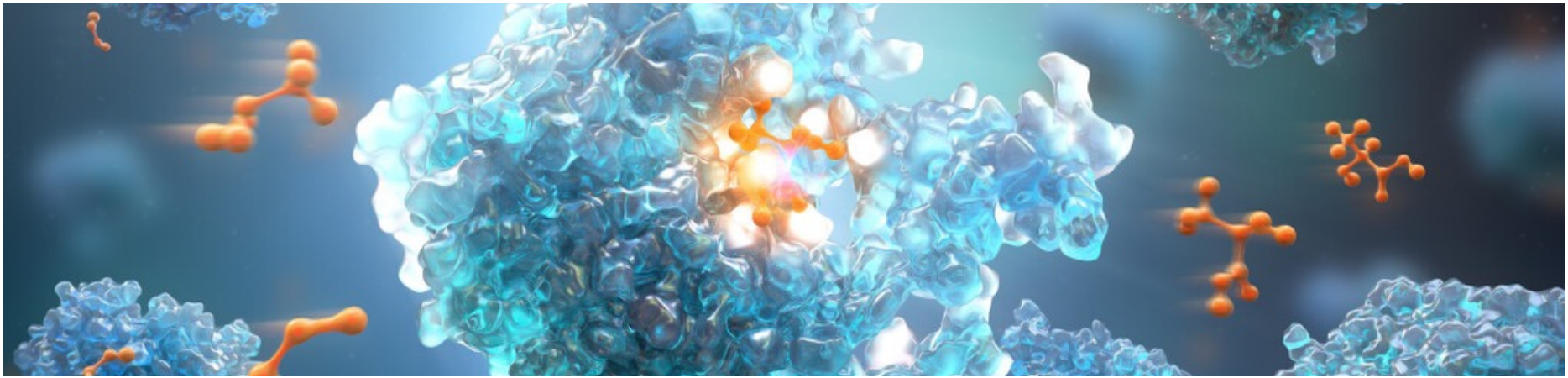
Statistical Challenges of Designing Covid-19 Therapeutic and Prophylaxis Studies

Julia Niewczas[1], Fredrik Öhrn[1], Carl-Fredrik Burman[1], Alexander Bore[1], David Wright[2], Ian Hirsch[3], Michael O'Kelly[4]

[1] Early Biometrics & Statistical Innovation, Data Science & Artificial Intelligence, R&D, AstraZeneca, Gothenburg, Sweden
[2] Early Biometrics & Statistical Innovation, Data Science & Artificial Intelligence, R&D, AstraZeneca, Cambridge, UK
[3] Biometrics, Late-stage Development, Respiratory and Immunology (R&I), BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK
[4] Center for Statistics in Drug Development, Decision Sciences, Data Sciences, Safety and Regulatory, IQVIA, Fairview Dublin, Ireland

BBSW Symposium

05 Nov 2020



Overview

1. Introduction
2. WHO Endpoints
3. Design and Analysis of vaccine trials
4. Conclusions

Introduction

- Covid-19 pandemic has imposed a lot of challenges in pharmaceutical research
  - Ongoing studies
    - Paused/delayed
    - Interpretability of results
    - Credibility of studies
  - Covid-19 therapeutic and prophylaxis studies
    - Delivering efficacious and safe compounds to patients fast

Introduction

- AstraZeneca has started both therapeutic and prophylaxis trials
  - ACCORD-2 Platform Study
  - CALAVI Study
  - DARE-19 Study
  - Vaccine studies

Therapeutic Trials

How can we define endpoints for therapeutic trials?

- WHO recommended ordinal scale[1]

Patient State	Descriptor	Score
Uninfected	No clinical or virological evidence of infection	0
Ambulatory	No limitation of activities	1
	Limitation of activities	2
Hospitalised – mild disease	Hospitalized—no oxygen therapy	3
	Oxygen by mask or nasal prongs	4
Hospitalised – severe disease	Noninvasive ventilation or high-flow oxygen	5
	Intubation and mechanical ventilation	6
	Ventilation and additional organ support—pressors, RRT, ECMO	7
Dead	Death	8

[1] World Health Organization (2020a), "WHO R&D Blueprint Novel Coronavirus COVID-19 Therapeutic Trial Synopsis," available at *https://www.who.int/blueprint/priority-diseases/key-action/COVID-19_Treatment_Trial_Design_Master_Protocol_synopsis_Final_18022020.pdf* .

How can we define endpoints for therapeutic trials?

- How can we assess patients based on WHO scale?
- Endpoints can be still defined in multiple ways[1]

Endpoint	Use of ordinal scale	Follow-up time
Improvement in the scale at a time point	Improvement from any point on the scale, conditional on being at that point or worse (proportional odds approach)	Day 14
Improvement in the scale at a time point	Mean ranks based on the scale, stratified by baseline score (Cochran–Mantel–Haenszel test)	Day 14
Improvement at a time point (Y/N)	2 scale points reduction from baseline, or achieving score ≤2	Day 14
Patient trajectory in scale over time	Ranking based on best and worst score achieved, and time on these	28 days
Time to improvement	Two scale points reduction at any time or achieving score ≤2	28 days
Time to discharge	Achieving score ≤2 at any time	28 days
Time to recovery	Achieving score ≤3 at any time	28 days
Time to worsening	Worsening by 2 scale points or death at any time	28 days
Time to death	Score indicates death	28 days

[1] O'Kelly, Michael, and Siying Li. "Assessing via Simulation the Operating Characteristics of the WHO Scale for COVID-19 Endpoints." *Statistics in Biopharmaceutical Research* (2020): 1-10.

Issues with some endpoints

- Not clear at which time point (binary) endpoints of improvement should be measured
- Time to improvement/worsening endpoint might face the issue of non proportional hazards
- Further, issues of semi-competing risk of death/discharge
  - Death may result in the censoring of time to improvement, but not vice versa; and
  - Discharge may result in the censoring of time to worsening, but not vice versa

Issues with some endpoints

- Measure of improvement
  - WHO scale might miss the efficacy of a treatment improving the overall experience of the patient
  - Improvement the same as SoC but time in the Intensive Care Unit (ICU) could be shortened
- Proportional odds analysis: estimates the odds of improvement from current score to the next best score, conditional on the current score
  - Assumption that the odds of improvement are the same for progress from any score to the next best score;
  - Might not detect the effectiveness of a treatment that inhibits worsening but is not expected to dramatically improve symptoms
  - Could make use of the ordinal scale in another way via the Cochran-Mantel-Haenszel (CMH) test

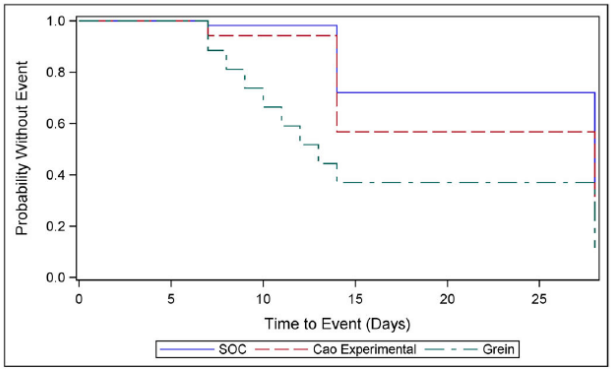
Patient trajectory in scale over time

- The following ranking could be used:
  - Sort patients in order of
    - Death (0/1 = No/Yes)
    - Score on WHO scale at Day 28
    - Best score on WHO scale occurring after the worst
    - (descending order) duration of the best score as defined in the previous bullet
    - Worst score on WHO scale
    - Days on worst score
- Patients are then ranked in sort order.
  - First rank is best
  - Patient experience worse with lower ranks.
  - In the above ranking scheme, death is most important factor; within those alive, improvement next most important; worsening is taken into account, but as a lower priority.
- Other ranking schemes could be used, tailored to the stage of COVID and/or the expected treatment effects; death should always have high(est) priority in ranking

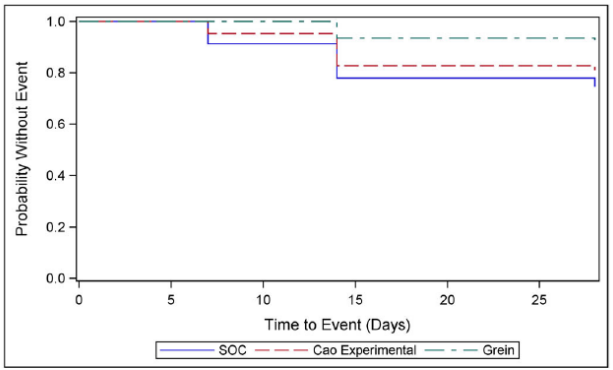
Example from O'Kelly and Li (2020)

- Simulated scenarios based on Cao et al (2020) and Grein at el (2020)

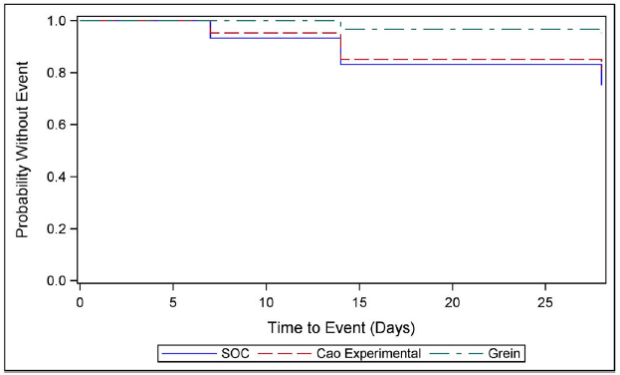
(a)



(b)



(c)



Kaplan–Meier estimates of survival probability, showing (a) time to Improvement; (b) time to worsening; (c) time to death

Patient trajectory in scale over time more robust

- Example from O'Kelly and Li (2020)
- Scenario A:
  - Experimental arm and SoC based on Cao et al. 2020
- Scenario C:
  - experimental arm from Grein et al. (2020) vs. experimental arm from Cao et al. (2020)

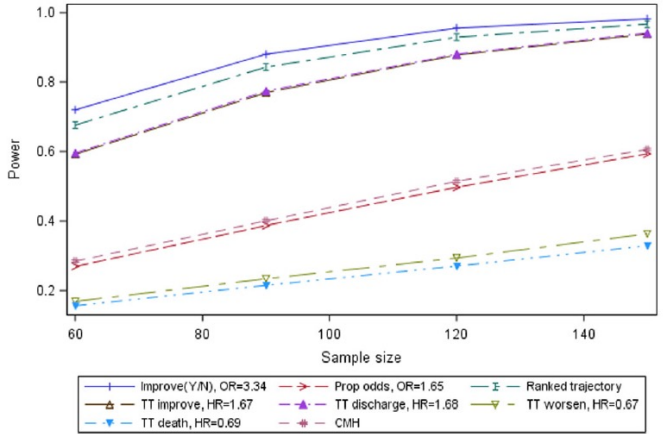


Figure 3. Power by sample size for selected endpoints, Scenario A.

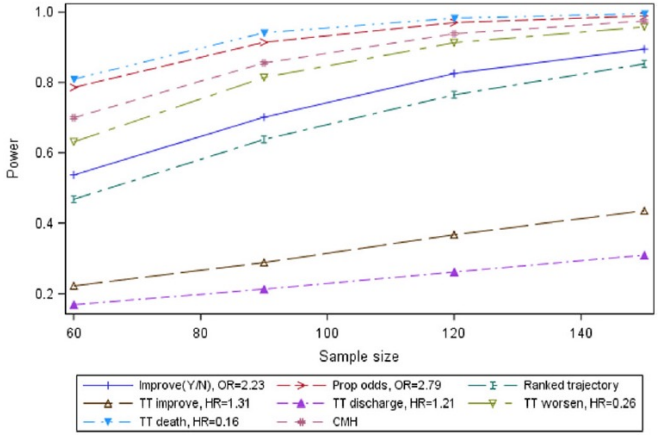


Figure 5. Power by sample size for selected endpoints, Scenario C.

O'Kelly, Michael, and Siying Li. "Assessing via Simulation the Operating Characteristics of the WHO Scale for COVID-19 Endpoints." *Statistics in Biopharmaceutical Research* (2020): 1-10.

Conclusions

- One primary endpoint will not fit all candidate treatments and SoCs for COVID-19 trials
- The primary endpoint may need to vary depending on the expected trajectory of a candidate treatment, and on how it is expected to show superiority to SoC
- Ranked trajectory endpoint takes worsening as well as improvement into account and is estimated to have moderate to good power for a wide range of trajectories of the WHO scale

Prophylaxis Studies

Defining Vaccine Efficacy

- Primary endpoint: Vaccine Efficacy (VE)
  - Commonly defined as 1 – Relative Risk (RR)
  - $VE = 1 - RR = 1 - \frac{d_V \times N_C}{d_C \times N_V}$, where d_i $(i = C, V)$ is the number of cases in the control/vaccine arm, and N_i $(i = C, V)$ corresponds to the number of patients
  - Standard error on log-scale using (Zou, 2004)
    - $SE(RR) = \sqrt{1/d_V - 1/N_V + 1/d_C - 1/N_C}$

- Null hypothesis is rejected when the lower bound of the confidence interval for the **VE** is higher than X% (e.g. 0%, 20%, 30%)
  - Super-superiority

Discreteness of the Data

- Suppose we perform an analysis after 60, 61, 62 and 63 cases at 2.5% alpha
  - Lower CI Bounds >= 30%
- If we observe 17 vaccines cases and perform an analysis at 60, we have a significant result
- If we were to perform an analysis at 61 or 62 but have 18 vaccine cases, the result is no longer significant

	17 cases on vaccine	18 cases on vaccine	19 cases on vaccine
CI Bound 60 cases	**0.307**	0.256	0.202
CI Bound 61 cases	0.324	0.275	0.223
CI Bound 62 Cases	0.340	0.293	0.242
CI Bound 63 Cases	0.356	**0.310**	0.261

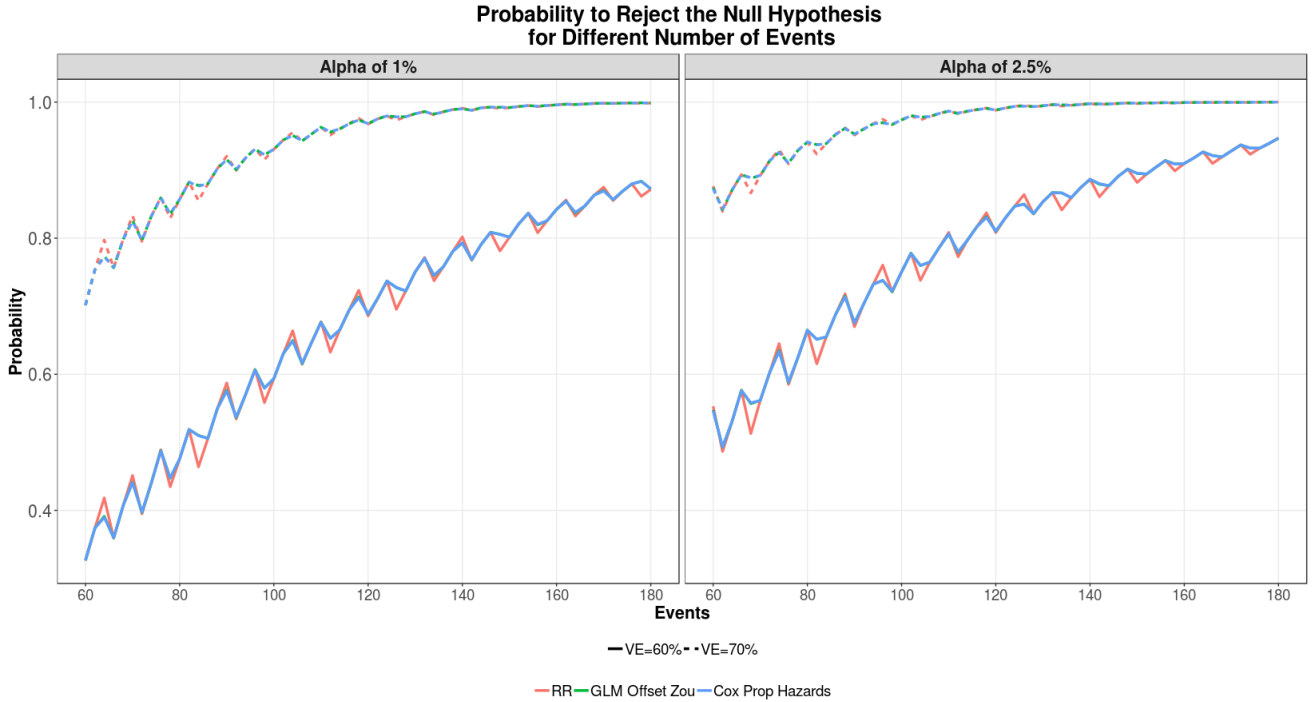
Defining Vaccine Efficacy

- VE could also be obtained from
  - Time-to-event analysis with Hazard Ratio,
  - Generalised Linear Models (GLMs) with Poisson or LogBinomial link
    - Could also be adjusted for exposure time

- For all of the approaches, the Lower Bound of CI of the estimate would be compared against threshold

Is there a difference in power for different methods?

- Relative Risk with Zou Standard Error
- GLM with Poisson link, offset term and sandwich estimator for SE
- Cox Proportional Hazards Model



Analysis Methods

- The full effect of the vaccine might not be observed from Day 1
  - Vaccine might reach full effect after e.g. 7, 14 or 21 days
  - Could be thought of as non-proportional hazards in a time-to-event setting
- How should this data be handled?
  - Patients having an infection prior day X could be removed from the analysis (per protocol)
    - Exposure time (if considered) would be counted after Day X
  - One could ”censor” the patients but still include them in the analysis

Would different endpoints and analysis methods lead to different conclusions?

- Consider a Phase III efficacy study
  - 10,000 patients per arm (1:1 randomisation)
  - 60% VE, with Lower CI Bound > 30%
  - Three analyses
    - At 60, 120 and 180 cases
    - Lan DeMets Spending function approximating Pocock Boundaries
    - One-sided alpha of 1.13%, 1.09%, 1.09%
  - 2-month linear recruitment
  - Responses simulated from an exponential model with yearly attack rate of 4%
  - Vaccine efficacy measured from Day 21

Operating Characteristics

Analysis	Method	VE	Prob Reject IA1	Prob Reject IA2	Prob Reject FA
Remove patients with events before 21 days	Relative Risk (Zou SE)	0.600	33.1%	68.9%	87.6%
	GLM with Offset	0.602	34.3%	72.0%	89.2%
	GLM with Offset Zou SE	0.602	34.3%	71.9%	89.2%
	Cox Prop Hazards	0.602	34.4%	72.0%	89.2%
Censor patients with events before 21 days	Relative Risk (Zou SE)	0.600	33.1%	67.6%	87.2%
	GLM with Offset	0.602	33.2%	71.5%	89.0%
	GLM with Offset Zou SE	0.602	33.3%	71.8%	89.1%
	Cox Prop Hazards	0.602	34.4%	72.0%	89.2%

Other group sequential testing strategies

- Instead of using LanDemets spending function with Pocock boundaries one could use
  - LanDemets spending function with O'Brien and Fleming boundaries
  - Logistic spending function,
    - i.e. fix alpha for e.g. IA1 and IA2 and use the remaining alpha at FA
    - Would allow to fix the probability of rejection at given analyses, e.g. if we would like to increase the probability of rejection at IA2

Logistic Spending Function

- This approach allows for more flexibility
- Possible to fix alpha for e.g. 2 out of 3 analyses

Method		Prob Reject IA1	Prob Reject IA2	Prob Reject FA
Lan DeMets O'Brien & Fleming (0.01%, 0.60%, 2.31%)	Relative Risk (Zou SE)	1.0%	58.9%	94.0%
	GLM with Offset Zou SE	1.0%	59.3%	94.2%
	Cox Prop Hazards	1.0%	59.3%	94.1%
Pocock (1.10%,1.10%, 1.10%)	Relative Risk (Zou SE)	32.3%	73.2%	89.8%
	GLM with Offset Zou SE	33.0%	71.9%	89.6%
	Cox Prop Hazards	33.1%	71.9%	89.6%
Logistic Spending Function (0.2%, 1.5%, 1.57%)	Relative Risk (Zou SE)	14.2%	73.3%	90.1%
	GLM with Offset Zou SE	14.2%	73.3%	91.6%
	Cox Prop Hazards	14.2%	73.3%	91.6%
Logistic Spending Function (0.5%, 1.0%, 1.75%)	Relative Risk (Zou SE)	22.4%	66.4%	92.4%
	GLM with Offset Zou SE	22.4%	67.0%	92.4%
	Cox Prop Hazards	22.5%	67.0%	92.4%

Conclusions

- Different methods to obtain an estimate of VE provide similar results
- Adjusting for exposure slightly removes the discreteness
- Logistic spending functions could optimise testing

References

- Cao,B.,Wang,Y.,Wen,D.,Liu,W.,Wang, J., Fan, G.,Ruan, L., Song, B.,Cai, Y., Wei, M. and Li, X. (2020), “A Trial of Lopinavir–Ritonavir in Adults Hospitalized With Severe Covid-19,” *New England Journal of Medicine*, 382, 1787–1799.
- Grein, J., Ohmagari, N., Shin, D., Diaz, G., Asperges, E., Castagna, A., Feldt, T., Green, G., Green, M. L., Lescure, F. X., and Nicastri, E. (2020), “Compassionate Use of Remdesivir for PatientsWith Severe Covid-19,” *New England Journal of Medicine*, 382, 2327–2336.
- O’Kelly, Michael, and Siying Li. "Assessing via Simulation the Operating Characteristics of the WHO Scale for COVID-19 Endpoints." *Statistics in Biopharmaceutical Research* (2020): 1-10.
- World Health Organization (2020a), “WHO R&D Blueprint Novel Coronavirus COVID-19 Therapeutic Trial Synopsis,” available at *https://www.who.int/blueprint/priority-diseases/key-action/COVID-19_Treatment_Trial_Design_Master_Protocol_synopsis_Final_18022020.pdf*

Confidentiality Notice