Statistical methods to compare functional outcomes in randomized controlled trials with high mortality

Elizabeth Colantuoni,[1,2] Daniel O Scharfstein,[1,2] Chenguang Wang,[3] Mohamed D Hashem,[1,4] Andrew Leroux,[2] Dale M Needham,[1,4,5] Timothy D Girard[6]

[1]Outcomes After Critical Illness and Surgery (OACIS) Group, Johns Hopkins University, Baltimore, MD, USA

[2]Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

[3]Division of Biostatistics and Bioinformatics, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[4]Division of Pulmonary and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[5]Department of Physical Medicine and Rehabilitation, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[6]Clinical Research, Investigation, and Systems Modeling of Acute illness (CRISMA) Center in the Department of Critical Care Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

Correspondence to:
E Colantuoni
ejohnso2@jhmi.edu

Additional material is published online only. To view please visit the journal online.

Mortality is a common primary endpoint in randomized controlled trials of patients with a high severity of illness, such as critically ill patients. However, researchers are increasingly evaluating functional outcomes, such as quality of life. Importantly, in such trials some patients may die before the assessment of a functional outcome, resulting in the functional outcome being "truncated due to death." As described in this paper, defining and testing treatment effects on functional outcomes in this setting requires careful consideration. Data from a completed trial of critically ill patients are used to highlight key differences among three statistical approaches used when analyzing such trials.

Introduction

It is critical to align the interests of patients, clinicians, and regulators when designing, analyzing, and interpreting the results of randomized controlled trials.[1] This is especially challenging in randomized controlled trials evaluating treatments for seriously ill patients, such as those with critical illness receiving care on an intensive care unit (ICU). In such trials mortality is a common primary outcome measure. However, given the high value that patients place on outcomes other than mortality, including cognition, physical function, and quality of life, such functional outcomes are increasingly being evaluated as coprimary or key secondary outcomes in such trials.[2][3] When evaluating the effects of treatments on primary and key secondary outcomes, regulators expect that key principles (eg, intention to treat) be adhered to and statistical procedures be defined a priori, whereas clinicians and patients expect that the defined statistical procedures yield meaningful information about the efficacy of the treatment. Defining the statistical procedure to evaluate the effect of randomized treatments on functional outcomes when patients die before such outcomes can be assessed is complex since the functional outcomes for patients who die do not exist—that is, these outcomes are "truncated due to death." [4-6]

We focus on three statistical approaches used to evaluate the effect of treatment on functional outcomes "truncated due to death": survivors analysis, survivor average causal effect (SACE),[5][6] and composite endpoint approach. We review each approach and use an example trial to show several key considerations in selecting an approach: the required assumptions; conditions under which an approach would satisfy established principles, such as intention to treat; and when approaches may yield different results. In addition, we discuss challenges in reconciling differences in the principles that govern the conduct of randomized controlled trials and clinician and patient perspectives when functional outcomes are "truncated due to death." The supplementary appendix provides technical statistical details, applies the three approaches to the data from the example trial, and describes freely available statistical software (www.improveLTO.com) that implements each of the three statistical approaches.

Illustrative example

Randomized controlled trials with functional outcomes "truncated due to death" occur in diverse areas of medicine (box 1 gives three examples). In such trials it is common for investigators to restrict their analysis to survivors at a specified follow-up time point. We explain and contrast this approach and two alternative approaches using the context and design of one of the three examples: the Awakening and Breathing Controlled (ABC) trial.[7]

SUMMARY POINTS

- In randomized controlled trials of seriously ill patients, functional outcomes (eg, cognition, physical function, and quality of life) are increasingly being measured and compared across treatment arms

- If patients die before outcomes are evaluated, the functional outcomes no longer exist, ie these outcomes are "truncated due to death," creating challenges in the definition and statistical evaluation of the treatment effect

- In this article, we describe three statistical approaches—survivors only, survivor average causal effect, and composite endpoint—and discuss their advantages and disadvantages, including the required assumptions, some of which are not testable within the observed trial data

- If the treatment does not affect mortality, the three statistical approaches yield similar conclusions about the treatment effect on functional outcomes

- If the treatment affects mortality, the three statistical approaches can yield different conclusions, highlighting the need for careful consideration when planning a trial during which functional outcomes may be "truncated due to death"

Box 1: Example of randomized controlled trials with functional outcomes "truncated due to death"

These three examples show that randomized controlled trials of seriously ill patients with functional outcomes "truncated due to death" occur in diverse areas of medicine. When analyzing each of these trials, the investigators restricted their analysis of functional outcomes to survivors at the specified follow-up time point.

The multicenter Awakening and Breathing Controlled trial

- The multicenter Awakening and Breathing Controlled (ABC) trial[7] randomized 336 patients with acute respiratory failure receiving mechanical ventilation in an intensive care unit to spontaneous awakening trials plus spontaneous breathing trials (intervention) versus usual care plus spontaneous breathing trials alone (control). The primary outcome was ventilator-free days (a commonly used composite outcome of mortality and duration of mechanical ventilation[8]) during the 28 day study period. Key long term secondary outcomes assessed among a single study site (roughly 200 randomized patients) included mortality, as well as cognitive, psychological, and physical function, at 12 month follow-up. Mortality at 12 month follow-up favored the intervention (41% v 62%; P <0.01), and investigators compared the 12 month functional outcomes between treatment groups.

The CheckMate025 trial

- The CheckMate025 trial randomized 821 patients previously treated for renal cell carcinoma to nivolumab versus everolimus.[9] The primary outcome was all cause survival to 30 months. A key secondary outcome included quality of life, which was assessed at baseline and then every four weeks to 104 weeks. The mortality results favored nivolumab versus everolimus (45% v 52%; hazard ratio 0.73, P=0.002). The investigators compared the 104 week change from baseline in quality of life between treatment groups.

The PARADIGM-HF trial

- The PARADIGM-HF trial randomized 8442 patients with class II, III, or IV systolic heart failure to receive LCZ696 versus enalapril.[10] The primary outcome was a composite of death from cardiovascular causes or first admission to hospital for heart failure up to 42 month follow-up. Key secondary endpoints included all cause mortality by 42 months and functional outcomes. All cause mortality was lower for LCZ696 versus enalapril (17.0% v 19.8%; hazard ratio 0.84, P<0.001), and the investigators compared a functional outcome, the change from baseline to eight month follow-up in the clinical summary score on the Kansas City Cardiomyopathy Questionnaire (KCCQ), between treatment groups.

The ABC trial randomized patients with acute respiratory failure receiving mechanical ventilation on an ICU to spontaneous awakening trials plus spontaneous breathing trials (intervention) versus usual care plus spontaneous breathing trials alone (control). The primary outcome was ventilator-free days (a commonly used composite outcome of mortality and duration of mechanical ventilation[7]) during the 28 day study period. For each patient, survival information, including the time of death and a binary indicator of surviving to 12 months, were recorded. For patients who survived to 12 months, several key secondary outcomes were measured, including cognition and a composite T score based on normative population data (mean=50 and SD=10, higher scores indicating better cognition).[7] Mortality at 12 month follow-up favored the intervention (41% v 62%; P<0.01), and among 12 month survivors, cognition was comparable between the intervention and control arms (41 v 42, respectively, P=0.66).

Definitions

To define the causal effect of treatment on cognition, we adopted Rubin's potential outcomes framework[11]; thus, defining what would happen to a patient under each of two treatment arms (box 2). We first defined potential survival experiences, both time to death and a binary 12 month survival indicator, under receipt of the intervention and control, and four stratums of patients.[12]

- "Always survivors" survive to 12 months regardless of the treatment assignment
- "Mortality benefiters" survive to 12 months under the intervention but not under control
- "Always diers" die before 12 months regardless of the treatment assignment
- "Specials" die before 12 months under intervention but survive to 12 months under control.

The always survivors may be considered the most robust patients, the mortality benefiters to be of typical health status, and the always diers as frail.

Having defined the potential survival experiences, we then defined 12 month cognition, which is only defined if the patient survives to 12 months (see box 2). Specifically, under the intervention, cognition at 12 months can be defined for the always survivors and mortality benefiters but not for the always diers and specials. Similarly, under the control, 12 month cognition can be defined for the always survivors and specials, but not for the mortality benefiters and always diers. This "truncation due to death" creates a challenge in defining the effect of treatment on the functional outcome. We review three proposed approaches; readers interested in statistical details, application of the proposed approaches to the ABC trial data, and a description of a freely available software package to implement the proposed approaches, are referred to the appendix.

Survivors analysis

A simple and commonly used statistical approach is to compare functional outcomes between randomized groups only among survivors (see box 1). In the ABC trial, the treatment effect as defined in a survivors analysis is the difference in mean 12 month cognition comparing the patients who survived to 12 months under intervention compared with those who survived to 12 months under control. The types of patients who survive under intervention (always survivors and mortality benefiters) may be inherently different from patients who survive under control (always survivors and specials) such that analyses based on only those who survive—a post-randomization event—may produce misleading comparisons across the treatment groups—for example, when survival differs across the treatment groups (table 2).[6 11 13]

Survivor average causal effect

As an alternative to the survivors analysis, the causal (direct) effect of the treatment on the functional outcome can be defined.[12] In a randomized controlled trial where no mortality occurs, the causal effect of treatment on 12 month cognition for an individual patient is defined, with respect to the potential outcomes, as the difference between the 12 month cognition under intervention and the 12 month

Box 2: Definitions of potential outcomes for functional outcome "truncated due to death"

We define the potential outcomes for a randomized trial with two treatment arms that follows patients until a specific follow-up time point (eg, 12 months after randomization). The trial records the time of death for patients dying by the follow-up time point and an indicator for survival to the follow-up time point. At the follow-up time point, the trial measures functional outcomes (eg, a continuous score measuring cognition) among survivors.

Each patient has two potential outcomes: what would happen to the patient in the randomized controlled trial if he or she were to receive the intervention, and what would happen to the patient in the randomized controlled trial if he or she were to receive the control

- Under each treatment arm, the potential outcomes are:
- Time (in days) from randomization to death and the indicator of 12 month survival set to 0 if the patient dies before 12 months
- Indicator of 12 month survival set to 1 and the functional outcome if the patient survives to 12 months
- Patients can be grouped based on their potential survival experiences (see table 1). The 12 month functional outcome is only defined if the patient survives to 12 months.

Table 1 | Patient groups based on potential survival experiences

	Survive to 12 months		Functional outcome at 12 months	
Patient stratum	Intervention	Control	Intervention	Control
Always survivors	Yes	Yes	A	C
Mortality benefiters	Yes	No	B	Undefined
Always diers	No	No	Undefined	Undefined
Specials	No	Yes	Undefined	D

We define three treatment effects:

- Survivors analysis: the treatment effect is the difference in the mean functional outcome among survivors under intervention and control—that is, A+B v C+D
- Survivor average causal effect (SACE): the treatment effect is the difference in the mean functional outcome among patients who would survive regardless of treatment assignment, also known as the always survivors, A v C
- Composite endpoint: define a new variable that orders the patients based on the time of death among patients who die and the functional outcome among survivors. Compare the distribution of the new variable under intervention versus control.

[1]See appendix for more technical details and statistical notation.

cognition under control. The average causal effect is defined as the average of the individual causal effects over the population of patients. However, in randomized controlled trials where mortality occurs, the individual causal effect of the intervention on 12 month cognition is not defined for all patients. The always survivors are the only subset of patients for whom 12 month cognition is defined under both intervention and control. The survivor average causal effect (SACE) is defined as the average of the individual causal effects among the always survivors (see box 1).

Since patients are only assigned to one treatment group during a trial, assumptions are required to identify the proportion of patients who are always survivors and to draw inference about the SACE.[5 14-17] To identify the proportion of patients who are always survivors, it is typically assumed that there are no specials—referred to as the monotonicity assumption since it assumes that the survival experience under the intervention must be at least as good as that under control. Under the monotonicity assumption, patients

randomized to control who survive to 12 months are always survivors, so the proportion of patients who are always survivors is estimated by the proportion of patients randomized to receive the control who survive to 12 months. Another consequence of the monotonicity assumption is that the mean 12 month cognition under control among the always survivors can be estimated by the sample mean 12 month cognition among the 12 month surviving control patients.

To estimate the SACE, we must make additional assumptions that differentiate the surviving patients who received intervention since these patients are a mix of always survivors and mortality benefiters. One such assumption may be that under intervention, the always survivors will have, on average, better cognition (higher average cognition scores) compared with the mortality benefiters, referred to as the ranked average score assumption.[4] This assumption allows for a lower or upper bound for the mean 12 month cognition under intervention among the always survivors (lower bound: sample mean 12 month cognition among the surviving intervention patients, upper bound: sample mean of the n×p largest 12 month cognition score among surviving intervention patients, where n is the number of intervention survivors and p is the proportion of always survivors).

The SACE approach requires making assumptions that cannot be tested using the observed data (table 2). Eliciting expert opinion may help to ensure these assumptions are valid, but it is unlikely that the assumptions will be valid in all trials; the monotonicity assumption, for example, may not be valid in acutely ill patients who may be more susceptible to previously unrecognized adverse effects of interventions. Several trials, such as the Transfusion Requirements in Critical Care (TRICC) trial[18] that compared liberal blood transfusion with restrictive blood transfusion triggers in critically ill patients, have uncovered harmful effects of interventions previously thought to be safe. In addition, although the SACE approach evaluates the causal effect of treatment directly on the functional outcome, it violates the intention to treat principle by estimating the effect of treatment only among a select subset of patients.

Composite endpoint approach

The composite endpoint approach eliminates the need to condition on survival by combining information about mortality and functional outcomes into a single variable. An ordering is required to create a composite endpoint. Lachin created an ordering using timing of death.[19] In the context of our example, the following ordering (from worst to best outcome) could be considered: (a) earlier mortality is considered worse than later mortality and (b) among those who survive to 12 months, a lower cognition score is considered worse than a higher score. Since higher cognition scores correspond to better cognition in our example, one can define a new outcome that is equal to the time of death if a patient dies before 12 months and,

Table 2 | Three statistical approaches for comparing the effect of randomized treatments on functional outcomes "truncated due to death"

Statistical approach	Target patient population	Advantages	Disadvantages
Survivors analysis	Patients who survive in each treatment group	− Simple to compute − If mortality is independent of treatment, provides an estimate of the causal effect of the treatment on the always survivors	− If the effect of the treatment is to sustain frail patients, then the estimated treatment effect is a biased estimate of the causal effect of the treatment on the always survivors and may be misleading − Randomization is not guaranteed to be preserved within the subset of survivors − Does not include all randomized patients, violating the intention to treat principle for randomized trials
Survivor average causal effect (SACE)	Always survivors—that is, patients who would survive regardless of which treatment they receive	− Estimates the causal effect of the treatment on the functional outcome among always survivors − Randomization is preserved within the subset of always survivors so the comparison of randomized treatment groups is unbiased − When the intervention does not affect mortality, SACE is identified from the observed data (ie, all survivors are always survivors) without making any untestable assumptions	− Comparison of the functional outcome is made among a subset of patients (ie, those who would survive regardless of which treatment they received) that is not directly identifiable − Requires eliciting expert opinion on assumptions that are not testable within the observed data (eg, there are no patients who would survive under control but who would die under intervention) − Does not include all randomized patients, violating the intention to treat principle for randomized trials
Composite endpoint	All randomized patients	− Simple to compute (both quantiles of and the rank statistic for the distribution of the composite outcome) − All randomized patients are included in the analysis, consistent with the intention to treat principle − Randomization is preserved so the comparison of randomized treatment groups is unbiased − Provides a hypothesis test for comparing the distribution or a specific quantile (eg, median) of the composite endpoint across treatment groups that has a causal interpretation − The distribution of the composite endpoint in each treatment group can be examined to identify if treatment groups differ in mortality and/or the functional outcome	− Requires eliciting expert opinion on ordering mortality and the functional outcome − The size of the rank statistic may be difficult to interpret

if a patient survives past 12 months, is equal to the cognition score plus a constant, where the constant is chosen to satisfy conditions (a) and (b); for example, if survival is measured in days, the constant can be a number greater than 365. Under these assumptions, the potential composite outcome under receipt of intervention and control can be conceptualized. Because each of these potential outcomes is the composition of two factors with different scales, it is inappropriate to draw inference about the difference in means between the potential composite outcomes. Rather, the researcher can compare the distributions of the potential composite outcomes and, in particular, quantiles of these distributions. The effect of the treatment may be quantified and compared statistically using a single predetermined quantile (eg, the median) of the treatment specific distributions. Alternatively, the treatment effect can be defined and tested using a rank statistic that builds on the concepts underlying the Mann-Whitney-Wilcoxon rank sum test with ties and is based on the entire treatment specific distributions. This statistic is defined as the difference in the probability that a randomly selected individual exposed to intervention will have a higher composite outcome than a randomly selected individual exposed to control and the probability that a randomly selected individual exposed to control will have a higher composite outcome than a randomly selected individual exposed to intervention. This rank statistic can vary from −1 to 1, with a value of 0 indicating no treatment effect and positive (negative) values favoring the intervention (control). The composite endpoint approach adheres to the intention to treat principle but requires an assumption, ideally elicited from experts, regarding ordering of patient outcomes (table 2).

Comparison of the three approaches

Using key features of the ABC trial, we demonstrate when the three statistical approaches provide the same or different conclusions about the effect of treatment. Consider the population of patients eligible for the ABC trial. For each of these patients, we assign potential outcomes under control to mimic the control arm data from the ABC trial. Specifically, 62% of control patients will experience death by 12 months, the mean cognition score among survivors is 42, and based on the composite outcome, 50% of patients survive past 72 days (25% survive to 12 months with a cognition score >39). Then, assuming the monotonicity assumption holds, we generate the potential outcomes under intervention for six scenarios defined by the effect of the intervention on mortality and cognition among always survivors. For each scenario, we present the proportion of patients in each stratum and the true population values of the three treatment comparisons (table 3). The appendix presents figures displaying the distribution of the potential composite outcomes and the technical details for each scenario.

When the intervention has no effect on mortality (scenarios 1 and 2), there are no mortality benefiters, only always survivors (38%) and always diers (62%) (table 3). In this case, the SACE approach and survivors analysis produce the same results regardless of whether there is no effect (scenario 1) or a positive effect of the intervention on cognition (scenario 2). Further, when the intervention has no effect on mortality, the SACE can be estimated directly from trial data without making any additional assumptions. The composite endpoint approach identifies that there is no difference in the distribution of the potential composite outcomes in scenario 1 (rank statistic=0) and the shift in

Table 3 | Illustrative example to show differences between three statistical approaches comparing the effect of treatments on functional outcomes "truncated due to death." We consider the population of patients eligible for the ABC trial and for each patient, generate the potential outcome under control by mimicking characteristics of the control group from the ABC trial such that 62% of the patients receiving the control would experience death by 12 months, the mean cognition among survivors is 42, and based on the composite outcome, as described in the manuscript, 50% of the patients will survive past 72 days (25% of patients will survive to 12 months with cognition score >39). Assuming the monotonicity assumption holds, potential outcomes under intervention were derived based on the effect of the intervention on mortality and the functional outcome among always survivors (for a total of six scenarios). For each scenario, the true population value of the treatment effect defined by the survivors analysis, the SACE and composite endpoint approach are displayed

Treatment improves cognition among always survivors	Intervention has no effect on mortality		Intervention promotes survival in patients with robust or high cognition		Intervention sustains patients who are less healthy or have poor cognition	
	No	Yes	No	Yes	No	Yes
Scenario	1	2	3	4	5	6
Always survivors	38%	38%	38%	38%	38%	38%
Mortality benefiters	0%	0%	21%	21%	21%	21%
Always diers	62%	62%	41%	41%	41%	41%
Treatment effects:						
Survivors analysis	0	15	0	15	-7	2
Survivor average causal effect	0	15	0	15	0	15
Composite endpoint*						
Rank statistic	0	0.12	0.36	0.55	0.29	0.41
Median under intervention	Survive >72 days	Survive >72 days	Survive 12 months with cognition >34	Survive 12 months with cognition >49	Survive 12 months with cognition >20	Survive 12 months with cognition >20
75th centile under intervention	Survive 12 months with cognition >39	Survive 12 months with cognition >54	Survive 12 months with cognition >41	Survive 12 months with cognition >56	Survive 12 months with cognition >39	Survive 12 months with cognition >54

*For composite endpoint analysis, the median under control is survive >72 days, the 75th centile under control is survive to 12 months with cognition score >39.

the distribution of cognition scores among always survivors in scenario 2 (rank statistic >0).

In scenarios 3 through 6, the intervention has a positive effect on mortality, which results in no change to the proportion of always survivors (38%) but creates a stratum of mortality benefiters (21%, table 3). When the intervention has a positive effect on mortality, the SACE approach and survivors analysis will not necessarily provide the same results; whether or not these approaches are equivalent depends on characteristics of the mortality benefiters. In scenarios 3 and 4, the effect of the intervention is to promote survival in more robust patients with high cognition such that the 12 month cognition scores among the always survivors and mortality benefiters under intervention have the same distribution. In these scenarios, the survivors analysis and the SACE provide the same results (table 3). Alternatively, in scenarios 5 and 6, the effect of the intervention is to sustain less healthy individuals, creating a heterogeneous group of survivors (hearty always survivors and frail mortality benefiters). In these scenarios, the survivors analysis underestimates the SACE regardless of whether or not the intervention improves cognition among always survivors (eg, scenario 5: survivors analysis −7 v SACE 0) and use of the survivors analysis could lead to erroneous conclusions about the efficacy of the intervention on cognition. In scenarios 3 through 6, the composite endpoint approach identifies the shift in the distribution driven by the effect of the intervention on mortality and, in scenarios 4 and 6, on cognition.

Discussion

Different approaches are available for defining a treatment effect on a functional outcome that is

"truncated due to death." We have focused attention on the perspective of the regulator—that is, on advantages and disadvantages of the approaches within the context of principled design and analysis of randomized controlled trials. However, the ability of clinicians to interpret and communicate the potential benefits of interventions to patients may require different considerations. In the setting of severely ill patients, what potential benefits of the interventions are most relevant to the patient? Knowing the average benefit of an intervention among survivors is important; however, the clinician must also advise on characteristics of those who survive. Communicating the SACE to patients offers unique challenges to the clinician. Even if the treatment shows benefit among the always survivors, the clinician does not have the ability to identify whether a specific patient is in this group but could provide information on the distribution of always survivors and mortality benefiters under certain assumptions.

The composite endpoint approach offers the unique ability to provide useful thresholds to patients—for example, in scenario 2, 50% of patients receiving control survive past 72 days, whereas 50% of patients receiving intervention survive to 12 months with cognition scores greater than 54 (above the population norm of 50). However, the composite endpoint approach relies on the ability to order mortality and the functional outcome, which requires extensive input from clinicians and patients before conducting a randomized controlled trial. Studies examining the values and preferences of severely ill patients suggest that some patients consider severe cognitive impairment as a state worse than death[20]; however, more research is needed to rigorously rank mortality and functional outcomes according to patient values

and preferences. We have described in detail the composite endpoint approach proposed by Lachin, yet other composite endpoint approaches exist (eg, the Glasgow outcome scale has been used in randomized controlled trials of patients with serious neurological illness,[21] and quality adjusted life years (QALYs)[22] have been used in critical care trials) with similar challenges to those described above.

We have highlighted that there is no perfect choice among the three statistical approaches for defining a treatment effect when functional outcomes are "truncated due to death." The choice should be driven by expected effects of the intervention and by the target patient population. Keeping this in mind, we make two recommendations for planning a randomized trial among patients with a high risk of death. First, when it is biologically unlikely for treatment assignment to impact mortality, the survivors analysis provides an unbiased estimate of the effect of treatment on the functional outcome and can be used. Second, when patient mortality is the primary outcome and is hypothesized to differ across treatment groups, the survivors analysis may produce misleading results. This would be the case if the intervention promoted survival among patients with greater morbidity than the always survivors. In such studies, alternatives to the survivors analysis should be considered. If mortality and functional outcomes can be ordered, we recommend the composite endpoint approach over the SACE approach in such settings. While SACE defines a causal effect, the assumptions required to estimate the SACE are restrictive and untestable from the observed data.

1 Akacha M, Bretz F, Ruberg S. Estimands in clinical trials - broadening the perspective. *Stat Med* 2017;36:5-19. doi:10.1002/sim.7033

2 Fried TR, Bradley EH, Towle VR, Allore H. Understanding the treatment preferences of seriously ill patients [PMID: 11932474]. *N Engl J Med* 2002;346:1061-6. doi:10.1056/NEJMsa012528

3 Turnbull AE, Rabiee A, Davis WE. Outcome measurement in ICU survivorship research from 1970-2013: a scoping review of 425 publications. *Crit Care Med* 2016;44:1267-77. doi:10.1097/CCM.0000000000001651

4 Rubin DB. Discussion of "Causal Inference with Counterfactuals" by A.P. Dawid. *J Am Stat Assoc* 2000;95:435-7.

5 Zhang JL, Rubin DB. Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *J Educ Behav Stat* 2003;28:353-68. doi:10.3102/10769986028004353

6 Rubin DB. Causal inference through potential outcomes and principal stratification: Application to studies with "censoring" due to death. *Stat Sci* 2006;21:299-309. doi:10.1214/088342306000000114

7 Girard TD, Kress JP, Fuchs BD. Efficacy and safety of a paired sedation and ventilator weaning protocol for mechanically ventilated patients in intensive care (Awakening and Breathing Controlled trial): a randomised controlled trial[PMID 18191684]. *Lancet* 2008;371:126-34. doi:10.1016/S0140-6736(08)60105-1

8 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;30:1772-7. doi:10.1097/00003246-200208000-00016

9 Motzer RJ, Escudier B, McDermott DF, CheckMate 025 Investigators. Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. *N Engl J Med* 2015;373:1803-13. doi:10.1056/NEJMoa1510665

10 McMurray JJV, Packer M, Desai AS, PARADIGM-HF Investigators and Committees. Angiotensin-neprilysin inhibition versus enalapril in heart failure[PMID 25176015]. *N Engl J Med* 2014;371:993-1004. doi:10.1056/NEJMoa1409077

11 Rubin DB. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *J Educ Psychol* 1974;66:688-701. doi:10.1037/h0037350

12 Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics* 2002;58:21-9. doi:10.1111/j.0006-341X.2002.00021.x

13 Rosenbaum PR. The consequences of adjustment for a concomitant variable that has been affected by the treatment. *J R Stat Soc [Ser A]* 1984;147:656-66. doi:10.2307/2981697

14 Chiba Y, VanderWeele TJ. A simple method for principal strata effects when the outcome has been truncated due to death. *Am J Epidemiol* 2011;173:745-51. doi:10.1093/aje/kwq418

15 Hayden D, Pauler DK, Schoenfeld D. An estimator for treatment comparisons among survivors in randomized trials. *Biometrics* 2005;61:305-10. doi:10.1111/j.0006-341X.2005.030227.x

16 Egleston BL, Scharfstein DO, Freeman EE, West SK. Causal inference for non-mortality outcomes in the presence of death. *Biostatistics* 2007;8:526-45. doi:10.1093/biostatistics/kxl027

17 Tchetgen Tchetgen EJ. Identification and estimation of survivor average causal effects. *Stat Med* 2014;33:3601-28. doi:10.1002/sim.6181

18 Hébert PC, Wells G, Blajchman MA. A multicenter, randomized, controlled clinical trial of transfusion requirements in critical care. Transfusion Requirements in Critical Care Investigators, Canadian Critical Care Trials Group[PMID: 9971864]. *N Engl J Med* 1999;340:409-17. doi:10.1056/NEJM199902113400601

19 Lachin JM. Worst-rank score analysis with informatively missing observations in clinical trials. *Control Clin Trials* 1999;20:408-22. doi:10.1016/S0197-2456(99)00022-7

20 Fried TR, Bradley EH, Towle VR, Allore H. Understanding the treatment preferences of seriously ill patients[PMID: 11932474]. *N Engl J Med* 2002;346:1061-6. doi:10.1056/NEJMsa012528

21 McMillan T, Wilson L, Ponsford J, Levin H, Teasdale G, Bond M. The Glasgow Outcome Scale - 40 years of application and refinement. *Nat Rev Neurol* 2016;12:477-85. doi:10.1038/nrneurol.2016.89

22 Ferguson ND, Scales DC, Pinto R, Canadian Critical Care Trials Group. Integrating mortality and morbidity outcomes: using quality-adjusted life years in critical care trials. *Am J Respir Crit Care Med* 2013;187:256-61. doi:10.1164/rccm.201206-1057OC

Supplementary information: The appendix provides technical details for the three statistical approaches and the 6 illustrative scenarios, as well as, application of the approaches to the ABC trial data