Supplemental Digital Content 1, http://links.lww.com/CCM/F63).

Simulations of Effect Sizes, Sample Sizes, and Statistical Power

We used simulations to assess the relative statistical power of the hierarchical AVF score versus the VFD score. Simulation parameters were obtained from each of the four clinical trial datasets. Frequencies of the resulting significance levels based on the Mann-Whitney U and Wilcoxon rank-sum statistic were plotted against the range of scenarios for each trial. Power calculations were performed by simulating 5,000 independent trials for each specification of parameters. The following parameters were specified: 28-day mortality rates, proportion of patients alive and mechanically ventilated at day 28, and distribution of days of mechanical ventilation among patients alive and not mechanically ventilated at day 28. Mortality rates and proportions of patients alive and mechanically ventilated at day 28 were simulated by a random binary function, whereas the distribution of days of mechanical ventilation among patients AVF at day 28 was simulated by a truncated normal distribution, which empirically fit well the distribution of observed values.

To further investigate the performance of AVF in the presence of varying effect estimates for the intervention, we simulated multiple treatment arm scenarios, compared to simulations based on the parameters estimated from the ARMA control group. These simulations (5,000 trials per simulation for all analyses) evaluated power at a sample size of 1,000 patients over a range of deviations (mortality improvement from 2% to 10% and days free of ventilation among survivors of 3–8 d) from the ARMA control group. In related simulations, we held the mortality rate constant and varied the days free of ventilation among survivors to evaluate power to detect differences in days free of ventilation for the two composite endpoints. Similarly, in other simulations, we held days free of ventilation constant and evaluated the association between differences in mortality rates and power for the two endpoints. In additional sensitivity analyses, again with 5,000 simulated trials of 1,000 patients each), we held the proportion of patients either dead or ventilated at day 28 constant, while decreasing the mortality rate, when compared with parameters estimated from the ARMA control group.

To explore tolerance to discordant effects on mortality and nonmortality endpoints within AVF and the VFD score, we performed additional simulations evaluating scenarios in which an increase in mortality was associated with shorter duration of mechanical ventilation among survivors. For this analysis, we used the ARMA control group estimates and compared simulated treatment groups with discordant outcomes across a range of differences in mortality and days free of ventilation. We held the proportion of patients alive and ventilator-dependent at 28 days constant.

All analyses were performed in the R Statistical Package 3.5.2 (R Core Team, Vienna, Austria) and in SAS version 9.3 (SAS Institute, Cary NC).

RESULTS

ARDSnet Trial Results

Mechanical ventilation with lower tidal volume (ARMA) decreased 28-day mortality as compared with ventilation with traditional tidal volumes (25% vs 35%, respectively) and increased the VFD score (median 13 vs 4, respectively) (**Table 3**). Mechanical ventilation with lower or higher PEEP levels (ALVEOLI) did not significantly affect mortality rates or the VFD score. Conservative versus liberal fluid management (FACTT-Fluid) did not affect mortality but increased the VFD score (median 18 vs 14, respectively). Management with a PAC versus CVC (FACTT-Catheter) did not significantly affect mortality or the VFD score.

The AVF hierarchical endpoint differed between treatment groups in ARMA (probability of superior outcome with lower tidal volume: 56.5%; 95% CI, 52.7–60.3%; $p = 0.003$) and FACTT-Fluid (probability of superior outcome with conservative fluid management: 57.5%; 95% CI, 53.9–60.9%; $p < 0.001$) (**Table 3**). The AVF score did not differ between treatment groups in the ALVEOLI and FACTT-Catheter studies.

TABLE 3. Results of the Alive and Ventilator Free As Compared With the Ventilator-Free Day Score Within the Acute Respiratory Distress Syndrome Network Trials

Measure	Treatment Group	Control Group	*p*
And Respiratory Management in Acute Lung Injury and Acute Respiratory Distress Syndrome	Low tidal volume	Traditional tidal volume	
Mortality, %[a]	25.2	35.2	0.002
Days free of mechanical ventilation among survivors, median (IQR)	20 (9–24)	20 (5–24)	0.46
VFD score, median (IQR)	13 (0–23)	4 (0–22)	0.003
VFD score, probability of superior outcome, % (95% CI)	56.2 (52.4–60.0)	43.8 (40.0–47.6)	0.003
AVF, probability of superior outcome, % (95% CI)	56.5 (52.7–60.3)	43.5 (39.7–47.3)	0.003
Assessment of Low tidal Volume and Elevated end-expiratory volume to Obviate Lung Injury	High PEEP	Low PEEP	
Mortality, % (*n*)	23.2	22.3	0.84
Days free of mechanical ventilation among survivors, median (IQR)	20 (11.5–23.5)	20 (13–24)	0.47
VFD score, median (IQR)	17 (0–23)	17 (0–23)	0.42
VFD score, probability of superior outcome, % (95% CI)	49.2 (44.4–54.0)	50.8 (46.0–55.6)	0.42
AVF, probability of superior outcome (95% CI)	48.4 (43.6–53.2)	51.6 (46.8–56.4)	0.51
FACTT-Fluid	Liberal fluid strategy	Conservative fluid strategy	
Mortality, % (*n*)	24.9	21.5	0.20
Days free of mechanical ventilation among survivors, median (IQR)	18 (9–22)	21 (15–24)	< 0.001
VFD score, median (IQR)	14 (0–21)	18 (0–23)	< 0.001
VFD score, probability of superior outcome, % (95% CI)	41.2 (37.8–44.8)	58.8 (55.2–62.2)	< 0.001
AVF, probability of superior outcome (95% CI)	42.5 (39.1–46.1)	57.5 (53.9–60.9)	< 0.001
FACTT-Catheter	Pulmonary artery catheter	Central venous catheter	
Mortality, % (*n*)	23.0	23.4	0.88
Days free of mechanical ventilation among survivors, median (IQR)	19 (12–23)	19 (12.5–24)	0.11
VFD score, median (IQR)	16 (0–22)	16 (0–23)	0.32
VFD score, probability of superior outcome, % (95% CI)	46.8 (43.3–50.4)	53.2 (49.6–56.7)	0.25
AVF, probability of superior outcome (95% CI)	48.1 (44.5–51.6)	51.9 (48.4–55.5)	0.29

AVF = Alive and Ventilator Free, FACTT = Fluids and Catheter Treatment Trial, IQR = interquartile range, PEEP = positive end-expiratory pressure, VFD = ventilator-free day.

[a]The mortality outcome for the And Respiratory Management in Acute Lung Injury and Acute Respiratory Distress Syndrome trial was hospital mortality, as opposed to the 28-d mortality presented here.

Note that the hierarchical endpoint should be presented with the distributions of the constituent outcomes.

Power Estimates As a Function of Sample Size and Effect Sizes

Figure 1 (data tabulated in **eTable 1**, Supplemental Digital Content 1, http://links.lww.com/CCM/F63) displays a plot of overall statistical power simulated using different sample sizes for the comparison of the AVF and VFD scores. Mortality rates, proportion of patients alive and mechanically ventilated at day 28, and distribution of days free of mechanical ventilation among survivors not ventilated on day 28 were obtained from the respective ARDSnet trials.

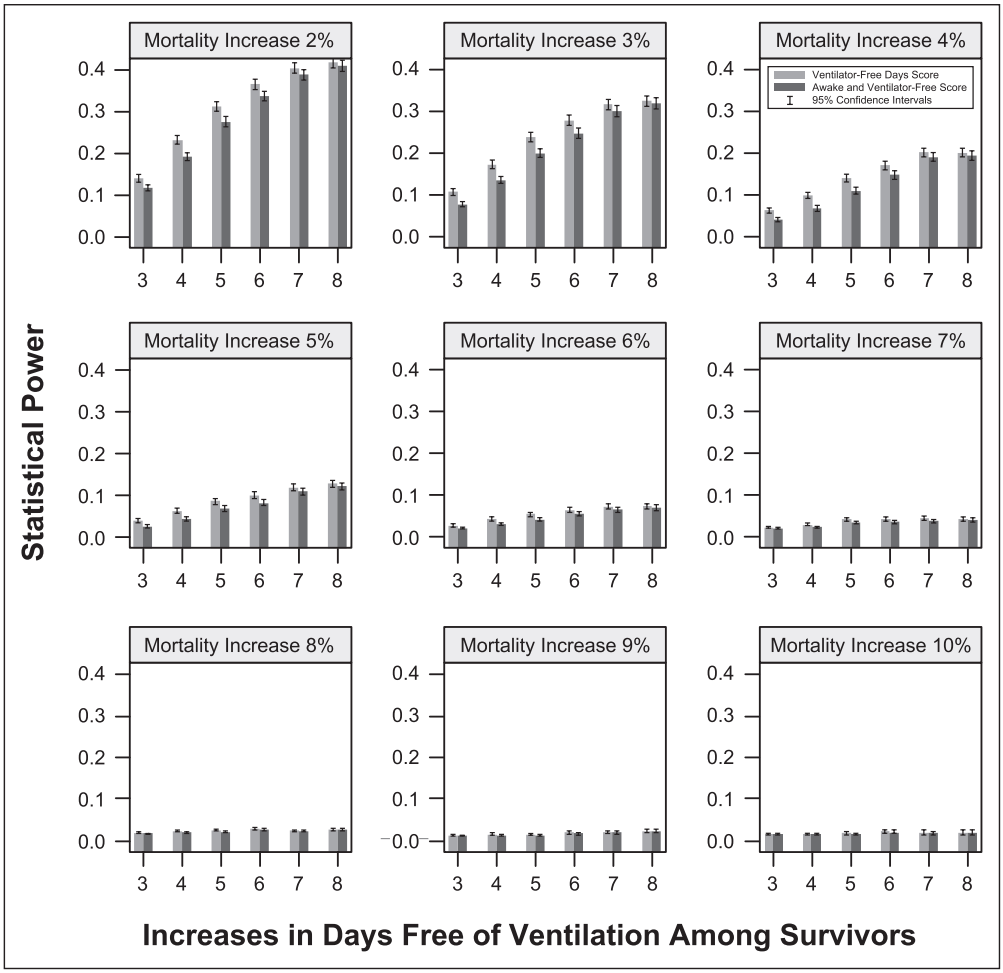

Figure 1. Statistical power for the ventilator-free day score versus Alive and Ventilator Free score, by total sample size, with 5,000 simulated trials per data point. Each pane displays results generated from simulations based on the designated Acute Respiratory Distress Syndrome Network trial. ALVEOLI = Assessment of Low tidal Volume and Elevated end-expiratory volume to Obviate Lung Injury, ARMA = And Respiratory Management in Acute Lung Injury and Acute Respiratory Distress Syndrome, FACTT = Fluids and Catheter Treatment Trial.

As expected, both the AVF and VFD scores had low power (5–14%) for simulations up to a sample size of 1,000 patients based on parameters estimated from the ALVEOLI and FACTT-Catheter trials. In simulations based on FACTT-Fluid (where the efficacy was in the distribution of days free of ventilation among survivors), the VFD score had similar power to AVF (e.g., 88% [87–89%] vs 87% [86–88%] with 600 total patients). By contrast, in simulations based on ARMA (where the efficacy was in mortality), the AVF score had slightly higher power than VFD score (e.g., 83% [82–84%] vs 80% [79–81%] with 900 total patients).

Results of other simulations are reported in the Online Data Supplement (Supplemental Digital Content 1, http://links.lww.com/CCM/F63); exemplary results are displayed in **Figure 2**. In general, the AVF score had similar power to the VFD score and was less prone to find in favor of a treatment that increased both mortality and days free of mechanical ventilation.

DISCUSSION

Mortality rates in the ARDSnet trials declined substantially over the course of the network's existence (21). Control group in-hospital mortality in ARMA (enrolling 1996–1999) was 40%, while control group in-hospital mortality to 28 days was 22% in Early Versus Delayed Enteral Feeding (EDEN) (enrolling 2008–2011) (22). This decrease in mortality in randomized controlled trials has important implications for future trials. To continue trials in broad populations of patients with ARDS may mean that design of such trials based on a mortality outcome will require ever-increasing sample sizes, which limits feasibility, increases cost, and risks delaying evaluation of promising therapies. One alternative approach would be to restrict enrollment to the most severely ill patients, as was done in The Oscillation for Acute Respiratory Distress Syndrome (ARDS) Treated Early Trial (OSCILLATE) (23) and ARDS et Curarisation Systematique (ACURASYS) (24), with associated control group hospital mortality of 35% and 41%, respectively. An alternative approach would be to test potential interventions in studies powered to include clinically and mechanistically relevant nonmortality endpoints.

Composite outcomes have become standard in cardiovascular trials and see some use (generally as the VFD score) in critical care trials. These composite outcomes improve power and efficiency of trials and allow incorporation of relevant nonmortality outcomes that are likely affected by candidate treatments. However, despite their widespread use, they have occasioned caution and criticism (25–30).

The NHLBI working group on future research directions in ARDS recently summarized the considerations for possible endpoints for clinical trials in ARDS (2). They concluded that there are no proven surrogate markers for intermediate or long-term mortality in ARDS. Furthermore, patient-important outcomes beyond survival such as prolonged organ support therapy, physical, cognitive, or vocational recovery may be complicated by variable and difficult-to-measure patient baseline impairments, differences in ICU and end-of-life decision making, and differential follow-up rates. Despite this uncertainty, the VFD score has been commonly applied as a composite outcome of mortality and duration of respiratory failure. The VFD score is problematic, though. Most importantly, the

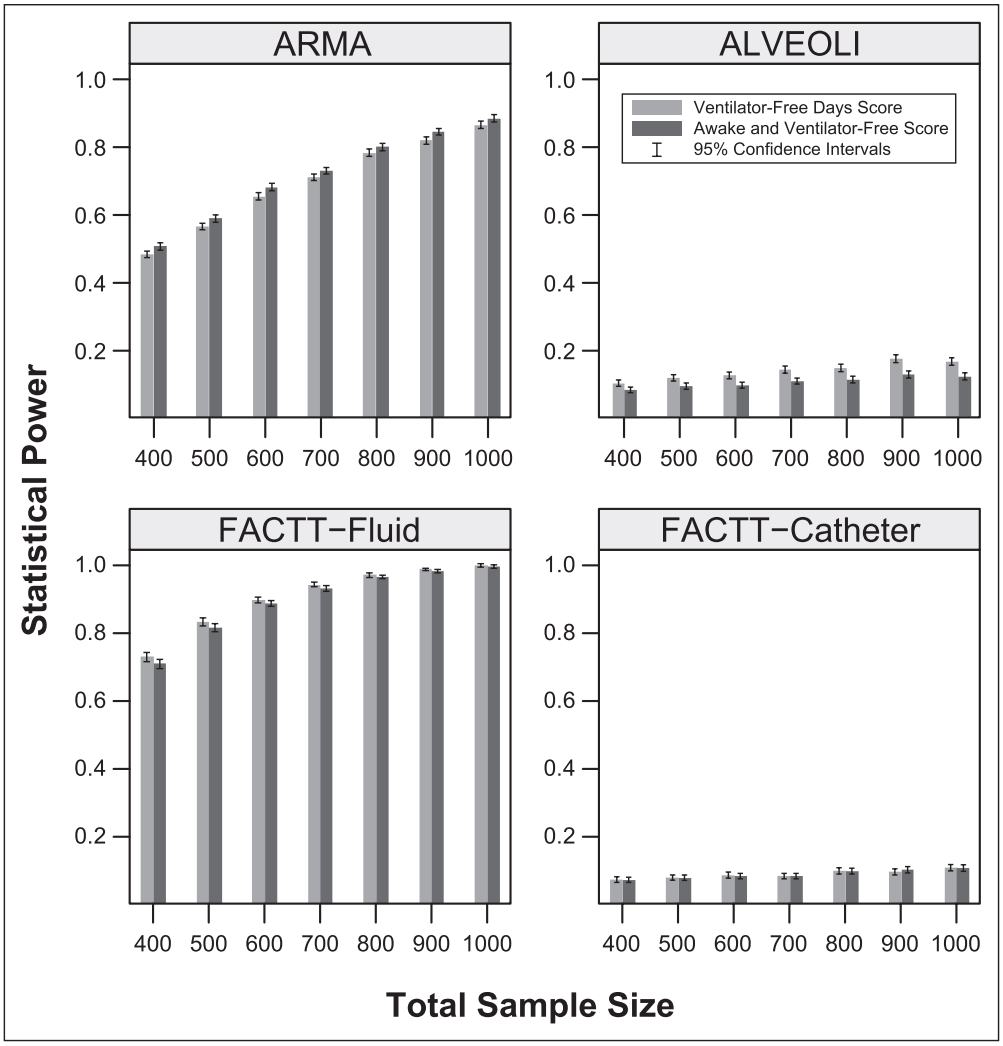

Figure 2. Power from 5,000 simulated trials for the hierarchical Alive and Ventilator Free endpoint versus the ventilator-free day score (500 patients in each arm) as a function of decreasing mortality rates and increases in days free of mechanical ventilation among survivors in the treatment group. Estimates for the control group are drawn from the And Respiratory Management in Acute Lung Injury and Acute Respiratory Distress Syndrome (ARMA) trial.

VFD score treats death and ventilation on day 28 as equivalent, a claim with limited face validity because most patients do not consider prolonged ventilation identical to death, even where they would prefer shorter ventilation to longer ventilation (31, 32). Similarly, the VFD score would not distinguish a treatment that increased survival while increasing by the same amount patients ventilated on day 28 (e.g., by saving very sick patients who would otherwise have died). Hierarchical outcomes address certain failings of simpler composite outcomes through improved face validity and interpretability (7–11, 33, 34). Hierarchical composite outcomes have better face validity because they explicitly rate death as more important than nonmortality outcomes; while the pairwise approach is more complex to calculate, it has in its favor the intuitive interpretation that patients are compared with each other to determine on balance which treatment arm is better. Such composite outcomes can accommodate multiple, hierarchically ranked outcomes into a single summary. Stakeholder groups could thus together establish outcome hierarchies, which could be implemented precisely within a hierarchical composite endpoint.

In this simulation-based study, the hierarchical composite outcome AVF score has similar power to the VFD score with better face validity. In addition, the AVF score has higher power to detect differences in mortality across a range of plausible increases in days free of ventilation. This basic attribute is manifest in the differences in power between ARMA and FACTT-Fluid: in ARMA, the AVF score has slightly more power because the difference in mortality was greater, while in FACTT-Fluid, the VFD score had slightly more power because mortality was similar but days free of ventilation was greater.

The AVF score may also have a more clinically intuitive interpretation than the VFD score, which as a trial summary is largely opaque as a merger of probabilities and distributions. The effect estimate for the AVF score is the probability of superior outcome with receipt of the studied intervention. True to its literal meaning, the probability of superior outcome is defined as the probability that a patient randomly selected from the study population would do better if assigned to a given study arm. An alternative, previously described metric for reporting treatment effect with hierarchical outcomes is the win/lose ratio (10). However, the win/lose ratio may be less easily interpreted if widespread misinterpretation of odds ratios is any indicator, whereas clinicians and the lay public naturally think in terms of probabilities (35, 36). Our reporting approach preserves face validity and robust statistical power while also prioritizing ease of interpretation (the probability of superior outcome), a crucial design feature of any clinical trial endpoint. As with any composite outcome, we recommend also reporting individual constituent endpoints. Although all composite outcomes represent compromises among competing priorities, the hierarchical AVF endpoint appears superior to the traditional VFD score.

We acknowledge that improvements in power with AVF are generally small. An increase in power was not our primary motivation, and we are reassured that improvements in power are most

marked in situations where, for example, an intervention increases mortality but decreases duration of ventilation among survivors or where mortality decreases but the proportion ventilated on day 28 increases by the same amount). We acknowledge that this endpoint has only been carefully evaluated in these four ARDSnet trials, although it has been used in other trials as a primary endpoint (13–15). We acknowledge that worst-rank ordinal endpoints have the same statistical characteristics as the AVF score and are as easy to understand when there is only one nonmortality outcome. We acknowledge that the VFD score could also be presented in terms of probability of superiority, although this would not solve its face validity problem. We acknowledge that interpretability of endpoints is always complex, there is no established minimum clinically important difference for AVF, and while a hierarchical composite is an improvement, it does not solve all problems. We also acknowledge that we did not formally engage patient collaborators for this specific project. We believe that this framework provides an infrastructure for building patient-centered composite outcomes and strongly recommend patient collaboration for the development of new outcomes within this proposed hierarchical framework.

In summary, we present a hierarchical composite endpoint for clinical trials in ARDS. This endpoint enhances face validity and ease of clinical interpretation. AVF can facilitate more efficient performance of ARDS clinical trials of without appreciable loss of power and may yield higher power as compared with the nonhierarchical composite outcome, the VFD score. A similar hierarchical endpoint, focused on mortality and the duration of nonpulmonary organ dysfunction, may similarly be relevant to clinical trials in other areas of critical care medicine.

REFERENCES

1. Ranieri VM, Rubenfeld GD, Thompson BT, et al; ARDS Definition Task Force: Acute respiratory distress syndrome: The Berlin definition. *JAMA* 2012; 307:2526–2533
2. Lieu TA, Au D, Krishnan JA, et al; Comparative Effectiveness Research in Lung Diseases Workshop Panel: Comparative effectiveness research in lung diseases and sleep disorders: Recommendations from the National Heart, Lung, and Blood Institute workshop. *Am J Respir Crit Care Med* 2011; 184:848–856
3. Spragg RG, Bernard GR, Checkley W, et al: Beyond mortality: Future clinical research in acute lung injury. *Am J Respir Crit Care Med* 2010; 181:1121–1127
4. Moss M, Huang DT, Brower RG, et al; PETAL Network Investigators: Early neuromuscular blockade in the acute respiratory distress syndrome. *N Engl J Med* 2019; 380:1997–2008
5. 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
6. Mendelsohn AB, Belle SH, Fischhoff B, et al; QOL-MV Study Investigators: How patients feel about prolonged mechanical ventilation 1 year later. *Crit Care Med* 2002; 30:1439–1445
7. O'Brien PC: Procedures for comparing samples with multiple endpoints. *Biometrics* 1984; 40:1079–1087
8. Finkelstein DM, Schoenfeld DA: Combining mortality and longitudinal measures in clinical trials. *Stat Med* 1999; 18:1341–1354
9. Moyé LA, Davis BR, Hawkins CM: Analysis of a clinical trial involving a combined mortality and adherence dependent interval censored endpoint. *Stat Med* 1992; 11:1705–1717
10. Pocock SJ, Ariti CA, Collier TJ, et al: The win ratio: A new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur Heart J* 2012; 33:176–182
11. Lachin JM: Worst-rank score analysis with informatively missing observations in clinical trials. *Control Clin Trials* 1999; 20:408–422
12. Colantuoni E, Scharfstein DO, Wang C, et al: Statistical methods to compare functional outcomes in randomized controlled trials with high mortality. *BMJ* 2018; 360:j5748
13. Fish E, Novack V, Banner-Goodspeed VM, et al: The Esophageal Pressure-Guided Ventilation 2 (EPVent2) trial protocol: A multicentre, randomised clinical trial of mechanical ventilation guided by transpulmonary pressure. *BMJ Open* 2014; 4:e006356
14. Beitler JR, Sarge T, Banner-Goodspeed VM, et al: Effect of titrating positive end-expiratory pressure (PEEP) with an esophageal pressure-guided strategy vs an empirical high PEEP-Fio2 strategy on death and days free from mechanical ventilation among patients with acute respiratory distress syndrome: A randomized clinical trial. *JAMA* 2019; 321:846–857
15. Bellingan G, Brealey D, Mancebo J, et al: Comparison of the efficacy and safety of FP-1201-lyo (intravenously administered recombinant human interferon beta-1a) and placebo in the treatment of patients with moderate or severe acute respiratory distress syndrome: Study protocol for a randomized controlled trial. *Trials* 2017; 18:536
16. Bernard GR, Artigas A, Brigham KL, 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
17. Brower RG, Matthay MA, Morris A, et al; The 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
18. Brower RG, Lanken PN, MacIntyre N, 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
19. Wheeler AP, Bernard GR, Thompson BT, et al; National Heart Lung and Blood Institute Acute Respiratory Distress Syndrome Clinical Trials Network: Pulmonary-artery versus central venous catheter to guide treatment of acute lung injury. *N Engl J Med* 2006; 354:2213–2224
20. Wiedemann HP, Wheeler AP, Bernard GR, 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
21. Erickson SE, Martin GS, Davis JL, et al; NIH NHLBI ARDS Network: Recent trends in acute lung injury mortality: 1996-2005. *Crit Care Med* 2009; 37:1574–1579
22. Rice TW, Wheeler AP, Thompson BT, et al: Initial trophic vs full enteral feeding in patients with acute lung injury: The EDEN randomized trial. *JAMA* 2012; 307:795–803
23. Ferguson ND, Cook DJ, Guyatt GH, et al; OSCILLATE Trial Investigators; Canadian Critical Care Trials Group: High-frequency oscillation in early acute respiratory distress syndrome. *N Engl J Med* 2013; 368:795–805
24. Papazian L, Forel JM, Gacouin A, et al; ACURASYS Study Investigators: Neuromuscular blockers in early acute respiratory distress syndrome. *N Engl J Med* 2010; 363:1107–1116
25. Freemantle N, Calvert M, Wood J, et al: Composite outcomes in randomized trials: Greater precision but with greater uncertainty? *JAMA* 2003; 289:2554–2559
26. Gent M: Some issues in the construction and use of clusters of outcome events. *Contemp Clin Trials* 1997; 18:546–549
27. Cordoba G, Schwartz L, Woloshin S, et al: Definition, reporting, and interpretation of composite outcomes in clinical trials: Systematic review. *BMJ* 2010; 341:c3920
28. Ferreira-González I, Busse JW, Heels-Ansdell D, et al: Problems with use of composite end points in cardiovascular trials: Systematic review of randomised controlled trials. *BMJ* 2007; 334:786
29. Ferreira-González I, Permanyer-Miralda G, Busse JW, et al: Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns. *J Clin Epidemiol* 2007; 60:651–657; discussion 658–662
30. Montori VM, Permanyer-Miralda G, Ferreira-González I, et al: Validity of composite end points in clinical trials. *BMJ* 2005; 330:594–596

31. Fried TR, Bradley EH, Towle VR, et al: Understanding the treatment preferences of seriously ill patients. *N Engl J Med* 2002; 346:1061–1066
32. Guentner K, Hoffman LA, Happ MB, et al: Preferences for mechanical ventilation among survivors of prolonged mechanical ventilation and tracheostomy. *Am J Crit Care* 2006; 15:65–77
33. Felker GM, Anstrom KJ, Rogers JG: A global ranking approach to end points in trials of mechanical circulatory support devices. *J Card Fail* 2008; 14:368–372
34. Ediebah DE, Galindo-Garre F, Uitdehaag BM, et al: Joint modeling of longitudinal health-related quality of life data and survival. *Qual Life Res* 2015; 24:795–804
35. Schwartz LM, Woloshin S, Welch HG: Misunderstandings about the effects of race and sex on physicians' referrals for cardiac catheterization. *N Engl J Med* 1999; 341:279–283; discussion 286–287
36. Persoskie A, Ferrer RA: A most odd ratio: Interpreting and describing odds ratios. *Am J Prev Med* 2017; 52:224–228