UPenn Conference: Comments on Afternoon Talks

Michael P. Fay

National Institute of Allergy and Infectious Diseases

April 17, 2023

- "If it ain't easy, no one will use it". Thanks to Terry for his survival R package!
- Recommend really studying data after the primary analysis. Example....

Adaptive COVID-19 Treatment Trail (ACCT-1)

- Primary analysis: stratified logrank test on time to recovery in hospitalized COVID-19 (non-recovery and deaths censored at 29 days). Remdesivir increased recovery time compared to placebo. Highly significant. Beigel, et al, (2020).
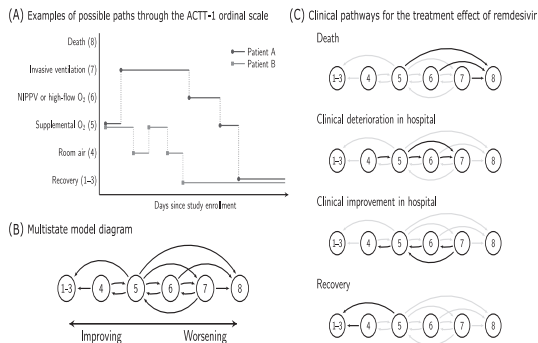- Secondary analysis: Multistate model. Fintzi, et al (2022).

from Fintzi, et al (2022)

(A) Examples of possible paths through the ACTT-1 ordinal scale

(B) Multistate model diagram

(C) Clinical pathways for the treatment effect of remdesivir

Figure 1. Multistate model for clinical progression for patients enrolled in the Adaptive COVID-19 Treatment Trial-1 (ACTT-1). *A*, Examples of possible paths through the ACTT-1 ordinal score (OS) scale. Both patients A and B are on supplemental oxygen (OS 5) at baseline. A standard time-to-event analysis assesses whether treatment with remdesivir shortens the expected time until the patients enter the recovered state (OS 1–3). Multistate analysis assesses whether treatment with remdesivir alters the dynamics of how patients travel throughout the ordinal scale over the course of the study. *B*, Multistate model diagram. Patients transition between states continuously in time. Arrows indicate which direct transitions are possible. For example, a patient starting on room air may transition to discharge or supplemental oxygen. However, the model assumes that a patient on room air would not be intubated without first receiving supplemental oxygen, whether "observed" or not from the perspective of data capturing. Note that the data are daily snapshots of each patient's status and that multiple transitions are possible within the same day. *C*, Clinical pathways for the treatment effect of remdesivir. Hazard ratio for remdesivir versus placebo is assumed to be common to all transitions within each transition group. For instance, we estimate that remdesivir slows down the rate of clinical deterioration within the hospital by a relative 26% (95% CI: 6%–43%) and that this effect applies to worsening from room air to supplemental oxygen (OS 4–5), supplemental oxygen to noninvasive positive-pressure ventilation (OS 5–6), or supplemental oxygen to invasive ventilation (OS 5–7). Sensitivity to groupings of transitions is explored in the supplement and the results are shown to be robust to how transitions are grouped. Abbreviations: CI, confidence interval; COVID-19, coronavirus disease 2019; NIPPV, noninvasive positive-pressure ventilation.
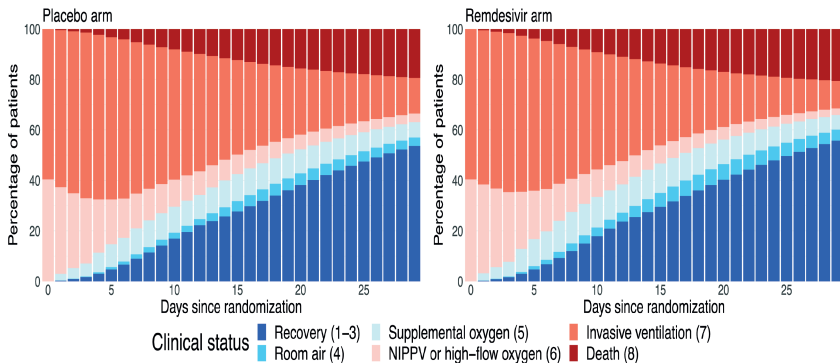
Clinical outcomes in patients receiving non-ICU respiratory therapies at baseline — ordinal scores 4 and 5

(A) Expected ordinal severity score distribution over the study period
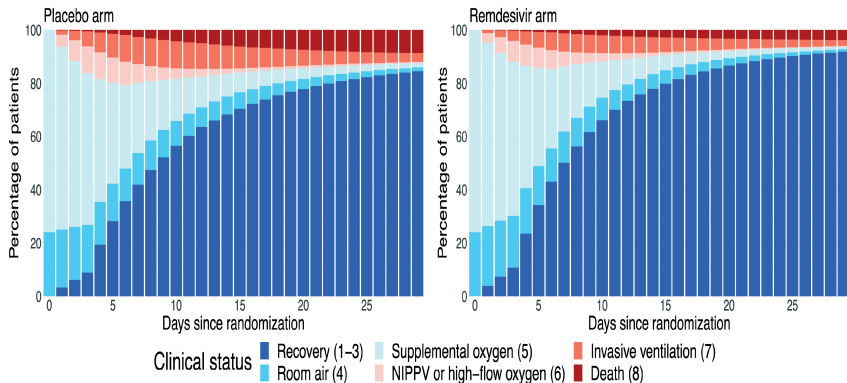
Clinical outcomes in patients receiving ICU respiratory therapies at baseline − ordinal scores 6 and 7

(A) Expected ordinal severity score distribution over the study period

Able to compare estimates of time in ICU for two arms.

Comments on Lu Mao's talk

Dr. Mao suggested two approaches to for constructing win ratio estimands:

- ► Nonparametric (specify τ): Restricted WR: $\frac{w_1(\tau)}{w_0(\tau)}$
- ► Semiparametric (proportional win-fractions model): Assume $\frac{w_1(t)}{w_0(t)} = \theta$ for all t
- ► For a primary endpoint estimand, it seems like it is safer to use the nonparametric estimand, because it does not requires the proportional win-fractions model. Would one ever use a semiparametric estimand for the primary endpoint estimand?
- ► nonparametric only requires independence assumptions on censoring to identify estimand. Seems like that is a less strict assumption than assuming proportional win-fractions.

Restricted mean time in favor

- $w_a(\tau)$ defined differently than in win ratio
- $w_1(\tau) =$
  $E(\text{ amount of time in } (0, \tau] \text{ when treated is better than control })$
- $\mu(\tau) = w_1(\tau) - w_0(\tau)$
- Very nice easy to interpret estimand, and R package available.

Comments on Anne Eaton's talk

- Beautiful idea
  - $P(D \geq t, Y(t) = 1) = P(D \geq t) \times P(Y(t) = 1 | D \geq t)$
  - $\hat{P}(D \geq t)$ with Kaplan-Meier
  - $\hat{P}(Y(t) = 1 | D \geq t)$ with kernel estimator
- Smooths out time to progression effect, so its effect is not so dependent on assessment visits

- ▶ Why use progression-free survival?
  - ▶ Progression is more common, effect easier to see.
  - ▶ Survival is more important, do not want to ignore it.
    (Do not treat death as censored!)
- ▶ Combining two endpoints does not help understand disease process better.
- ▶ Better for primary endpoint

Simple Example

- ▶ Interval censoring, twice as often in Trt B than in Trt A.
- ▶ Suppose progression-free survival and no one dies.
- ▶ What if treatment A is just a pain medication, so disguises pain, and you get assessed less often?
    - ▶ What is recommended in this case for a treatment effect estimand for progression-free survival?
    - ▶ If no deaths, then progression at observation time is invalid, but certain versions of logrank test are approximately valid (Fay and Shih, 2012).

Survival

time
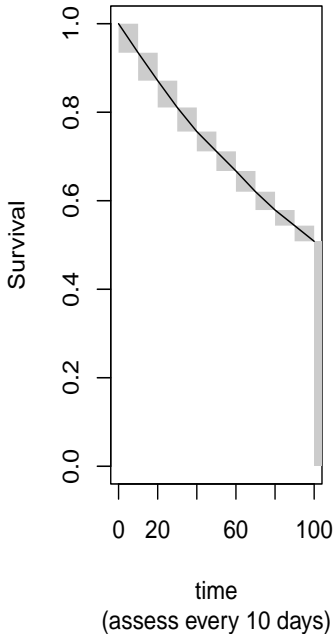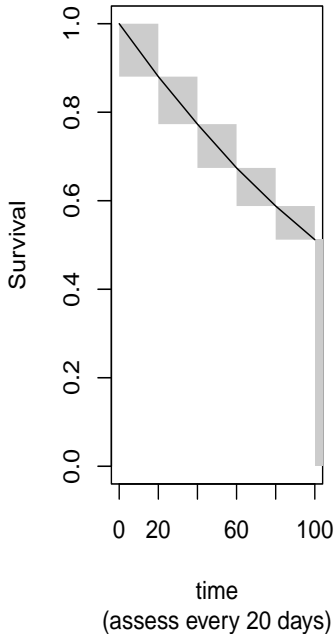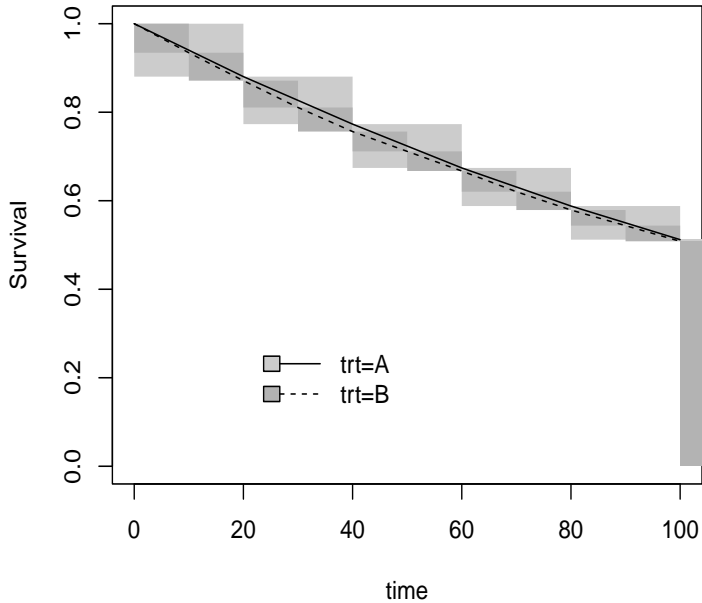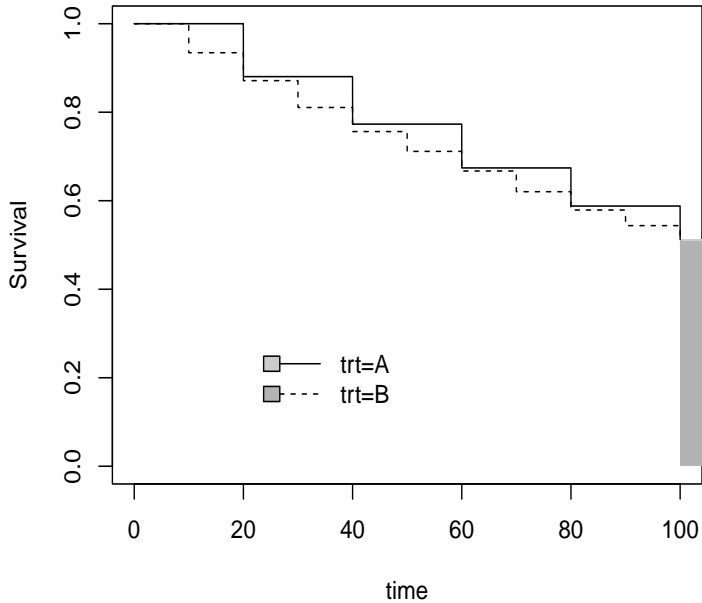
▶ What if the assessment process depends on treatment? Talk about recommendations (Eaton and Zabor, 2022).

Comments on Richard Cook's talk

▶ You recommend not making untestable assumptions. I would like to point out that we often assume independence of the censoring with the endpoint. Some assumptions are easier to accept than others.

▶ "Clinical trials are not primarily designed to enhance understanding of causal mechanisms but rather to test and estimate effects on marginal process features and facilitate regulatory decision making."

▶ Restatement: Clinical trials are a robust (i.e., relatively model independent way) for establishing causal effects on populations, not for understanding causal mechanisms on individuals. In usual two-arm trial, each individual is observed only under 1 arm. Compare treatment effect on arms, not on each individual.

Comments on Richard Cook's talk

- It was good to emphasize the problems with conditioning. That have been an issue with the usual proportional hazards model, and for this generalization it is good to mention the issue still applies!
- Collider bias, and hazard of hazard ratios.
- Example: Decreasing vaccine efficacy over time. Condition on being at risk for second half of study, then calculate vaccine efficacy for second part of study. Cannot interpret lower vaccine efficacy later to mean that the vaccine is losing its efficacy over time, could be that more frail/higher risk are eliminated early on from placebo arm.

▶ Hazard ratios as estimands
  ▶ hazard at any specific time is not a marginal estimand process feature, but ratio of cumulative hazards is a comparison of marginal process features (Vansteelandt, Dukes, Lancker, and Martinussen, 2022).
  ▶ Under proportional hazards assumption, that is a hazard ratio.

References

- Beigel, et al, (2020). (New England J of Medicine, Nov 5, 2020; 383:1813-26).
- Fay and Shih (2012) Weighted logrank tests for interval censored data when assessment times depend on treatment. Statistics in Medicine 31:3760-3772.
- Fintzi, et al (Clinicial Infectious Diseases, 2022; 74(112):2209-17).
- Vansteelandt, S, Dukes, O, Van Lancker, K, and Martinussen, T (2022): Assumption-Lean Cox Regression, Journal of the American Statistical Association, DOI: 10.1080/01621459.2022.2126362