RE-APPRAISAL OF VENTILATOR-FREE DAYS IN CRITICAL CARE RESEARCH

Nadir Yehya, MD; Michael O. Harhay, PhD; Martha A.Q. Curley, RN, PhD, FAAN; David A. Schoenfeld, PhD; Ron W. Reeder, PhD

SUPPLEMENTARY METHODS AND DATA

ADDITIONAL METHODS

Competing Risk Regression

An alternative approach to analyzing VFDs is to employ Fine and Gray competing risk regression (1). Traditional survival analysis (log-rank tests, Cox proportional hazard) considers a single endpoint (2), with subjects either experiencing the event of interest, not experiencing the event, or not being available for assessment (censoring). In this framework, if the outcome of interest was intact extubation, death is censored. Thus, the test determines whether the risk of extubation among subjects that remain alive differs between treatments. Censoring due to death is inappropriate here for two reasons. First, it removes the importance of mortality differences between the groups so that a treatment group may have higher mortality but be deemed better because survivors were extubated earlier. Second, the processes contributing to duration of ventilation are not independent of the processes contributing to mortality, thereby violating a critical assumption needed for valid inference, namely, that censoring must be non-informative (censoring should be unrelated to the outcome)(3).

Competing risk addresses the situation when more than one mutually exclusive endpoint is possible. A new hazard function, the subdistribution hazard, is defined, which handles the at-risk population by *retaining* (not censoring) subjects experiencing the competing event (death) in the risk set. This is analogous to setting VFD = 0 for non-survivors. In effect, subjects experiencing the competing event (death) act as placeholders for subjects that can never experience the event of interest (extubation) by remaining in the risk set, thus constraining this hazard function. This results in a subdistribution hazard ratio (SHR), the magnitude of which is affected by both the time to extubation among survivors and the probability of death. By censoring the analysis at 28 days, SHR imparts information analogous to VFDs at 28 days. SHR assesses the association between an intervention and extubation *accounting for the existence* of

the alternative outcome of death. SHR measures effect size, and the regression readily accommodates confounders for multivariable analysis.

As with Cox models, Fine and Gray is based on proportional hazards (1, 2). The alternative Gray's test is a non-parametric test that does not rely on the proportional hazards assumption; however, it does not offer an effect size or the ability to adjust for confounders (4), analogous to the log-rank test.

VFDs thus can be considered time-to-event analyses right-censored at 28 days, with the event of interest as extubation, and mortality a competing event. Extrapolating VFD = 0 for non-survivors to a time-to-event analysis, non-survivors are assigned > 28 days of ventilation and are thus "never free of ventilation." A Cox regression setting non-survivors to > 28 ventilator days and censoring at 28 days will provide a hazard ratio for extubation *identical* to a competing risk SHR treating mortality as the competing risk (4, 5). Therefore, any software which can perform a log-rank test or Cox regression can perform competing risk regression by setting all non-survivors past the date of censoring, typically > 28 days.

In the main manuscript, we provide further details on how to define VFDs. Specifically, we suggest assigning all non-survivors who died within 28 days 0 VFDs, irrespective of intubation status at time of death. In the competing risk scenario, this would most commonly result in subjects being assigned death on the day they died, likely while still intubated. However, there may be situations in which a subject is extubated on day 10 and dies on day 20, a scenario in which we suggest assigning 0 VFDs in order to appropriately penalize non-survival (main manuscript Figure 1). This seemingly violates the competing risk framework, in which the two outcomes are exclusive (death *or* successful extubation). We reconcile this by clearly defining "successful extubation" as requiring survival to 28 days.

Simulations

Simulations were performed in SAS 9.4 (SAS Institute, Cary, North Carolina). The relative power of different tests in 3000 simulations of a two-arm trial with n = 300 per arm is provided in Table 2 (censored at 28 days), Supplementary Table 1 (60 days), and Supplementary Table 2 (90 days). All tests are performed for a two-sided alternative hypothesis with $\alpha = 0.05$. Mortality is simulated according to a Bernoulli distribution, and ventilation duration according to an exponential distribution. Fewer than 3% of survivors were ventilated > 28 days in the simulated data. We varied whether the effect was driven by mortality, ventilator duration, both, or in opposite directions.

For comparison, Fisher's exact test is provided to test differences in mortality rate without accounting for the duration of ventilation. We also included a log-rank test, setting non-survivors to have a duration of ventilation longer than any duration of ventilation among survivors, to demonstrate the similarity of this construct to Gray's test and Fine and Gray competing risk regression.

Supplementary Table E1: Power calculations for different statistical tests in which the primary outcome of interest is ventilator duration censored at 60 days.

			Power[1]					
Effects	**Mortality[2]**	**Mean ventilator days among survivors[3]**	**Fine and Gray**	**Gray**	**Log-rank[4]**	**Rank-sum[5]**	**T-test**	**Fisher exact[6]**
Mortality only								
Treatment	15%	7	79%	78%	79%	56%	84%	**85%**
Control	25%	7						
Strong mortality and weak duration			93%	**94%**	93%	89%	93%	85%
Treatment	15%	6						
Control	25%	7						
Moderate mortality and duration			76%	78%	76%	**84%**	62%	33%
Treatment	15%	5.0						
Control	20%	6.5						
Weak mortality and strong duration			77%	81%	77%	**96%**	46%	5%
Treatment	15%	5						
Control	16%	8						
Duration only								
Treatment	15%	5	68%	73%	69%	**94%**	34%	4%
Control	15%	8						
Conflicting								
Treatment	15%	6.5	5%	5%	5%	13%	14%	**33%**
Control	20%	5						

The highest power for any scenario is **bolded**.

[1] Results are each based on 3000 simulated trials with 300 subjects in each of two treatment groups, a two-sided alternative hypothesis, and a type I error rate of $\alpha = 0.05$.

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

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

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

[5] Due to computational limits, the normal approximation with continuity correction was used for the Wilcoxon rank-sum test.

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

Supplementary Table E2: Power calculations for different statistical tests in which the primary outcome of interest is ventilator duration censored at 90 days.

			Power[1]					
Effects	**Mortality[2]**	**Mean ventilator days among survivors[3]**	**Fine and Gray**	**Gray**	**Log-rank[4]**	**Rank-sum[5]**	**T-test**	**Fisher exact[6]**
Mortality only								
Treatment	15%	7	79%	78%	79%	56%	**86%**	85%
Control	25%	7						
Strong mortality and weak duration			93%	**94%**	93%	89%	91%	85%
Treatment	15%	6						
Control	25%	7						
Moderate mortality and duration			76%	78%	76%	**84%**	54%	33%
Treatment	15%	5.0						
Control	20%	6.5						
Weak mortality and strong duration			77%	81%	77%	**96%**	27%	5%
Treatment	15%	5						
Control	16%	8						
Duration only								
Treatment	15%	5	68%	73%	69%	**94%**	17%	4%
Control	15%	8						
Conflicting								
Treatment	15%	6.5	5%	5%	5%	13%	20%	**33%**
Control	20%	5						

The highest power for any scenario is **bolded**.

[1] Results are each based on 3000 simulated trials with 300 subjects in each of two treatment groups, a two-sided alternative hypothesis, and a type I error rate of $\alpha = 0.05$.

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

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

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

[5] Due to computational limits, the normal approximation with continuity correction was used for the Wilcoxon rank-sum test.

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

REFERENCES

1. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association* 1999; 94: 496-509.
2. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *American journal of epidemiology* 2009; 170: 244-256.
3. 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. *Medical care* 2010; 48: S96-105.
4. Gray R. A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk. *Ann Stat* 1988; 16: 1141-1154.
5. Bateman ST, Borasino S, Asaro LA, Cheifetz IM, Diane S, Wypij D, Curley MA, Investigators RS. Early High-Frequency Oscillatory Ventilation in Pediatric Acute Respiratory Failure. A Propensity Score Analysis. *Am J Respir Crit Care Med* 2016; 193: 495-503.