Sample Size Re-estimation for COVID-19 Trials and Analysis Technique for Unique Endpoints

Pranab Ghosh & Ling Wang

Pfizer Inc.

September 10, 2020

Disclaimer

The examples and accompanying data in this presentation is based solely on hypothetical and simulated scenarios for illustration purposes only. No data or information presented in this presentation is derived from actual clinical studies, past or current.

Presentation Outline

- Challenges with trial design due to lack of prior information about the disease, endpoints.
- Appropriate choice about endpoint, that are meaningful.
- Sample size restriction, solution with adaptive design.
- Simulation results to convince the use of adaptive design.
- Ordinal scale endpoint and how to analyze the data.

Prior Information/Data

- The study will start before the natural history of the disease is fully understood.
- Ambiguity about the choice on standard of care (SOC)/placebo group.
- Highly accelerated timelines and poor study design resulted in confusion about the treatment effect in the SOC arm.
- Effect of the treatment mechanism with COVID-19 patients.
- Most of the prior data derived from observational studies, not from randomized clinical trials.

Challenges with Endpoint

- Heterogeneity in choice of endpoints among different trials.
- A Ordinal endpoint represents disease spectrum from full recovery to death was recommended by World Health Organization (WHO).
- There was different variation of this ordinal endpoint 6, 7 or 8 point.
- Effective size derived from hazard rate, time to death, time to worsening or time to recovery based on patient population in the trial.
- More subtle endpoint will be mortality rate/ Recovery rate during a fixed follow-up time period

Primary Challenges in Designing the Trial

- Patient population that is appropriate for the treatment evaluation.
- Time restriction and the study should start quickly to track with the epidemic curve.
- Producing meaningful information within a short time-span.
- Rapid change in treatment landscape, resulting in uncertainty about extracting the magnitude of a meaningful effect size.
- Relatively smaller number of patients due to being a proof-of-concept study.

Why Adaptive Design?

- Starting a large study without knowing much details about the treatment effect involves risk.
- Lack of prior information efficacy endpoint and placebo effect.
- Sample size restriction depending on the patient population and enrollment center.
- Prior use of adaptive design in the past during Ebola outbreaks in 2014 (PREVAIL trial - Dodd et al. (2016)).

Design Assumptions

- Primary endpoint : Mortality Rate at the end of treatment period.
- Type one error of 0.05 (1-sided).
- Initially planned to enroll 112 subjects for detecting 85% power, assuming 20% improvement in mortality rate (30% vs 10%).
- Interim Analysis with 56 subjects and sample size will be increased based on interim results.
- Possible futility stopping at this interim in case mortality rate is much higher than SOC with this new treatment.

Statistical Concepts around Adaptive Design

- Promising zone design based on conditional power computed from interim data.

$$CP^{n^*} = P\left(Z_2 > \Phi^{-1}(\alpha)|Z_1 = z_1\right)$$
$$= \Phi\left(\Phi^{-1}(\alpha)\sqrt{\frac{n}{n-n_1}} - \left[\sqrt{\frac{n_1}{n-n_1}} + \sqrt{\frac{n^*-n_1}{n_1}}\right] z_1\right)$$

- $n = 78, n_1 = 34$ are planned sample size at two analysis point.
- n^* is the targeted sample size for the decision making/conditional power computation.
- z_1 is the observed sample size at the interim.
- Final analysis will combine data before interim analysis Z_1 and after the interim analysis $Z_{(2)}$ with weights $w_1 = w_2 = \sqrt{0.5}$ (Cui, Hung, and Wang (1999))

$$Z^{CHW} = w_1 Z_1 + w_2 Z_{(2)}$$

Promising Zone Concepts

- Possible sample size increase is capped at 180.
- Favorable zone : Conditional power with planned sample size ($n^* = 78$) is higher than 80%.
- Promising Zone : If conditional power detected with maximum sample size is at least 50%, results will be considered as promising.
  - Sample size increased to detect 80%.
- Unfavorable Zone : CP with maximum sample size increase is still less than 50%.

Simulation Parameters

- Significance Level : one-sided 0.1
- Number of Simulations : 10,000
- Favorable Zone : $CP^{112} \geq 0.85$
- Promising Zone : $CP^{112} < 0.85$ and $CP^{160} > 0.5$
- Unfavorable Zone : $CP^{160} \leq 0.5$
- Futility Zone : $\hat{\delta} > 0.1$
- Interim Analysis : Based on mortality rate data from first 50% (n = 56) participants.
- Total Sample Size : 112
- Maximum Sample Size in the promising Zone : 160

Simulation Results : Optimistic Scenario (30% vs 10%)

Zone	P(Zone)	Power		Avg. SS	
		No-Adapt	Adapt	No-Adapt	Adapt
Futility	0.001	0.00	0.00	56	56
Unfavorable	0.16	0.40	0.40	112	112
Promising	0.22	0.82	0.89	112	142
Favorable	0.61	0.95	0.95	112	112
Overall	1.00	0.84	0.89	112	119

Simulation Results : Pessimistic Scenario (30% vs 20%)

Zone	P(Zone)	Power		Avg. SS	
		No-Adapt	Adapt	No-Adapt	Adapt
Futility	0.01	0.00	0.00	56	56
Unfavorable	0.36	0.33	0.33	112	112
Promising	0.43	0.69	0.8	112	145
Favorable	0.21	0.89	0.89	112	112
Overall	1.00	0.84	0.89	112	122

Simulation Results : Null Scenario (30% vs 30%)

Zone	P(Zone)	Power		Avg. SS	
		No-Adapt	Adapt	No-Adapt	Adapt
Futility	0.24	0.00	0.00	56	56
Unfavorable	0.60	0.02	0.02	112	112
Promising	0.11	0.16	0.19	112	147
Favorable	0.06	0.42	0.42	112	112
Overall	1.00	0.048	0.051	96	102

Ordinal Endpoints

- NIAID 8-point ordinal scale of disease severity ranging from patients.
  1. Discharged with no limitations on activities
  2. Discharged but with limitation on activities
  3. Hospitalized, not requiring supplemental oxygen
  4. Hospitalized, not requiring supplemental oxygen but requiring ongoing medical care;
  5. Hospitalized, requiring supplemental oxygen;
  6. Hospitalized, on non-invasive ventilation (NIV) or high flow oxygen device;
  7. Hospitalized, on invasive mechanical ventilation or ECMO;
  8. Death;

How to Analyze this Ordinal Endpoint Data

- Proportion of participants at the baseline and end of treatment period in each category of the ordinal endpoint.
- Stacked plot representing categorical/scale distribution over time.
- Fitting Proportional odds / Adjacent categories model raise concerns regarding model validity.
- Interpretation of odds ratio and reflection of the estimate becomes challenging in this scenario.
- Time to 1 or 2-point improvement in the 8-point ordinal scale using cox proportional hazard model.

Multi-State Model

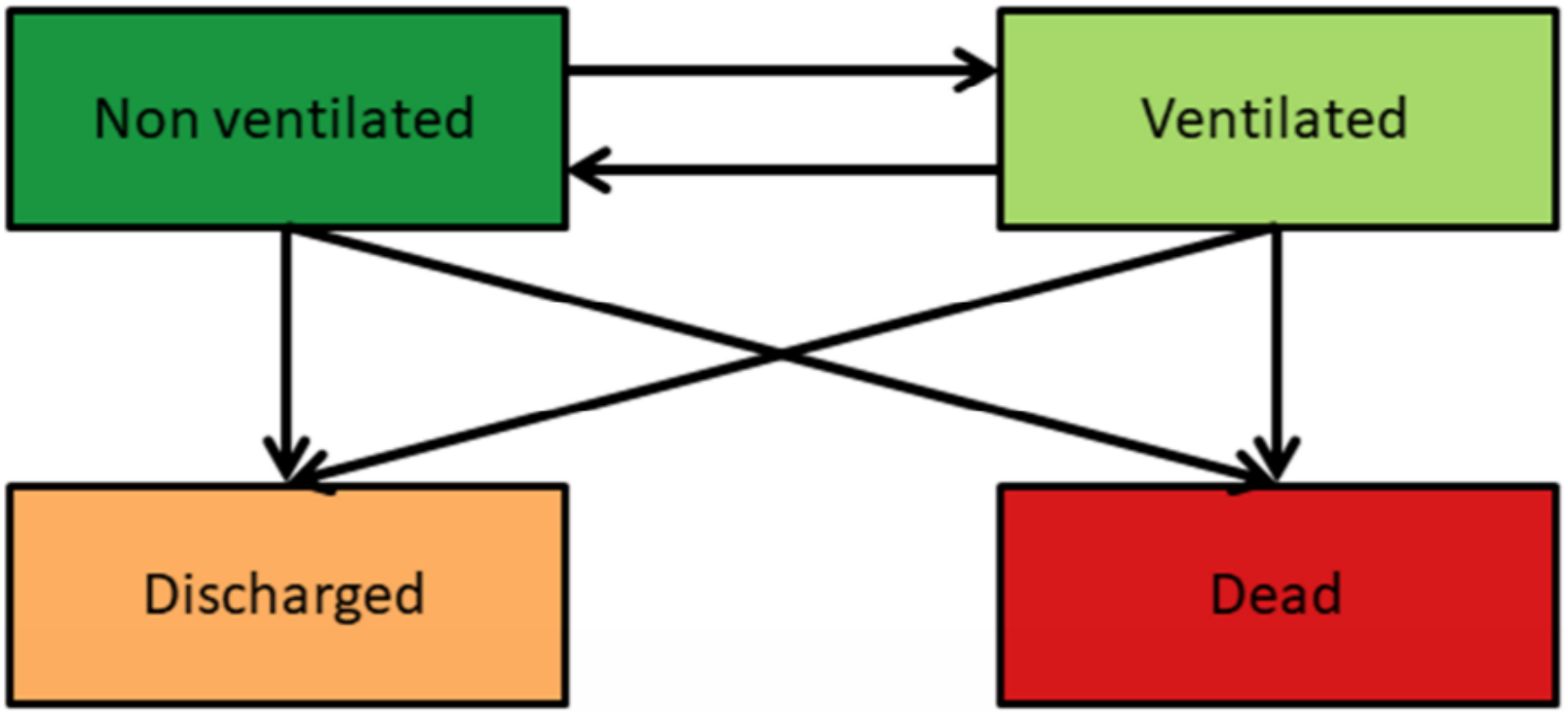
- Multi-state model to study different endpoints (death, discharged) and also time-to-event endpoints such as (hospital duration, ventilation time).
- Becoming more commonly used to analyze the natural history data with COVID-19 patients.
- Allows for a detailed investigation of treatment effects for various endpoints over the course of time thereby harmonizing differing endpoints and lengths of follow-up.
- Simulation experiment to explore the multi-state model in this settings.

Simulation Assumptions

- Only 4 disease category, by collapsing ordinal scales -
  1. Discharged (Scale 1 and 2)
  2. Hospitalized without Ventilation (Scale 3, 4 and 5).
  3. Hospitalized with Ventilation (Scale 6 and 7).
  4. Death (Scale 8)
- Baseline population comes from hospitalized patients with and without ventilation (based on collapsed scale 2 and 3) with 20% non-ventilated and 80% ventilated patients.
- Simulation experiment to explore the multi-state model in this settings.
- After baseline visits, patient's disease state will be recorded at several time point (continuous time point was not considered for the simulation exercise)
- R package **mstate** was used for the analysis.

Simulation Assumptions (cont.)

- Transition probability represents the rate of moving from one state to another state.
- Assumed to be constant over time and independent of any previous state.



Simulation Assumptions (cont.)

- Transition probability for the SOC arm (row represents current state and column is the very next state)

	Discharged	Non-Ventilated	Ventilated	Death
Discharged	0.9	0.05	0.03	0.02
Non-Ventilated	0.4	0.25	0.2	0.05
Ventilated	0.15	0.45	0.3	0.1
Death	0	0	0	0

- Transition probability for the Treatment arm

	Discharged	Non-Ventilated	Ventilated	Death
Discharged	0.9	0.05	0.03	0.02
Non-Ventilated	0.65	0.2	0.13	0.02
Ventilated	0.25	0.55	0.17	0.03
Death	0	0	0	0

Simulation Results

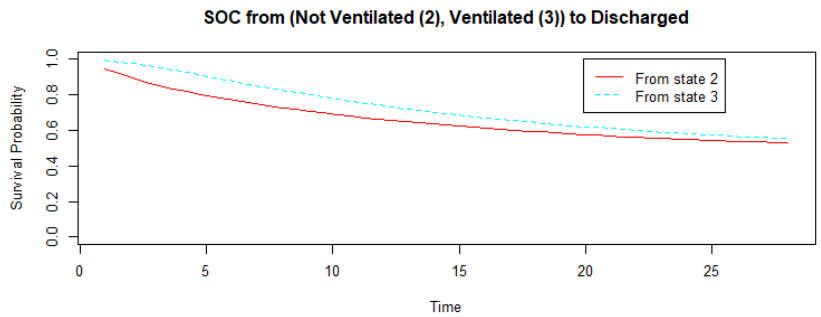
- Multi-state model used treatment group as only covariate

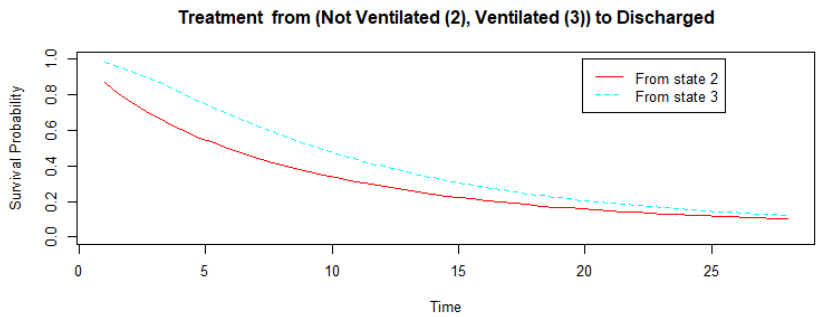
```
Call:
msm(formula = state ~ Day, subject = PTNUM, data = sim.data,     qmatrix = Q, gen.inits = TRUE, covariates = ~TRT.ID, deathexact = 4)

Maximum likelihood estimates
Baselines are with covariates set to their means

Transition intensities with hazard ratios for each covariate
                             Baseline                                    TRT.ID
Not Ventilated - Discharged     9.580e-02 (  7.871e-02,  1.166e-01) 2.321461 (1.574e+00,3.424e+00)
Not Ventilated - Not Ventilated -2.258e-01 ( -2.830e-01, -1.801e-01)
Not Ventilated - Ventilated     1.053e-01 (  6.855e-02,  1.619e-01) 0.845618 (3.568e-01,2.004e+00)
Not Ventilated - Death          2.462e-02 (  1.612e-02,  3.761e-02) 0.176562 (7.386e-02,4.221e-01)
Ventilated - Discharged         1.161e-06 (  2.260e-51,  5.959e+38) 0.526062 (9.186e-93,3.013e+91)
Ventilated - Not Ventilated     2.777e-01 (  2.197e-01,  3.509e-01) 1.266324 (7.925e-01,2.023e+00)
Ventilated - Ventilated        -2.777e-01 ( -3.509e-01, -2.197e-01)
Ventilated - Death              9.226e-07 ( 7.048e-171, 1.208e+158) 0.005334 (0.000e+00,      Inf)
```

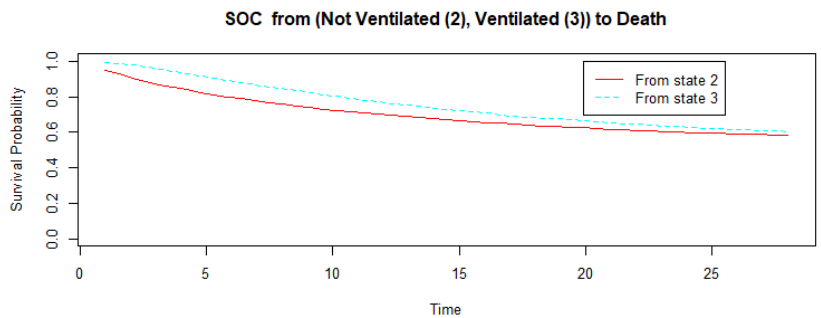
- Higher hazard ratio in moving to severe state means improvement with the treatment.
- Estimate average time in each state among arms
- Transition probability in the treatment group

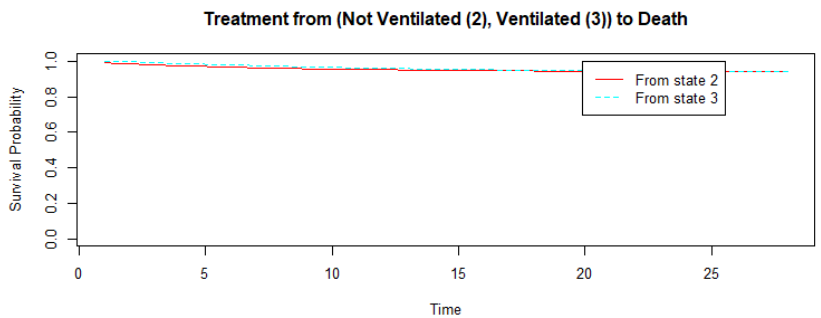
Arm	State	Estimate	SE	L	U
Placebo	Non-Ventilated	6.9	1.58	3.9	9.8
	Ventilated	5.2	1.55	3.3	9.3
Treatment	Non-Ventilated	4.9	1.8	2.1	8.8
	Ventilated	4.4	1.6	1.89	7.93

Simulation Results (cont)









- Higher hazard ratio in moving to severe state means improvement with the treatment.

References

- Cui, L., Hung, H. M. J., and Wang, S.-J. (1999), "Modification of Sample Size in Group Sequential Clinical Trials," Biometrics, 55, 853–857.
- Cook RJ, Lawless JF. Multistate Models for the Analysis of Life History Data. CRC Press; 2018.
- Beyersmann J, Gastmeier P, Grundmann H, Bärwolff S, Geffers C, Behnke M, et al. Use of multistate models to assess prolongation of intensive care unit stay due to nosocomial infection. Infection Control. 2006;27(05):493–499.
- de Wreede LC, Fiocco M, Putter H (2011). "mstate: An R Package for the Analysis of Competing Risks and Multi-State Models." Journal of Statistical Software, 38(7), 1–30
- Maja von Cube, Marlon Grodd, Martin Wolkewitz, Derek Hazard, Jerome Lambert. "Harmonizing heterogeneous endpoints in COVID-19 trials without loss of information - an essential step to facilitate decision making", medRxiv 2020.03.31.20049007; doi: https://doi.org/10.1101/2020.03.31.20049007