Volume 27, No. 3 • Fall 2020

BIOPHARMACEUTICAL REPORT

CONTENTS

Note from the editors

In the past months, all aspects of our work and life have been significantly impacted by the ongoing COVID-19 pandemic. Many clinical trials that we are involved in as statisticians are facing unprecedented difficulties in patient recruitment and treatment compliance. Many of you are also closely involved in designing new COVID-19 vaccine trials, treatment trials and master protocols In order to facilitate the understanding of the disease diagnosis, prevention and treatment and increase our ability in tackling the common challenges. As you found in Bruce's message to the BIOP members in the Spring Issue, we invited authors from government, industry and academics to write up their perspectives on the impact of COVID-19 on the ongoing clinical trials, and on the statistical issues of designing new COVID-19 trials. The contributed articles have been reviewed by the BIOP Report editors. We hope this special issue will add to the ongoing conversations on the impact of COVID-19 pandemic and design and analysis of COVID-19 vaccine and treatment studies.

In the first article of the special issue, we reported a survey of COVID-19 related clinical trials based on the data extracted from *clinicaltrials.gov* as of early October. We hope the survey set a stage for the discussion on the current status and developing trend and possible design and statistical analysis issues in the future on developing safe and effective vaccine and treatment for COVID-19. **LJ Wei** (Harvard) and his colleagues summarized some recent interesting work in his group on how to quantify treatment effect and competing risk in COVID-19 clinical trials. **Peter Mesenbrink** (Novartis) shared his experience in designing several COVID-19 trials. **Paul Berg** (Lilly) and his colleagues conducted simulation studies to gain insight about how to choose optimal study endpoints in a proof-of-concept trial. **Sheng Feng** (Parexel) and his colleagues contributed a short article on the challenges and possible solutions of conducting real world evidence studies on COVID-19 treatment. The article by **Yongming Qu** (Lilly) and **Ilya Lipkovich** (Lilly) provided a comprehensive review of estimands and estimation of treatment effect of clinical trials in the period of COVID-19 pandemic in the context of ICH E9.

Meg Gamalo (Pfizer) was recently appointed as the Editor in Chief for *Journal of Biopharmaceutical Statistics* (JBS). As the second installment of the series of interviewing applied statistical journal editors, we asked Meg to share her perspectives in improving the impact of this journal and how to engage the contributing authors and its readers. We hope her responses and the initiatives that she is taking will create sustaining positive impact on *JBS* and our broader community of biopharmaceutical statistics.

In this issue, you will also find a summary report from BIOP Statistical Methods in Oncology Scientific Working Group on the virtual discussion with regulators on type I error considerations in master protocols with common control in oncology clinical trials. Congratulations on the 13 BIOP members who were elected as ASA Fellow in JSM 2020. We would like to recognize their contribution to BIOP and the broader statistical community by publishing their citation. We also publish the minutes of our EC meetings in March and in August at JSM 2020. The minutes are provided by **Janelle Charles** (PPD) and were approved by the EC meetings. We also add a Conference Updates section for the key conferences and meetings sponsored by BIOP. We hope the readers find some useful information in this special issue and stay healthy and safe in this unusual times.

A SNAPSHOT OF COVID-19 STUDIES REPORTED ON *CLINICALTRIALS.GOV*

Steve Wolf (Duke University), Peter Mesenbrink (Novartis), Herbert Pang (Genentech), and Xiaofei Wang (Duke University)
Contact: steven.wolf@duke.edu

Introduction

Since it was first discovered in January 2020 in Wuhan, China, the COVID-19 pandemic has changed most countries of the world in unprecedented levels. Compared to other pandemics in early human history, the advancement of technology and the world's interconnectedness has created a perfect set of circumstances for the elusive and contagious virus to spread to every corner of the world in a very short period of time. Diagnosis of new cases, hospitalizations for moderate and severe cases, human loss from deaths, as well as economic cost to maintain social distancing and shutdown of businesses have created unprecedented negative effect on our modern society. Since January, research organizations, governments and pharmaceutical companies across the world have raced to find fast and accurate diagnostic tests and safe and effective vaccine and treatments for COVID-19.

The goal of this short article is to provide a survey of COVID-19 related clinical trials and studies reported in ClinicalTrials.gov since February of 2020. Two other studies have reviewed the data on clinicaltrials.gov. Both concluded that the number of active studies that could generate credible evidence was low (1-2). This conclusion was drawn by Piovani et al. (2020) using the estimated sample sizes of trials and by Pundi et al. using the Oxford Centre for Evidence-Based Medicine (OCEBM) framework. An important caveat with these studies are their timeframes; Piovani et al. collected data up to April 29th 2020, Pundi et al. collected data between March 2011 and May 19th 2020. In our survey, we collected data between February 1st 2020 and October 6th 2020. Our survey is focused on distribution of clinical trial phase, treatments under investigation, the type of outcomes as well as their temporal changes from January to early October. We are hoping the survey will provide a more accurate and comprehensive snapshot of how the U.S. as well as the world are tackling the COVID-19 pandemic by discovering new diagnostic tests, vaccine and treatments through conducting clinical trials.

Methods

Data was abstracted from Clinicaltrials.gov, maintained by the National Library of Medicine (NLM) at the National Institutes of Health (NIH). This was done by downloading the data on all clinical trials related to COVID-19 in XML format and parsing reported clinical trial information such as start date, type of study, enrollment, treatments, outcomes and phase. We only included trials that primarily studied COVID-19, started between February 1st and October 6th 2020, were Interventional and had a documented phase (phase I, phase I/II, phase II, phase II/III, phase IV).

Treatments and outcomes were reported with significant heterogeneity; we parsed this information using keywords to generate broader treatment and outcome definitions. Treatment categories were anti-malarials (such as hydroxychloroquine and chloroquine), antiviral/retrovirals, anti-microbials (excluding anti-malarial or anti-virals, e.g. remdesivir which is only treatment approved in US or EU), other (traditional Chinese medicine, steroids and stem cell therapy), vaccine, immunomodulators and plasma therapy. Outcomes were defined as adverse events, hospitalization, ICU, ventilation and mortality.

We report descriptive statistics for clinical trial features and plot cumulative change in the number of trials over time based on phase, treatments, outcomes and estimated enrollment. Finally, as clinicaltrials.gov contains most trials conducted in the U.S., the records of the COVID-19 conducted in other countries are underrepresented. We made a map of the prevalence of trials based on location to show the states of the U.S. where COVID-19 is being studied in the clinical trial setting. SAS 9.4 and R 3.6.1 were used for our analysis.

Results

2,872 trials were identified with 1,691 (59%) as interventional. Of these trials, 73.1% (1,236/1,691) had a documented phase. 5.3% (65/1,236) were completed,

57% (705/1,236) were recruiting, 25.7% (318/1,236) were not yet recruiting and 12% (148/1,236) were other (active, not recruiting; available; no longer available; suspended; withdrawn or terminated). 37.7% (466/1,236) expect to enroll over 200 patients, 36.2% (447/1,236) expect to enroll between 51 to 200, 20.1% (248/1,236) expect to enroll between 20 to 50 patients and 6.1% (75/1,236) expect to enroll less than 20 patients. 65.9% (814/1,236) had 1, 19% (235/1,236) had 2 and 15.1% (187/1,236) had 3 or more primary outcomes. More detailed information about these trials can be found in **Table 1**.

	Total (N=1236)
Status of Clinical Trial	
Completed	65 (5.3%)
Recruiting	705 (57.0%)
Not yet recruiting	318 (25.7%)
Other	148 (12.0%)
Estimated Enrollment	
<20 Patients	75 (6.1%)
20-50 Patients	248 (20.1%)
51-200 Patients	447 (36.2%)
201+ Patients	466 (37.7%)
Number of primary outcomes	
1 Primary Outcome	814 (65.9%)
2 Primary Outcomes	235 (19.0%)
3+ Primary Outcomes	187 (15.1%)
Allocation Type	
Randomized	1040 (84.1%)
Non-Randomized	76 (6.1%)
NA	120 (9.7%)
Phase of Trial	
Phase 1	256 (20.7%)
Phase 2	710 (57.4%)
Phase 3	425 (34.4%)
Phase 4	85 (6.9%)
Number of Sites	
0	173 (14.0%)
1 Site	612 (49.5%)
2-5 Sites	226 (18.3%)
6+ Sites	225 (18.2%)
Region	
At least one site in the United States	395 (32.0%)
International Only	841 (68.0%)
Treatments	
Anti-Malarial Drug	196 (15.9%)
Anti-Viral/Retroviral Drug	91 (7.4%)
Anti-Microbial Drug	80 (6.5%)
Immunomodulators	101 (8.2%)
Other (TCM, stem cell, steroid)	104 (8.4%)
Plasma Therapy	128 (10.4%)
Vaccine	78 (6.3%)
Primary Outcomes	
Adverse Event	202 (16.3%)
Hospitalization	73 (5.9%)
ICU	41 (3.3%)
Mortality	283 (22.9%)
Ventilation	148 (12.0%)

Table 1. Characteristics of COVID-19 Clinical Trials Reported in *clinicaltrials.gov*

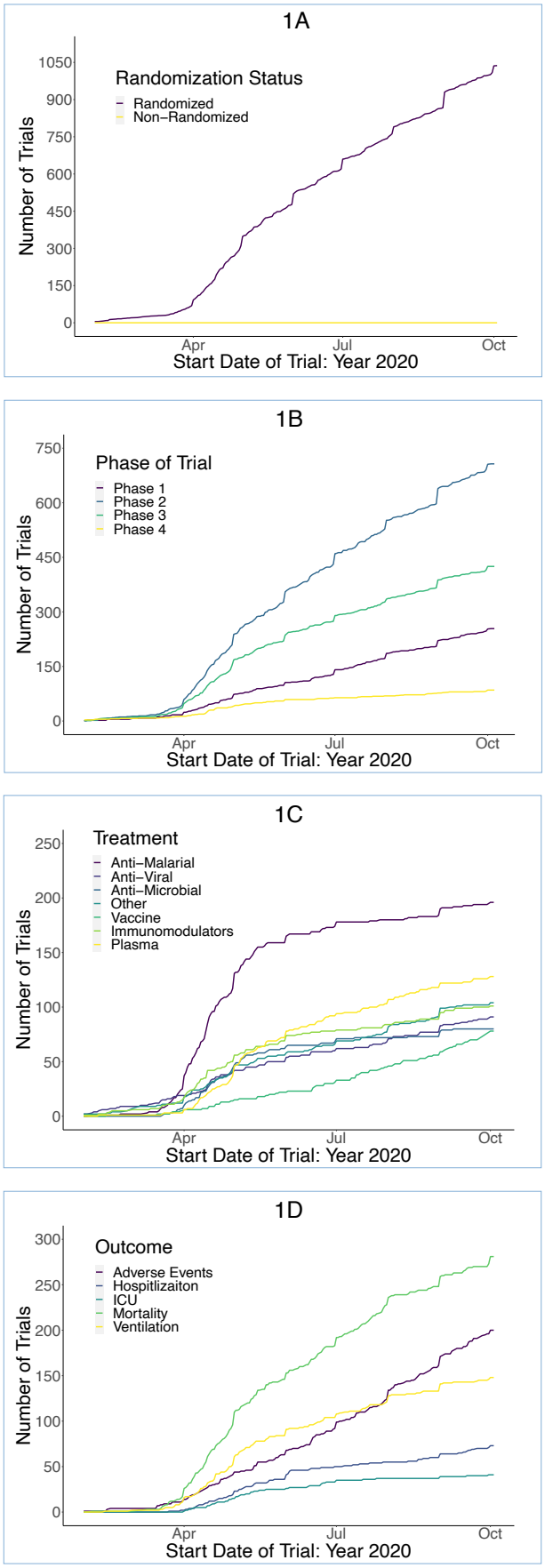


Figure 1. Number of Clinical Trials Reported over Time by Randomization, Phase, Treatment and Outcome

84.1% (1040/1,236) of trials were randomized with phase 2 (57.4%) followed by phase 3 (34.4%; 20.7% phase 1 and 6.9% phase 4) as the most prevalent (**Table 1**). Between February and April, all phases of trials had a similar prevalence, but between April and July, the prevalence of phase 2 and phase 3 trials increased exponentially (**Figure 1B**). 49.5% of trials were single site, with 18.3% between 2 to 5 and 18.2% with 6 or more sites and 32% (395/1,236) of trials had at least one site from the United States. Of the single site trials, 66.8% (409/612) were international trials and of the multi-site trials 57.4% (259/451) were international (**Table 1**). By July, the estimated total enrollment for phase 3 trials was over 200,000 and close to 200,000 for phase 2 trials (**Figure 2**).

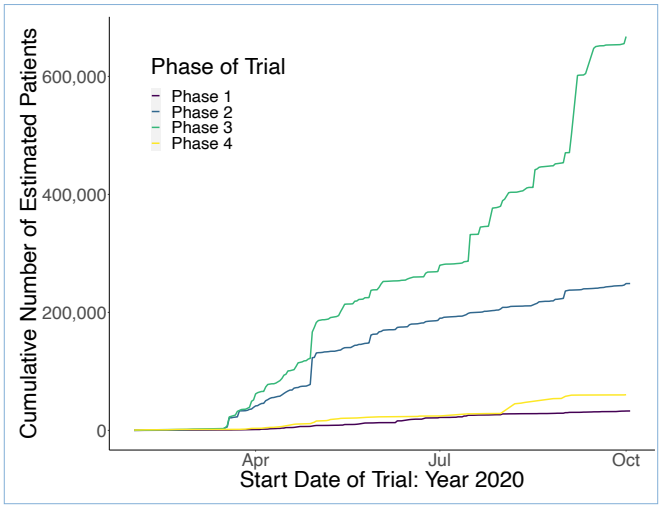


Figure 2. Cumulative Number of Estimated COVID-19 Patients over Time

As seen in **Table 1**, anti-malarials were the most prevalent treatment studied (15.9%), followed by plasma therapy (10.4%), other (8.4%), immunomodulators (8.2%), anti-viral/retroviral (7.4%), anti-microbial (6.5%) and vaccines (6.3%). Of note, between April and July the number of clinical trials studying anti-malarials (such as hydroxychloroquine) jumped up exponentially. After July, the growth of trials studying these types of drugs plateaued (Figure 1C). HCQ and CQ gained media attention in the middle of March.

Mortality (22.9%) was the most prevalent outcome studied followed by adverse events (16.3%), ventilation (12%), hospitalization (5.9%) and ICU (3.3%). Mortality became the most prevalent outcome after April and adverse events became the 2nd most prevalent after July (**Figure 1D**). It should be noted that within these categories endpoints are not necessarily defined consistently which creates an additional challenge in conducting meta-analyses and network meta-analyses from the data reported from these studies.

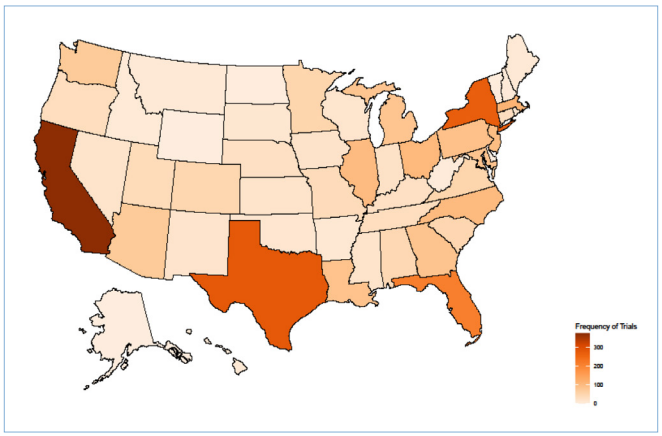


Figure 3. Number of Clinical Trials Available by Country

In the U.S., California, Texas, New York, Florida, Massachusetts, Ohio, North Carolina, Illinois, New Jersey and Pennsylvania have the most sites (**Figure 3**). Globally, the United States followed by Brazil, Spain, France, Russia, Canada, Italy, United Kingdom and Germany have the most registered sites studying COVID-19 on clinicaltrials.gov.

Discussion

Since February 1st of 2020, 1,236 interventional trials have been submitted to *clinicaltrials.gov*. Most were either phase 2 or 3. This indicates that most clinical trials are focused on establishing efficacy on different treatments and comparing this efficacy against a control. We've also observed some interesting trends in treatments being assessed in the clinical trial setting. For example, there was a notable spike in the prevalence of hydroxychloroquine after it received media attention in mid-March.

The largest randomized controlled trial that assessed the efficacy of hydroxychloroquine was conducted by Tang et al and they found no significant difference in viral negative conversion between standard of care and standard of care plus HCQ (3-4). Additionally, they found the cohort that received HCQ had higher adverse events. This data, observations of serious adverse

events associated with HCQ and the lack of replication of decreased viral shedding resulted in the FDA revoking their EUA for HCQ and CQ (3). Despite this early data, the safety and efficacy still needs to be established through completed phase 2 and 3 trials. HCQ and CQ may have a benefit as a preventative or adjuvant treatment, but this may not be established since both the RECOVERY and SOLIDARITY trials have discontinued treatments arms for hydroxychloroquine (5-6). This resulted from both trials finding no evidence of clinical benefit for the drug during interim analyses. Novartis also decided to discontinue their HCQ sponsored trial due to acute enrollment challenges (7). These results are consistent with our observation that the growth of trials studying anti-malarials plateaued after July.

Additionally, based on preliminary data from the ACTT study conducted by NIAID, Remdesivir has been approved by the FDA on October 22nd for patients 12 years and older who are hospitalized with COVID-19 disease (8-9). It has also been approved as treatment for patients with severe COVID-19 in Japan, Taiwan, India, Singapore, the United Arab Emirates and the European Union (10). This has led to most of the new platform trials that have started since then to focus more on the evaluation of different treatments in combination with remdesivir (e.g. ISPY-Covid ACTT-2, and ACTT-3). Remdesivir is being studied in 1.9% of trials in our survey.

As of October 2nd, there were 193 vaccines being developed, 42 in clinical evaluation and 151 in preclinical evaluation (11-12). One of these vaccines, an Ad5 based vaccine, has been approved for limited use in China for state-owned company employees and military personnel (13). Russia has also approved a vaccine for use, despite not completing a phase III trial (14). The results of these early approvals could prove to be instrumental in developing an effective treatment for COVID-19. It is important to note that none of these early approval drugs have completed a phase III trial which is considered a hallmark to demonstrate the efficacy of a treatment. Until one is completed, we will not have compelling evidence of an effective treatment against COVID-19. Based on our survey, Ad5 is being studied in 0.9% of trials.

It is reassuring that both the United States and Brazil contain the most included sites in clinical trials. These two countries, according to data from Johns Hopkins, are two of the top three countries with the most reported cases (15). As of October 8th, The U.S. has had over 7.5 million cases and Brazil has had more than five million cases. It's important to study interventions in areas where the pandemic affects the most not only for the potential benefit to participants but also to have sufficient data to assess efficacy. More clinical trial sites should launch in India given they have had over 6.7 million confirmed cases. Additionally, as of October 8th, 7 out of 10 of the states with the most sites in clinical trials are also in the top reported COVID cases in the United States. These include California, Texas, Florida, New York, Illinois, North Carolina and New Jersey (16). As the number of cases have exponentially grown, so have the number of clinical trials studying this deadly disease.

This study provides a snapshot of clinically trials currently evaluating treatments for COVID-19. We were able to describe the number of trials by phases, treatments and outcomes longitudinally. This provides information on our progression towards finding an effective treatment. Despite these important findings, our study had notable limitations. First, we only reviewed trials submitted to *clinicaltrials.gov*. It's important to note that only drug, device or biological studies are required to report to the website. Second, trials are not consistently updated on the website. Finally, information reported in treatment arms and outcomes were not uniform and we had to create our own definitions by searching for keywords. This may result in us underestimating the prevalence of treatments and outcomes.

Conclusion

Our study indicates that most clinical trials are currently either in phase 2 or 3, suggesting a focus on establishing efficacy for different treatments and comparing this efficacy to controls. Anti-malarials are the most commonly examined treatment with mortality, adverse events and ventilation as the most common outcomes. There has been a significant increase to the number of trials as the number of COVID-19 cases have continued to grow.

Acknowledgements

We would like to acknowledge the contribution from Siqi Zhang, Zhengxi Chen, Kenny Zhang and Yinglu Zhou who helped develop the program to abstract clinical trial features off the *clinicaltrials.gov* website.

References

1. Pundi K et al. Characteristics and Strength of Evidence of COVID-19 Studies Registered on ClinicalTrials.gov. July 27, 2020. doi:10.1001/jamainternmed.2020.2904

2. Piovani, D., Pansieri, C., Peyrin-Biroulet, L., Danese, S., & Bonovas, S. (2020). A snapshot of the ongoing clinical research on COVID-19. F1000Research, 9

3. https://www.fda.gov/drugs/drug-safety-and-availability/fda-cautions-against-use-hydroxychloroquine-or-chloroquine-covid-19-outside-hospital-setting-or

4. Hydroxychloroquine in patients with mainly mild to moderate coronavirus disease 2019: open label, randomised controlled trial. BMJ 2020;369:m1849 (Published 14 May 2020)

5. https://www.recoverytrial.net/news/statement-from-the-chief-investigators-of-the-randomised-evaluation-of-covid-19-therapy-recovery-trial-on-hydroxychloroquine-5-june-2020-no-clinical-benefit-from-use-of-hydroxychloroquine-in-hospitalised-patients-with-covid-19

6. https://www.who.int/news-room/detail/04-07-2020-who-discontinues-hydroxychloroquine-and-lopinavir-ritonavir-treatment-arms-for-covid-19

7. https://www.novartis.com/news/media-releases/novartis-discontinues-hydroxychloroquine-clinical-trial-based-slow-enrollment-remains-committed-pandemic-research-efforts

8. https://www.accessdata.fda.gov/drugsatfda_docs/label/2020/214787Orig1s000lbl.pdf

9. Beigel JH, et al. Remdesivir for the Treatment of Covid-19 — Preliminary Report. NEJM; May 22, 2020 DOI: 10.1056/NEJMoa2007764.

10. https://www.gilead.com/news-and-press/press-room/press-releases/2020/7/european-commission-grants-conditional-marketing-authorization-for-gileads-veklury-remdesivir-for-the-treatment-of-covid19

11. https://www.nytimes.com/interactive/2020/science/coronavirus-vaccine-tracker.html

12. https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines

13. https://www.nytimes.com/2020/07/16/business/china-vaccine-coronavirus.html

14. https://www.nytimes.com/2020/08/11/world/coronavirus-covid-19.html

15. https://coronavirus.jhu.edu/data/new-cases

16. https://en.wikipedia.org/wiki/Template:COVID-19_pandemic_data/United_States_medical_cases ■

WHAT WE LEARNED FROM RECENT COVID-19 CLINICAL STUDIES REGARDING STATISTICAL METHODOLOGY

Zachary R. McCaw (Google), Lu Tian (Department of Biomedical Data Science, Stanford University), and Lee-Jen Wei (Harvard T.H. Chan School of Public Health)

Contact: zmccaw@alumni.harvard.edu | lutian@stanford.edu | wei@hsph.harvard.edu

A few years ago, Professor Donald A. Berry said at a public meeting "Clinical trial methodology research is led by colleagues in the pharmaceutical industry and regulatory agencies, not by academicians anymore." Unlike the past, with the recent focus on bioinformatics and genomics, there is currently less industry and government funding to support academic research into clinical trial methodology. Recently, Professor Berry also noted that "COVID-19 has brought people's attention to clinical trial methodology." All of the recently published COVID-19 clinical studies, whether observational or randomized controlled trials, have drawn tremendous interest, as well as concern about their validity. Moreover, the general public deserves transparent, intuitive, and clinically meaningful interpretations of the study results. It is our responsibility to make this possible. It is important to translate the results from statistical analysis of clinical studies into language understandable to clinicians and the general public as a whole. This is part of so-called "translational statistics," like translational medicine, which we have recently tried to promote (McCaw et al., 2018). We take this opportunity to share what we learned from COVID-19 randomized clinical trials.

In designing a clinical study to explore whether a new treatment is better than standard care, trialists need to determine the study population, define the endpoints, and select the summary measures that will be used to quantify the treatment effect (ideally including both efficacy and harm together). These are 3 of the 5 components that form the "estimand" described in the ICH-E9(R1) (ICH E9[R1] Expert Working Group, 2019). Here, we discuss issues related to the selection of study endpoints and the quantification of the between-group difference for randomized clinical studies of COVID-19.

In response to COVID-19, the US National Institutes of Health (NIH) has conducted and set in motion various "Adaptive COVID-19 Treatment Trials" (ACTT) to investigate the safety and efficacy of new therapeutic agents. For example, ACTT-1 compared Remdesivir with standard care (Beigel et al., 2020), and the recently launched ACTT-3 will compare Remdesivir plus Interferon Beta-1a with Remdesivir alone among approximately 1000 patients hospitalized COVID-19 (National Institute of Allergy and Infectious Diseases [NIAID], 2020). For these trials, the primary endpoint is time to recovery, and among the numerous secondary endpoints are overall survival and the duration of hospitalization. Let us first consider the primary endpoint, focusing on ACTT-1 as an example. For this study, clinical status was assessed daily during hospitalization and categorized on an 8-point ordinal scale, with category 1 being the most favorable outcome (discharge from hospital with no limitation of activities), and category 8 being death. Recovery was defined as the first day, during the 28 days of follow-up, where the patient reached clinical categories 1, 2, or 3, defined as 1) not hospitalized and with no limitations on activities; 2) not hospitalized, but with limitations on activities and/or requiring home oxygen; and 3) hospitalized, but no longer requiring supplemental oxygen or ongoing medical care. Like other critical care studies, this trial was short in duration and needed to account for the competing risk of death, since the mortality rate was non-negligible, at 12 to 15%. Since the time to recovery of a patient who had died before day 28 was not well defined (technically infinite), the study team assigned a censored value of 29 days to these patients. The investigators then adopted standard survival analysis techniques to analyze the time to recovery, reporting a hazard or "rate ratio" to compare the two arms (Beigel et al., 2020). The observed rate ratio (Remdesivir vs Placebo) of 1.32 (95% CI 1.12 to 1.55; $P<0.001$) was highly statistically significant in favor of Remdesivir. However, in the presence of a competing risk from death, the 32% increase in the rate of recovery with Remdesivir is difficult to interpret. For instance, this 32% increase does not mean that patients receiving Remdesivir were 32% more likely to recover because rate is not a probability measure like risk. Moreover, as a relative measure of treatment

effect, a rate ratio alone lacks context without a reference value for the recovery rate from the Placebo arm. The difficulty of interpreting the hazard or rate ratio in the presence of competing risks has been discussed by Fine and Gray (Fine and Gray, 1999) and Zhao et al (Zhao et al., 2018). Specifically, since the time to recovery is not a proper random variable, the distribution function for the time to recovery never reaches one as time increases. The hazard ratio proposed by Fine and Gray is based on two sub-distributions and is even more difficult to interpret than its counterpart when no competing risks are present.

In a recently published paper in the *Annals of Internal Medicine (McCaw el al., 2020)*, and a correspondence to the *New England Journal of Medicine (McCaw et al., 2020)*, we discussed this translational issue and presented an alternative measure, which is more intuitive and clinically meaningful. Here, we summarize the results from these two publications. Since we did not have the original data from ACTT-1 (Beigel et al., 2020), we reconstructed the individual patient-level time-to-event data by scanning (Guyot et al., 2012) the cumulative recovery and mortality rate curves from the published article. We then estimated the cumulative incidence curves reproduced here in **Figure 1 (A)** (McCaw et al, 2020). The Remdesivir curve (blue) is always above the Placebo curve (red), suggesting that Remdesivir was better than Placebo, at least numerically, for hastening the time to recovery. The resulting rate ratio, estimated from the cumulative incidence curves while accounting for the competing risk of death, was 1.30 (95% CI, 1.11 to 1.51; P=0.001), very similar to that reported by ACTT-1. The median recovery times were 11 and 15 days for Remdesivir and Placebo, identical to those reported in the publication. However, no formal comparison of

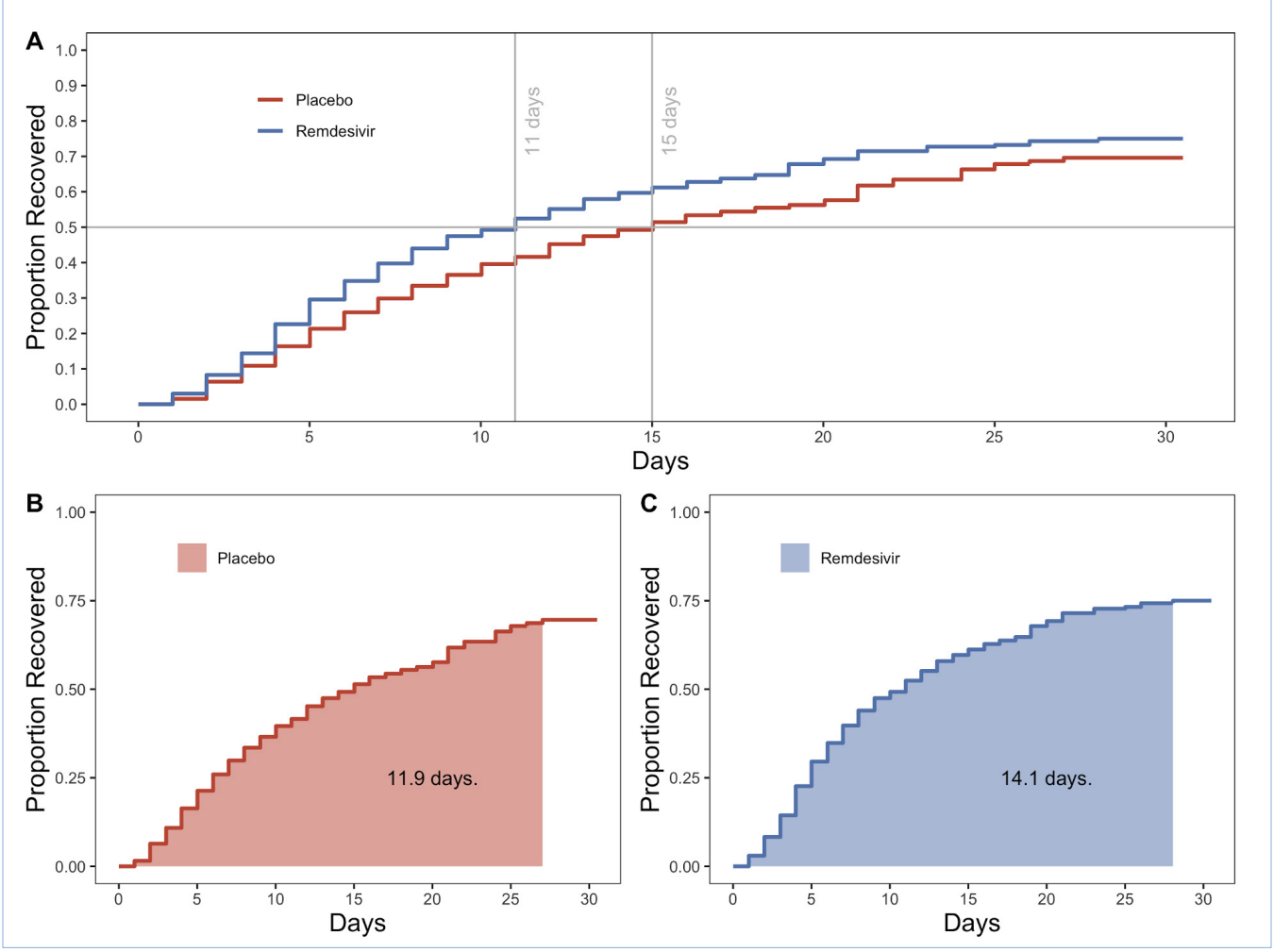


Figure 1. (A) Cumulative Incidence Curves from ACTT-1 for the Proportion of Patients Recovered, treating Death as a Competing Risk. (B & C) Mean Time in Recovery, as the Area Under the CIC, across the 28 Days of Study Follow-up. (McCaw et al., 2020; Used with permission. © 2020 American College of Physicians.)

medians was reported for ACTT-1. Using reconstructed data, the difference of medians was 4 days (95% CI: 1.0 to 7.0, P = 0.003). The 95% confidence interval is quite large, reflecting the instability of the estimated median recovery time. Moreover, the median is only a local measure of the recovery rate curve in **Figure 1 (A)** and cannot capture the entire profile. It is important to note that due to the competing risk from death, the mean time to recovery is not well-defined and cannot be estimated (the distribution of time to recovery has a point mass at infinity). In addition, if the recovery rates had been below 50% on day 28, then the median recovery time would not have been empirically estimable.

The question to be answered is how to quantify the difference between the two cumulative recovery curves in Figure 1 in a manner that is interpretable, stable, and takes the entire recovery rate curve into account. Since recovery is a desirable outcome, the higher the curve in **Figure 1(A)**, the better the treatment. Intuitively then, a better treatment will have a larger area under the curve. Furthermore, the area under the curve across the 28 days of follow-up has a straightforward interpretation as the mean time span post-recovery. That is, the expected time patients spent having recovered by day 28; the longer, the better. For Remdesivir and Placebo, the post-recovery time spans were 14.1 and 11.9 days, which are presented graphically in **Figures 1(B)** & **(C)**. The difference of 2.2 days (95% CI, 0.89 to 3.52, P<0.001) significantly favored Remdesivir. That is, patients hospitalized with COVID-19 that were randomized to Remdesivir and followed for 28-days spent 2.2 days longer, on average, having recovered. This time-scale summary of the treatment effect is more translatable than the roughly 30% increase in the rate of recovery from Remdesivir. We encourage the investigators of future COVID-19 trials to report this alternative metric when quantifying the treatment effect.

For a key secondary endpoint, the overall survival, ACTT-1 again summarized the treatment effect using a hazard ratio, which was 0.70 (95% CI, 0.47 to 1.04; P = 0.07), comparing Remdesivir with Placebo. Unlike the recovery time analysis, for short-term studies in critical care medicine, the absolute mortality benefit at the end of follow-up seems more relevant than a reduction in the time-to-event. Using reconstructed data, the 28-day overall survival rates were 88% and 85% for Remdesivir and Placebo.5 The absolute difference of 3.1% (95% CI, -2.2% to 8.3%, P = 0.25) did not differ significantly between the two arms. The lesson here is that the hazard ratio may not be an appropriate measure for evaluating patients' survival in short-term comparative studies; that is, absolute survival at the end of a short-term study is more clinically relevant than the relative hazard rate, which instead reflects how much the new treatment reduces hazard during the study period compared with the control.

There are 33 secondary endpoints listed under ACTT-3 (NIAID, 2020). Let us consider those secondary endpoints that involve the duration of certain events, such as the duration of hospital stay or the duration of supplemental oxygen. Direct comparison of the two arms with respect to these endpoints may not be an appropriate measure of efficacy in the presence of death. A treatment that is more effective at preventing death may prolong a patient's stay in hospital or duration on supplemental oxygen, extending an apparently undesirable outcome, and potentially resulting in misleading conclusions. A straightforward way to handle the presence of death would be to compare, for example, hospital-free survival times rather than the durations of hospital stay. This complementary endpoint would account for the potential imbalance in survival rates between the treatment and control arms.

Note that most published COVID-19 trials have based their endpoints on a 6- to 8-point ordinal outcome. For example, ACTT-1 classified clinical status using an 8-point scale, spanning from discharged from hospital with no limitation of activities to deceased, with increasing requirements for medical attention in between (Beigel et al., 2020). To increase statistical power, some COVID-19 trials have proposed an ordinal logistic model and its common odds ratio to quantify the treatment difference. However, in at least two major studies (Goldman et al., 2020; Spinner et al., 2020), this model's proportional odds assumption was not met, and the reported analysis was mainly based on p-values, which have no clinical meaning. The lesson we learned is that the prespecified summary measure of treatment efficacy should not depend on modeling assumptions: if the model does not fit the data at the end of the trial, it is unclear how to interpret the treatment difference.

Last though not least, the conventional approach to analyzing the data would be to summarize the treatment difference for each of the many endpoints separately, resulting in 34 separate summaries at the end of the trial. However, these separate analyses provide a disconnected view of safety and efficacy because we do not know whether the outcomes were correlated at the individual patient-level. Moreover, these separate

summaries do not reflect the way clinicians approach treatment selection in practice, by jointly weighing risks and benefits for individual patients. Instead, we need a global composite outcome including efficacy and safety events at the individual patient-level. Leveraging the patient's baseline information, this composite endpoint can also help us to identify "high value" subgroups of patients who would benefit most from the new treatment (Claggette et al., 2015; Angelidou et al., 2018). There are many other interesting and challenging issues that we will learn more about from on-going and future studies of COVID-19.

References

1. McCaw, Z. R., Orkaby, A. R., Wei, L. J., Kim, D. H., & Rich, M. W. (2019). Applying Evidence-Based Medicine to Shared Decision Making: Value of Restricted MeanSurvival Time. *Am J Med*, 132(1), 13-15. https://www.amjmed.com/article/S0002-9343(18)30750-2/fulltext.

2. E9(R1) EWG Addendum: Statistical Principles for Clinical Trials. https://www.ich.org/page/efficacy-guidelines#9-2.

3. Beigel JH, Tomashek KM, Dodd LE, et al. Remdesivir for the treatment of Covid-19 - A preliminary report. *NEJM*. 2020 May 22; doi: 10.1056/NEJMoa2007764.

4. Adaptive COVID-19 Treatment Trial 3 (ACTT-3). https://clinicaltrials.gov/ct2/show/NCT04492475.

5. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *J AmStat Assoc*. 1999; 94(446):496-509. doi:10.1080/01621459.1999.10474144.

6. Zhao L, Tian L, Claggett B, et al. Estimating Treatment Effect With Clinical Interpretation From a Comparative Clinical Trial With an End Point Subject to Competing Risks. *JAMA Cardiol*. 2018 Apr 1; 3(4):357-358. doi: 10.1001/jamacardio.2018.0127.

7. McCaw, Z. R., Tian, L., Vassy, J. L., Ritchie, C. S., Lee, C. C., Kim, D. H., & Wei, L. J. (2020). How to Quantify and Interpret Treatment Effects in Comparative Clinical Studies of COVID-19. *Ann Intern Med*. 2020; M20-4044. doi: 10.7326/M20-4044. https://www.acpjournals.org/doi/10.7326/M20-4044.

8. McCaw ZR, Kim DH, Wei LJ. Remdesivir for the Treatment of Covid-19 - Preliminary Report. NEJM. 2020; 383(10):993. https://www.nejm.org/doi/full/10.1056/NEJMc2022236.

9. Guyot P, Ades AE, Ouwens MJ, et. al. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Med Res Methodol*. 2012; 12:9.

10. Goldman JD, Lye DCB, Hui DS, et al. Remdesivir for 5 or 10 Days in Patients with Severe Covid-19. NEJM. 2020 May 27; doi: 10.1056/NEJMoa2015301.

11. Spinner CD, Gottlieb RL, Criner GJ et al. Effect of Remdesivir vs Standard Care on Clinical Status at 11 Days in Patients With Moderate COVID-19: A Randomized Clinical Trial. *JAMA*. 2020;e2016349. doi: 10.1001/jama.2020.16349.

12. Claggett B, Tian L, Castagno D, Wei LJ. Treatment selections using risk-benefit profiles based on data from comparative randomized clinical trials with multiple endpoints. *Biostatistics*. 2015; 16(1):60-72.

13. Angelidou K, Palumbo P, Lindsey J, et.al. International Maternal Pediatric Adolescent AIDS Clinical Trials Group, P1060 Study Team. Defining Study Outcomes That Better Reflect Individual Response to Treatment. *Pediatr Infect Dis J*. 2018; 37(3):258-62.

THE CHALLENGES AND COMPLEXITIES IN DESIGNING CLINICAL TRIALS TO DEVELOP A SAFE AND EFFECTIVE TREATMENT FOR COVID-19

Peter Messenbrink (Novartis)

1. Introduction

The year 2020 has been an epic year remembered for generations with the emergence of a global pandemic caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) virus. It has brought great opportunities throughout the scientific community and the pharmaceutical industry to find effective treatments to help in the recovery of patients who are infected with the virus and to develop vaccines to help prevent the global population from becoming infected and potentially having a fatal or long-term debilitating outcome due to the virus in the future.

Anytime one enters into drug development in a new disease area, there will be challenges and obstacles that will create obstacles for us as statisticians in the design, execution, and analysis of clinical trials to demonstrate that treatments are safe and effective to treat the new condition. These challenges become over accentuated during a pandemic when everyone is in competition to design the best trial to answer clinical hypotheses in the shortest time possible especially when the pandemic started and many of us did not have a good idea how best to design such trials to answers these questions with a high degree of confidence.

Throughout the course of the pandemic, not only has the patient population continued to change; but also, the standard of care has continued, making the determination of the target population, best endpoint to determine if a treatment is effective, and the number of required patients to demonstrate effectiveness an incredibly challenging task. I will discuss some of these issues.

2. Determining the right patient population

When the pandemic started and many of the hospitals in this country were overflowing with patients, the focus was on studying those patients hospitalized due to their COVID-19. This may seem very straightforward to anyone who has not been actively involved in the development of treatments for COVID-19, but has been actually a major source of the problem in finding the best treatment for patients hospitalized due to this illness. As part of the patient screening process for clinical study eligibility, patients are administered a seven- or eight-point ordinal scale (Beigel 2020). Within this ordinal scale, hospitalized patients were in one of four categories ranging from a baseline score of 4 to 7. These categories are:

4. Hospitalized, not requiring supplemental oxygen, requiring ongoing medical care (COVID-19-related or otherwise)
5. Hospitalized, requiring supplemental oxygen
6. Hospitalized, receiving non-invasive ventilation or high-flow oxygen devices
7. Hospitalized, receiving invasive mechanical ventilation or extracorporeal membrane oxygenation (ECMO)

These categories represent a very heterogeneous patient population with highly variable degrees of baseline risk of mortality and morbidity. Thus, effectiveness with a single therapeutic agent has not been demonstrated across all levels of disease severity, which brings into question the robustness of the efficacy shown by these treatments. Two treatments have demonstrated effectiveness and published their study results: remdesivir (Beigel et. al., 2020) and dexamethasone (RECOVERY group, 2020). In the remdesivir trial, the primary efficacy outcome measure was time to recovery during the 28 days after enrollment on which a patient reaches categories 1, 2, or 3 on the eight-category ordinal scale for which the categories are as follows:

1. Not hospitalized, no limitations of activities
2. Not hospitalized, limitation of activities, home oxygen requirement, or both
3. Hospitalized, not requiring supplemental oxygen and no longer requiring ongoing medical care

In the RECOVERY clinical study cited above that evaluated dexamethasone, the primary efficacy outcome was all-cause mortality within 28 days after randomization. Both studies used a rate ratio as their summary measure to determine if effectiveness.

Both clinical trials, met their primary efficacy endpoint. In the remdesivir trial (N=1059), the rate ratio for recovery was 1.32 (95% CI, 1.12 to 1.55) in favor of remdesivir. However, the effectiveness was not consistent across baseline disease severity as indicated in **Table 1** below.

Ordinal Scale at Baseline	Remdesivir	Placebo	Rate ratio (95% CI)
4	61/67	47/60	1.38 (0.94, 2.03)
5	177/222	128/199	1.47 (1.17, 1.84)
6	47/98	43/99	1.20 (0.79, 1.81)
7	45/125	51/147	0.95 (0.64, 1.42)
OVERALL	334/538	273/521	1.32 (1.12-1.55)

Table 1. Recovery rate at Day 28 by baseline disease severity

Patients receiving no supplemental oxygen or receiving low flow supplemental achieved recovery rates higher than the overall treatment effect compared while those receiving mechanical ventilation or ECMO who actually had a lower recovery rate numerically relative to standard of care. These patterns were consistent with the 28-day mortality rate comparisons across baseline disease severity where those with baseline ordinal scale values of six or seven were more likely to die on remdesivir than with standard of care (both hazard ratios were greater than one, indicating an increase in the risk of death within those subgroups).

In the RECOVERY trial which is a multi-treatment platform trial to evaluate several treatments in hospitalized patients in the United Kingdom, 482/2104 (22.9%) of patients receiving dexamethasone died over 28 days of follow-up from randomization compared to 111/4321 (25.7%) of patients receiving standard of care which corresponded to a rate ratio of 0.83 (95% CI, 0.75- 0.93). In this study, the results were very different depending on the level of respiratory support received. For those patients not receiving any oxygen, the rate ratio for 28-day mortality was 1.19 (95% CI, 0.91-1.55) indicating an increase in the risk of death for those treated with dexamethasone. In contrast to this, for those receiving invasive mechanical ventilation, the rate ratio was 0.64 (95% CI, 0.51-0.81) and for those receiving oxygen only the rate ratio was 0.82 (95% CI, 0.72-0.94).

What this tells us is that the virus behaves vary differently in different patient populations. With remdesivir, the recovery was best in those who were less severe and still hospitalized because potentially the inflammation in the tissue and the organs that has been observed with many COVID-19 patients has not occurred yet and with those patients, an anti-viral like remdesivir has a reasonable chance of being effective in treating such patients. In contrast to this a potent steroid like dexamethasone, works best in the most severe patients who have undergone a reaction known as cytokine release syndrome which if not treated can often lead to multi-organ failure prior to death. Thus, by reducing the inflammation, even those in the most morbid state prior to death have a chance of recovery.

It is clear that monotherapy will not likely be most effective in treating such patients, which is why many of the clinical studies ongoing are looking combination therapies often combining different classes of drugs with the anti-viral treatment remdesivir to treat all of the dimensions of the disease.

This leads asking the difficult question within the estimand framework, what is the treatment of interest?

3. What is the treatment of interest and how can you define intercurrent events?

In October 2020, there are no approved treatments for the SARS-CoV-2 virus by the Food and Drug Administration (FDA). At the start of the pandemic when designing clinical trials to treat COVID-19, there was much discussion on what was the standard of care. The answer provided often by physicians, who are involved in the treatment of COVID-19 patients, anything that will provide benefits to the patients and help in their recovery will be tried. Under these considerations any concomitant treatments or procedures used in treating the patient to help them recover from COVID-19 is part of the standard of care. This implies that it is theoretically possible that a patient could receive two, three, or more additional treatments, if they are accessible, that are currently being evaluated as possible COVID-19 treatments as part of the standard care group in the clinical trial in which they are participating. Some clinical trials were designed to estimate the treatment effect with minimal bias by excluding anti-viral treatments as part of the standard of care when an anti-viral treatment was the experimental treatment being evaluated in the clinical trial. This paradigm changed when the FDA issued an emergency use authorization for remdesivir, making it the standard of care for hospitalized patients where it is available. The challenge that has existed is that remdesivir has become a standard of care in hospitalized patients where it is available. However, it not always available in all regions of the world where COVID-19 clinical trials are being conducted in hospitalized patients. This has the potential to introduce substantial differences in the response rate within the standard of care group across geographical regions.

Given that there is no approved treatments, even when new concomitant treatments are added, such changes in the standard of care of for the patients would be handled as "treatment policy" if such changes would be considered as intercurrent events and thus include as part of the treatment of interest definition. Efforts continue towards finding the best treatment(s) for the different severity levels of those infected with the SARS-CoV-2 virus. Increased efforts will focus on finding the best combination of treatments to eradicate the virus and its symptoms and the complications that may ensue as the severity of the virus worsens or as patients try to return to normalcy when the virus is no longer detectable when tested. Improved treatment options for patients confirmed in the clinical trial setting are hopefully on the near-term horizon. With this, the standard of care as part of the treatment of interest will continue to evolve and increase the importance of the use of adaptive platform trials that will allow for open entry of new potential treatments, as well as adaptations to the control arms and the exiting of ineffective treatments.

4. What is the best endpoint for demonstrating effectiveness of COVID-19?

In most of the randomized controlled trials for COVID-19 treatments, the endpoints evaluated in these trials are derived in some way from the WHO ordinal scale that has been developed for use in COVID-19 clinical trials. The problem that exists is almost every implementation of this ordinal scale deviates from the 9-point scale that was developed for use (WHO 2020). This creates a different problem in an area that has been perfectly setup for ongoing meta-analyses to be performed, the question becomes are the endpoints across the different clinical trials actually comparable? A detailed summary of the endpoints derived from the WHO disease severity ordinal scale is provided in O'Kelly and Li (2020) along with the operating characteristics of those endpoints in different settings.

All of the endpoints derived from the WHO ordinal scale have strengths and weaknesses. Time to recovery has been used the most frequently as the primary endpoint across clinical trials followed by 28-day mortality. However, as mortality rates decrease over time with those who have been infected, the size of a trial required to demonstrate a statistically significant reduction in mortality becomes unachievable. Note, even in RECOVERY, over 6000 patients were required (randomized in a 1:2 ratio) to demonstrate a mortality improvement for dexamethasone when the 28-day death rates were greater than 20% in both the experimental treatment and control arms. In the remdesivir results reported by Beigel et. al. (2020), the original endpoint based on improvement in the ordinal scale was adapted to time to recovery after an early interim analysis, which also noted that the sample size needed to be increased from a sample size of around 450 patients to a sample size of over 1000 patients.

The challenges faced by all of these endpoints is that they do not capture the patient journey to their clinical outcome (recovery free of virus or death) often it is plausible for a patient to experience both improvement and worsening of their disease state over time. Lin et. al. (2020) proposed averaging the disease states in which a patient resides over time taking an area under curve approach to evaluate differences in the time spent in different disease states across patients. This would appear to be far more sensitive to change and can detect meaningful differences potentially with fewer patients and requires further examination in many of the platform trials and master protocols that are currently being executed to evaluate a wide range of different types of COVID-19 treatments.

As the standard of care improves, the demand for doing more outpatient trials will continue to increase. In this patient population, the current FDA recommendations for registration (FDA 2020) are to demonstrate faster recovery from the symptoms of COVID-19. However, as the number of patients being hospitalized remains lower than at the start of the pandemic, is it more important for those treated in the outpatient setting to prevent hospitalization for their recovery? Evaluation through meta-analysis across studies will be necessary to evaluate what are the best endpoint(s) in this patient population.

5. Discussion

Since the start of 2020, the development of treatments for COVID-19 has traveled a long journey and faced many obstacles along the way. Many companies have tried to repurpose medicines that were being used to treat other conditions with the hope that either they could eradicate the virus or reduce the cytokine activity experiences by the more severely ill patients. The results of these efforts have been mixed, with failures outnumbering the successes.

There are many reasons why these clinical trials have been less than successful. The causes of failure are not necessarily always the fault of the Sponsors. This may be because it has been almost impossible to keep up with the dynamic environment over which the SARS-CoV-2 virus has mutated over time. In addition, the patients infected with the virus have also evolved in their demographic and baseline characteristics, as have the medical knowledge of the pathophysiology and ability to treat the patients and their symptoms, which has increased the probability of eventual recovery.

The race to find highly effective treatments is clearly a marathon and not a sprint to the finish line. All who are involved in this effort share the passion to do all that is possible to find the best trial design, patient population, and endpoint to test their treatment of interest to demonstrate that an effective and safe treatment can be developed and made available to save lives around the world as vaccine development continues in parallel. As we continue to learn from the data gathered across clinical trials, meta-analysis and network meta-analysis will be conducted in earnest to determine which treatments are the most appropriate to treat the millions who will become infected by the virus in the coming months.

References

1. Beigel, JH, Tomashek, KM, Dodd LE, et al (2020). Remdesivir for the treatment of Covid-19—preliminary report. *New Eng J Med*, May 22 (DOI: 10.1056/NEJ- Moa2007764).

2. RECOVERY Collaborative Group (2020), "Dexamethasone in Hospitalized Patients with Covid-19 - Preliminary Report. *N Engl J Med*. 2020 Jul 17:NEJMoa2021436. doi: 10.1056/NEJMoa2021436. Epub ahead of print. PMID: 32678530; PMCID: PMC7383595.

3. World Health Organization (2020), "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*

4. Michael O'Kelly & Siying Li (2020) Assessing via Simulation the Operating Characteristics of the WHO Scale for COVID-19 Endpoints, *Statistics in Biopharmaceutical Research*, DOI: 10.1080/19466315.2020.1811148.

5. Food and Drug Administration (FDA) (2020), "COVID-19: Developing Drugs and Biological Products for Treatment or Prevention," available at *https://www.fda.gov/media/137926/download*. ■

BRIEF REPORT ON SIMULATIONS FOR A PROOF-OF-CONCEPT STUDY OF A POTENTIAL TREATMENT FOR COVID-19

Mitchell A Thomann (Eli Lilly), Michael G Durante (Eli Lilly), and Paul H Berg (Eli Lilly)

Introduction

The most severely affected patients with COVID-19 can progress to Acute Respiratory Distress Syndrome (ARDS), which is often treated with intermittent mechanical ventilation (IMV). Past studies of ARDS have used ventilator-free days (VFD) over a 28-day follow-up period, with patients who die during the 28-day period being assigned a value of -1, as the primary efficacy measure (Bellingan, et al., 2017; Ranieri et al., 2020). This VFD score, an ordinal endpoint ranging from -1 (patient died) to 28 (patient survived without ever going on to IMV) is analyzed via the Wilcoxon Rank-Sum test, employing the Van Elteren method to account for any relevant stratification variables (Finkelstein and Schoenfeld, 1999). For a small proof-of-concept study, the VFD outcome may be preferred over simpler outcome measurements, for example 28-day mortality rate or 28-day IMV progression rate, because it allows for detecting a treatment effect for multiple dimensions of ARDS (IMV progression, time on IMV and mortality). While all may be potentially clinically meaningful, reliance on a single dimension for the primary analysis may lead to underestimated totality of benefit.

The clinical trial being simulated in this report has 200 patients randomized 1:1 in a double-blind fashion to placebo or experimental treatment, either of which is added to standard of care. The experimental treatment is hypothesized to reduce the rate of progression to IMV, mortality, the time on IMV or a combination of those factors. The study population is proposed to be U.S. patients who are hospitalized with pneumonia and presumed or confirmed COVID-19. In this report we describe the simulations used to determine the operating characteristics of such a phase 2 proof-of-concept study.

Simulation Methods

To generate the endpoints in this simulation, a simplifying assumption is used that a patient must progress to IMV prior to death from COVID-19, i.e. death is assumed to be conditional on IMV progression. This may not necessarily be true in practice due to potential limitations in resources and other forms of disease progression. However, in planning this study, we believed that this would be the progression of the majority of COVID-19 patients who progress to ARDS. This also matched the available data at that time on clinical progression and the elicitation of information from the study team.

Using this assumption and considering the VFD endpoint definition, data are generated for patient i=1,…,200 on treatment t=0,1 as follows:

$$y_{it}^{IMV} \sim Bernoulli(p_t^{IMV})$$
$$y_{it}^{death|IMV} \sim Bernoulli(p_t^{death|IMV})$$
$$y_{it}^{VFD|IMV} \sim \exp(\lambda_t^{VFD|IMV})$$

for the endpoints IMV progression, death conditional on IMV progression, and ventilator-free days conditional on IMV progression. IMV progression, y_{it}^{IMV}, is calculated directly from the generated data. Death, y_{it}^{death}, is defined as 0 if y_{it}^{IMV} is 0 else it is $y_{it}^{death|IMV}$. Finally, y_{it}^{VFD} is defined as $y_{it}^{VFD|IMV}$ rounded to the nearest day and truncated at 28 if y_{it}^{IMV} is 1 and y_{it}^{death} is 0. If y_{it}^{IMV} is 0 then VFD is 28, and if y_{it}^{death} is 1 then VFD is -1.

For this simulation, many scenarios are investigated, including scenarios of no treatment benefit and harm compared with placebo to investigate false positive rates. These rates are well controlled at the levels specified by the tests (1-sided $\alpha < 0.05$) for each candidate endpoint and will not be discussed further. The scenarios of interest, however, for this report compare the power of the candidate endpoints with varying levels of plausible treatment benefit based on expert elicitation and the rapidly emerging

Scenario	p_0^{IMV}	p_1^{IMV}	$p_0^{death\|IMV}$	$p_1^{death\|IMV}$	$\left(\lambda_0^{VFD\|IMV}\right)^{-1}$	$\left(\lambda_1^{VFD\|IMV}\right)^{-1}$
Improve: VFD	0.25	0.25	0.5	0.5	14	21
Improve: IMV	0.25	0.125	0.5	0.5	14	14
Improve: VFD and IMV	0.25	0.125	0.5	0.5	14	21
Improve: All	0.25	0.125	0.5	0.25	14	21

Table 1. Simulation Scenarios

data at the time. Specifically, these scenarios are in **Table 1** with subscripts of 0 and 1 for placebo and experimental treatment parameters, respectively. Through discussions with both internal and external thought leaders, it was determined that an effective treatment should demonstrate at least a 50% relative improvement on the endpoints of interest. This underlying assumption is the basis for powering this study. Relative improvement on treatment at this level for COVID-19 has not yet been demonstrated for mechanical ventilation endpoints in randomized trials (Siemieniuk et al., 2020).

Patient data are generated using R version 3.6.3. Simulated trial data are analyzed using a Fisher's Exact test for dichotomous data and a Wilcoxon Rank Sum test for ordinal data at a one-sided 0.05 significance level. The statistical power of each endpoint will be reported for these scenarios. The simulation results for each scenario are based upon 10,000 independent replications.

Simulation Results

Figure 1 shows the final trial results of the three candidate endpoints, VFD, IMV progression, and Death across the columns and the scenarios described above with increasing efficacy down the rows. Success is defined as a statistically significant result, so probability of success can be considered study power for these scenarios and endpoints. In totality, the results show that VFD analyzed using the Wilcoxon Rank-Sum test is more powerful than the other endpoints across all the scenarios that were investigated, although, the degree of improvement varies.

When the treatment benefit is only demonstrated by the number of ventilator-free days, none of these endpoints are sensitive given the data generation parameters and study size. This is driven by the small percentage of patients progressing to IMV in this scenario that do not die (12.5% of the total patients) so there is only a treatment difference on a small subset of the simulated patients. Under these assumptions, a large trial would be needed to be well powered for this scenario.

There is a large increase in sensitivity when treatment prevents progression to IMV. The largest degree of separation in power of the VFD endpoint compared to the other endpoint occurs when there is a treatment benefit on the ventilator-free days, progression to IMV and conditional mortality, shown in bottom row of **Figure 1**. This is expected as, under this assumption, there is a treatment-driven change in the VFD endpoint across all dimensions of the ordinal scale.

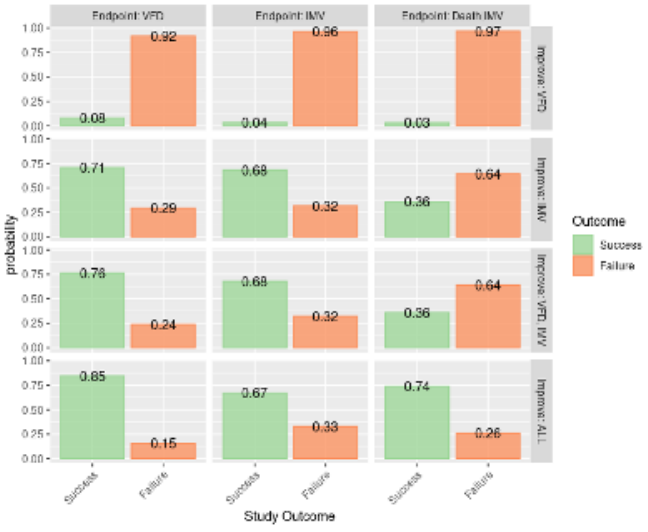


Figure 1. Statistical power of VFD, IMV and Death endpoints in planned study

Discussion

The understanding of clinical endpoints and trial design for the treatment of COVID-19 is rapidly developing. Through conversation with regulatory agencies and health care providers, a variety of clinically important endpoints were identified. These endpoints, however, were poorly understood and little was known about the best way to leverage them in the development of meaningful treatments.

Dichotomous endpoints, such as death or IMV, that capture direct disease progression may be more clinically meaningful to caregivers and patients but may require larger, more expensive trials to clearly demonstrate effectiveness than can be invested in the early stages of drug development. These endpoints are likely to be required in well-powered large confirmatory studies.

Ordinal endpoints, like VFD as defined in this report, provide an alternative to dichotomous endpoints and are more able to show efficacy in smaller studies. This VFD endpoint also has the advantage of being able to capture the hypothesized impact of treatment across combinations of multiple meaningful categories, IMV progression, mortality and time on ventilator. While the results of this simulation study indicate that time on ventilator does not seem to be a large driver for this endpoint's sensitivity, its inclusion does provide marginally more power. Exploring other magnitudes of treatment effect could also demonstrate additional utility of the endpoint.

Note that in simulations using death as the endpoint, success was achieved 36% of the time when there was not a direct treatment benefit on this endpoint. This is due to the endpoint being conditioned on progression to IMV, which results in a there being a treatment difference on the marginal death rates. This also is the driver behind the results showing that if there is a treatment benefit on both IMV and death, this endpoint is more powerful than IMV alone.

Given the uncertainty and evolving understanding of this infectious disease, understanding the sensitivity of endpoints in clinical trials for COVID-19 to capture meaningful treatment efficacy is vital for successful trial design. Simulation studies such as this allow for a better understanding of the interrelatedness of important clinical endpoints. They help inform trial design and potentially expedite the development and availability of meaningful treatments for patients.

References

1. Bellingan G., Brealey D., Mancebo J., Mercat A., Patroniti N., Pettila V., Quintel M., Vincent J., Maksimov M., Jalkanen M., Piippo I., Ranieri V.M. (2017). Comparison of the efficacy and safety of FP-1201-Iyo (intravenously administered recombinant human interferon beta-1a) and placebo in the treatment of patients with moderate or severe acute respiratory distress syndrome: study protocol for a randomized controlled trial. *Trials*, 18, 536.

2. Finkelstein D.M., Schoenfeld D.A. (1999). Combining Mortality and Longitudinal Measures in Clinical Trials. *Statistics in Medicine*, 18, 1341-1354.

3. Ranieri V.M., Pettila V., Karvonen M.K. (2020). Effect of Intravenous Interferon β-1a on Death and Days Free From Mechanical Ventilation Among Patients With Moderate to Severe Acute Respiratory Distress Syndrom: A Randomized Clinical Trial. *JAMA*, 323(8), 725-733.

4. Siemieniuk R., Bartoszko J., Ge L., Zeraatkar D., Izcovich A., et al. (2020). Drug treatments for covid-19: living systematic review and network meta-analysis. *BMJ*, 370, m2980. ■

REAL-WORLD DATA AND COVID-19 CLINICAL TRIALS: COMPARISON OF TWO NATIONAL APPROACHES

Sheng Feng (Parexel), Amanda Shield (Parexel) and Andy Wilson (Parexel)

On January 7, 2020, the Chinese National Medical Products Administration (NMPA) released its first official guidance on using real world evidence (RWE) to support drug development. This guidance laid out a bold vision to integrate RWE into the regulatory process in a scientifically rigorous way designed to move the field forward (NMPA guidance 2020).

On January 11, 2020, Chinese state media reported the first known death from an illness caused by the coronavirus, which occurred in Wuhan, a city of 15 million people in central China (Taylor 2020). Between January and May 2020, more than 500 COVID-19 related clinical trials were registered in China (Wang 2020). Multiple innovative study designs centering RWD/RWE were considered during the first few weeks of the pandemic. Single arm studies with historical or external control arms and Bayesian adaptive designs were discussed in two phase 3 clinical trials testing a promising medication. A platform design that was used to simultaneously test four investigational drugs for Ebola virus (Mulangu et al. 2019) was adapted to evaluate three different investigational drugs repurposed for treating COVID-19 patients in China. The World Health Organization (WHO) suggested a two-step study design—pilot (which could easily be replaced by RWD) and pivotal. Though randomized controlled trials (RCT) vs RWD/RWE approaches were debated among clinical trial design experts in China, ultimately none of these RWD/RWE-enhanced studies were implemented.

Real world study designs have clear advantages that could be of significant help in any pandemic and specifically in the fight against COVID-19, where traditional methods may be insufficient. RWD studies are useful for rapidly characterizing conditions that are new and/or not well-defined, for identifying potential risk factors, and for determining comparative effectiveness and safety of treatments. Real world techniques, including advances in targeted learning (van der Laan 2018), are designed to manage complexity in both patients (e.g. heterogeneous populations) and in the data, and are therefore well-suited to understanding clinical trial data captured under pandemic conditions. In spite of these advantages, RWD and RWE have not played any significant role in COVID-19 clinical development in China.

The field of RWD/RWE is much more mature in the US than in China. RWD and RWE have not played a very significant role in the US COVID-19 solution either. RWD and RWE have had limited roles in both the direct response (e.g. COVID-19 treatments) to the pandemic and in the response to indirect effects, like filling gaps in traditional clinical trials that have been disrupted by the pandemic. China and the US have very different RWD/RWE landscapes. Why have RWD and RWE been similarly underutilized for clinical development in the COVID-19 pandemic?

As the site of the first known outbreak in the pandemic, China had no ability to prepare. Tests for the disease had to be invented; manufacturing capacity for those tests had to be expanded; staff to administer them had to be deployed; and logistical and supply chain challenges had to be resolved. At the same time, disease presentation was not well understood, and multiple case definitions were applied (Tsang 2020). COVID-19 diagnosis was made using dozens of toolkits, some developed prematurely. Accuracy of the tests varies. There were no structured mechanisms (like diagnosis codes) for capturing COVID-specific data. At the height of the outbreak, staff were overwhelmed. RWD, especially collected in this timeframe earlier in the pandemic, lacked adequate details, quality, and scientific grounding. Those RWD are not robust for regulatory purposes. In addition, clinical practices and data standards vary remarkably among hospitals. Research that provided potential treatments to patients was prioritized, standardizing patient documentation for secondary use understandably was not. This all occurred in a context in which the infrastructure for RWD/RWE was still being developed.

The US has well-developed RWD/RWE infrastructure, and had the advantage of drawing on the Chinese experience, especially with regard to testing and better understanding of the disease. In addition, the pandemic has spurred enormous research efforts; governments, academic institutions, healthcare providers, and industry are all working to understand, contain, and treat the disease. However, lack of coordination has led to duplicative efforts in some areas while other areas are underutilized and/or underserved, e.g., the lack of effort of generating decision-making-enabled RWD and making it available to the public.

The US government response has been chaotic (Garrett 2020), both in its variability between federal, state, local, and individual agency efforts and in the quality of its guidance. Political and socioeconomic considerations interact with scientific concerns and affect how resources are allocated. These conflicting priorities have limited the ability to coordinate.

Scientific research and the provision of care are based on collaboration. Academia, providers, and industry have been able to organize some initiatives. For example, multiple large COVID-specific claims databases have been quickly created using previously established pathways and newly defined coding systems. Detailed patient data, including clinical notes, lab values, imaging, etc., that could provide reliable regulatory-grade evidence to answer research questions like comparative effectiveness of treatments, though, requires separate, painstaking collection particular to each research study. Because of the operational difficulties and expense of this approach, it has not been coordinated at scale, and individual efforts to discern things like effectiveness collect data of variable quality. This variable quality can result in increased bias in estimates produced from real-world data. Hundreds of COVID-19 RWD/RWE studies have been published so far, many on pre-publication-domain without peer-review. Taken as a whole, this output has a lot of noise, making it harder to detect clinically meaningful signals.

Because of these limitations, trial designers have not proposed and advocated for RWD/RWE as key components in a clinical trial, e.g., external controls. However, RWD has helped COVID-19 clinical studies outside of the regulatory environment. Most importantly, RWD has helped scientists, clinicians, and researchers gain general and practical knowledge, including the natural history, progression, and treatments, of COVID-19, which has helped them to design more efficient trials.

Looking forward as the pandemic continues to unfold, the question becomes, can RWD perform better in terms of assisting the design of clinical trials and controlling virus spread? We believe the answer is yes. Clinicians, researchers, officials, and indeed the whole world has learned from the first wave and has gained needed experienced for subsequent waves. A consensus is that it will take concerted action on the part of multiple groups to make RWD more useful fighting COVID-19. Governments, academia and industry each has a role to play, and this drive forward has begun.

Governments and regulatory bodies are being flexible and innovative in their thinking. The COVID-19 pandemic is forcing FDA to accelerate their progress in incorporating RWD/RWE in rapid regulatory decision-making and clinical trial design. FDA is discussing this shift and resultant plans publicly. NIH is partnering with 16 drug companies in studies for COVID-19 treatments and vaccines, where the single arm design (and historical control with RWD) is being considered. EMA is ready to set up infrastructure for real-world monitoring of treatments and vaccines in Europe. China's regulatory agency will release RWD guidance soon (draft circulated since 5/28/2020). CFDA has published "Experts' Opinion of Digital Trials" as well. These guidances are paving the road to clinical utility of RWD in China.

RWD design is continuing to adapt. In academia, universities have been at the front line in designing and executing novel clinical trials. For example, the platform "Recovery" clinical trial (*https://www.recoverytrial.net/*) was being conducted at the University of Oxford; Duke University has been hosting the "Grand Rounds Rethinking Clinical Trials" (*https://rethinkingclinicaltrials.org/*), where pragmatic trials, RWD/RWE, and digital trials are presented every week. In China, in February the West China University Hospital and Parexel developed a protocol of a platform trial, similar to the "Recovery" trial, the first of its kind in China.

Data collection and collaboration is deepening. Industry is making every effort to advance the field. In May 2020, 22 biotech companies announced the "COVID-19 research databank", sharing RWD with the public. Parexel is developing a clearinghouse of real world COVID-19 data sources, that will aid researchers in finding and evaluating these resources. In China, artificial intelligence, smartphone apps and digital devices have been used to trace the sporadic sparkles of the second wave of COVID-19 and control the virus from spreading.

In addition to the individual effort from each party, combined efforts are critical. Best data, best analysts, trial designers, sponsors and regulators should be brought together, to have the best chances of defeating COVID-19 with the assist of RWD/RWE.

References

1. Garrett, L. (2020), "COVID-19: The medium is the message." *The Lancet*, 395(10228), 942-943. DOI: 10.1016/S0140-6736(20)30600-0

2. Mulangu, S, Dodd, LE, Davey, RT, Mbaya, OT, Proschan, M, Mukadi, D, et al. (2019), "A randomized, controlled trial of ebola virus disease therapeutics." *New England Journal of Medicine*, 381:2293-2303. DOI: 10.1056/NEJMoa1910993

3. National Medical Products Administration (NMPA). (2020), "Guiding principles for drug development and evaluation supported by real world evidence." Available from: *http://www.nmpa.gov.cn/WS04/CL2138/373175.html*

4. Taylor, DB. (2020), "A timeline of the coronavirus pandemic." *New York Times*. Available from: *https://www.nytimes.com/article/coronavirus-timeline.html*

5. Tsang, TK, Wu, P, Lin, Y, Lau, EHY, Leung, GM, Cowling, BJ. (2020), "Effect of changing case definitions for COVID-19 on the epidemic curve and transmission parameters in mainland China: a modelling study." *The Lancet Public Health*. DOI: 10.1016/S2468-2667(20)30089-X

6. van der Laan, MJ and Rose, S, (2018), Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies, Softcover reprint of the original 1st ed. 2018 edition. Springer.

7. Wang, Mingyue. (2020), "State council rectifies the chaos of COVID-19 clinical research: 43 projects have been voluntarily cancelled." *CN-Healthcare*. Available from: *https://www.cn-healthcare.com/article/20200410/content-534294.html?from=timeline&isappinstalled=0* ■

ESTIMANDS AND ESTIMATION FOR CLINICAL TRIALS IMPACTED BY THE COVID-19 PANDEMIC

Yongming Qu and Ilya Lipkovich (Eli Lilly)

Contact: Yongming Qu, Department of Biometrics, Eli Lilly, Lilly Corporate Center, Indianapolis, IN 46285, USA.
Email: qu_yongming@lilly.com

1. Introduction

On March 11, 2020 the World Health Organization (WHO) declared the coronavirus disease 2019 (COVID-19) pandemic, caused by SAS-Cov-2 virus infection. The pandemic has had a great impact on ongoing clinical trials from multiple aspects, including study treatment interruptions, treatment or study discontinuations, and missing clinical visits due to COVID-19 control measures (patient or site personnel's quarantine and travel restrictions, site closure, disruption on drug supply chain, or transportation) and COVID-19 illness. This has spurred many discussions in the scientific community including guidance from regulatory agencies, scientific communities, publications, webinars, podcasts, and posts on social media (ACRO, 2020; McDermott and Newman, 2020; EMA, 2020a, 2020b; FDA 2020a; Meyer et al., 2020; Collins and Levenson, 2020).

Despite the fact that the ICH E9 (R1) addendum released in November 2019 laid a solid foundation for estimands, the pandemic has uncovered much confusion and uncertainty within the clinical trials community surrounding defining and re-defining estimands and revising the estimation methods for handling missing values. The pandemic has caused an increasing amount of intercurrent events (ICEs) (e.g. treatment interruptions and discontinuations, AEs related to COVID-19 illness, etc.) and missing values (e.g. missing clinical visits). Current discussions within the scientific community have focused on the immediate need to revise statistical analysis plans for handling ICEs and missing values due to the COVID-19 pandemic (e.g., Meyer et al., 2020). It is important to realize the pandemic should not make us change the original target of inference that applies to "normal" situations. Instead of rushing to make changes, we need to ask ourselves: "as our study objectives have not changed, why cannot the original statistical analysis plan cover the pandemic period in terms of the primary estimand?"

In this paper, we will discuss estimands and estimation from a holistic perspective and provide a novel framework for estimands and estimation that will not only cover the current COVID-19 pandemic and future unexpected events but will also improve statistical analysis even under normal circumstances.

This article is organized as follows. In Section 2, we briefly review estimands in the causal inference framework. In Section 3, we provide in-depth discussion of all strategies for handling ICEs under the causal inference framework. In Section 4, we discuss options for handling ICEs and missing values according to the cause of ICEs. Finally, in Section 5 we summarize the key points and propose some enhancements of the current estimand framework in ICH E9 (R1).

2. Using causal language to define estimands

One gap in existing guidance on estimands (ICH E9 [R1], 2019) is that it does not use causal language and potential outcomes (POs) in discussing estimands and strategies for dealing with intercurrent events (ICE) defined as post-randomization events that may confound and complicate interpretation of the treatment effect.

The framework of POs was introduced by Neyman (1923), and was later accepted by most researchers in causal inference (see Rubin, 1978; Robins, 1986; Pear 2001) to formulate the target of inference in observational and randomized clinical trials. Hernán and Robins (2020) provide a comprehensive review of causal inference. We briefly review defining estimands with the PO language in this section and use the framework in discussing handling ICEs and missing values in the remainder of the article (see also Lipkovich et al., 2020).

The PO is a random variable for the outcome of a treatment regimen, if applied to a given subject i, even when contrary to the fact. Hence, it is often also referred to as a "counterfactual" outcome (see Pearl, 2001; Robins, 1998). Let Y_i denote the outcome of interest and

Y_i (a,b) denote the PO with assigned treatment regimen *a*, but actually taking treatment regimen *b* during the study. Assume we only have two treatment regimens of interest in a study and let A_i denote the treatment regimen to be studied, such that A_i=0 for the control treatment and A_i=1 for the experimental treatment. Then, the causal treatment difference for a subset of patients (*S*), if they would have adhered to their assigned treatments, is the average treatment effect (ATE) in the response between the two potential treatments averaged across all patients in *S*:

$$ATE = N_s^{-1} \sum_{i=1}^{N_s} E[Y_i(1,1) - Y_i(0,0) \,|S],$$

where N_S is the sample size for *S*. If the outcomes are independently and identically distributed, the ATE becomes the treatment difference in the population means μ_1-μ_0, where μ_0 and μ_1 are the population means under the control and experimental treatment, respectively. Any causal estimand can be written in the form of contrasting the outcomes under two "parallel" treatments at the individual level conditioning on the same population. For simplicity, in the rest of the article, we omit the averaging over finite (sub)populations and write the causal estimand as

$$ATE = E\{Y_i(1,1) - Y_i(0,0)|S\} = E\{Y_i(1,1)|S\} - E\{Y_i(0,0)|S\}. \quad (1)$$

We may omit the subset "*S*" if the population of interest is comprised of all randomized patients. Except in a cross-over study, for each subject, only one of the PO's can be observed; however, randomization allows us to estimate $E\{Y_i(1)\}$ and $E\{Y_i(0)\}$ using data from respective randomized arms (provided there are no missing values) and the causal inference can be drawn from observed data.

3. Using PO's to determine strategies for handling ICEs

ICH E9 (R1) provides a framework for defining estimands and handling missing values in estimation. The treatment of interest and the handling of ICEs are the two most important components in defining an estimand. The treatment of interest should include not only the randomized treatment, but also the treatment regimen (e.g., adjusting for other concomitant medications). ICH E9 (R1) proposes five strategies for handling ICEs: treatment policy, hypothetical, composite variable, while-on-treatment (WOT), and principal stratum (PS). Subsequent comprehensive discussions on defining and choosing estimands and handling missing values according to ICH E9 (R1) are provided by Ratitch et al. (2020a, 2020b) and Mallinckrodt et al. (2020).

According to the ICH E9 (R1) any post-randomization change in treatment is an ICE, which may lead to confusion, especially for treatments requiring flexible dosing. For example, in clinical trials evaluating insulin treatments, doses may be adjusted frequently to maintain the optimal glucose level. Considering every insulin adjustment an ICE seems to unnecessarily complicate a more natural interpretation of the dose adjustment as part of a single treatment regimen. Using the treatment policy strategy (Section 3.2) will result in ignoring all such ICE's which makes us wonder: why should we first define so many ICEs just to ignore them later? As an alternative, one may define treatment regimen first (in this example, insulin treatment with any glucose-driven insulin dose adjustment per investigators and patients' decision) and not classify those events that are part of the treatment regimens as ICEs. As there is no ambiguity in handling events that are part of treatment regimens (i.e., ignoring them), in the rest of this article we primarily focus on events that are not part of the treatment regimens.

The potential ICEs related to the COVID-19 pandemic include:

- Prolonged treatment interruptions either due to COVID-19 illness or the associated "controlled measures." The definition of "prolonged" should depend on the disease state, study objectives, and mechanism of action of the study medications and should be the same for treatment interruptions due to COVID-19 or other reasons.
- Study treatment discontinuations either due to COVID-19 illness (an adverse event [AE]) or the controlled measures.
- Death as a result of COVID-19 illness.
- Use of protocol prohibited medications to treat COVID-19 illness.

ICH E9 (R1) clearly points out methods for handling ICEs may be different depending on the nature of the ICE; however, before the pandemic it is common that only one strategy is used to handle all ICEs within a

single estimand. For example, in the PIONEER 2 study (Rodbard et al., 2019), two estimands are used: the treatment policy estimand (using the treatment policy strategy for all ICEs) and the trial product estimand (using the hypothetical strategy for all ICEs). While convenient, it is overly simplistic and is not consistent with the spirit of ICH E9 (R1). One of the major reasons the COVID-19 pandemic has led to so much discussion within study teams regarding amending estimands is arguably the use of a single strategy in handling all ICEs in a study.

As an example of how different strategies for handling ICEs may be justifiable depending on their causes, ICH E9 (R1) (2019, page 12) states, "… the question of what the values for the variable of interest would have been if rescue medication had not been available may be an important one. In contrast, the question of what the values for the variable of interest would have been under the hypothetical condition that subjects who discontinued treatment because of adverse drug reaction had in fact continued with treatment, might not be justifiable as being of clinical or regulatory interest." The COVID-19 related ICEs may belong to the latter situation in which we are interested in the PO if the pandemic would not have occurred.

One of the key aspects of the ICH E9 (R1) is (in agreement with earlier National Research Council, 2010) that it distinguishes between missing data and ICEs (e.g. change of treatment) that often cause missing data. In doing so, it encourages sponsors to document the nature of ICEs and the reason of missing values, as much as possible. As stated in ICH E9 (R1) (page 14), "… a prospective plan to collect informative reasons for why data intended for collection are missing may help to distinguish the occurrence of intercurrent events from missing data." We will discuss handling of ICEs using different strategies by the underlying reasons in defining estimands following the work by Akacha et al. (2017) and Qu et al. (2020b). This amounts to classifying the ICEs into three categories: due to adverse events (AEs), due to lack of efficacy (LoE), and due to administrative reasons that are not related to efficacy and safety.

One response of pharma to the ICH E9 (R1) is a tendency to include in every SAP all 5 strategies so as not to miss any opportunity. In our opinion, not all strategies may be equally important, and sponsors should critically evaluate them in light of clinically meaningful goals, which is in agreement with the spirit of the Addendum. Here we make several comments about common misuses of the guidance. Some of these points are elaborated later.

- Treatment policy strategy stated vaguely as incorporating any deviation from initial treatment often results in treatment comparison that cannot be meaningfully generalized to real clinical practice nor describe features of specific treatment regimens.
- Although the principal stratum is mentioned as a strategy to handle ICEs in ICH E9 (R1), the principal stratum should be a strategy for defining a population(s) of interest, not for handling ICEs (Scharfstein, 2019).
- The while-on treatment-strategies have little practical use, in our opinion, and are often used as a disguise for the "old good" last observational carried forward analysis.
- The composite strategy is a method for creating a composite endpoint, so we suggest "composite" outcomes should be explicitly listed as part of endpoint definitions.

In the rest of this article, we will revisit the hypothetical and treatment policy strategies for handling ICEs, using examples based on the three categories of ICE.

3.1. Hypothetical strategy

ICH E9 (R1) uses the term "hypothetical strategies" to refer to special cases when we are interested in the PO under the "hypothetical" treatment regimen which a patient may or may not follow. Stating a hypothetical strategy always requires posing an assumption (a "hypothetical" scenario) that in combination with other assumptions (e.g. on the missing data mechanism) make the resulting estimand identifiable. Based on the causal framework, there could be many hypothetical strategies for PO (Lipkovich et al., 2020). We only discuss a few special cases here.

The first type of hypothetical strategy is interested in the PO when a patient would have followed his or her initially assigned randomized treatment until study end despite an ICE. Using the notation in Section 2, the POs for a patient who has such an ICE are Y_i (a,a),$a \in \{0,1\}$; therefore this estimand is given by

$$E\{Y_i(1,1) - Y_i(0,0)\}.$$

In causal literature, this type of causal effect is often called a controlled direct effect (Pearl, 2009) of treatment where "controlled" means we force the ICE not to occur and the initial treatment to continue. We call this hypothetical strategy the controlled direct hypothetical (CDH) strategy. With this strategy, the estimand is the treatment difference if patients would have adhered to the designed treatment regimen.

This type of hypothetical strategy is useful in scenarios when such POs are generalizable to a real-world population because ICEs may reflect special conditions that would not necessarily be reproduced in the future or in a real-world setting. There are a few scenarios where the CDH strategy may be used. This strategy is most applicable when patients discontinue treatment for reasons unrelated to the experimental treatment, e.g., due to administrative reasons including the COVID-19 controlled measures. This strategy may also be applicable for the ICEs related to using rescue medications due to ethical reasons and using concomitant medications to treat COVID-19 illness, which could potentially impact the outcome. In these cases, one may be interested in the PO if the patient would not have used the (rescue) concomitant medications, as the (rescue) concomitant medications in the clinical trials may not reflect the real world or "normal circumstances." The CDH strategy may also be applied to prolonged treatment interruption or treatment discontinuation due to COVID-19 illness. As the COVID-19 illness does not occur under normal circumstances, one may be interested to understand the PO and the treatment effect in the absence of the COVID-19 pandemic.

The second type of hypothetical strategy is interested in the PO assuming patients who discontinue treatment (an ICE) due to an AE would have no benefit: as if the patients were left untreated starting from randomization. In this case, the estimand can be written as

$$E[\{Y_i(1,-1)\Delta_i(1) + Y_i(1,1)(1-\Delta_i(1))\} - \{Y_i(0,-1)\Delta_i(0) + Y_i(0,0)(1-\Delta_i(0))\}].$$

where "-1" in the second parameter $Y_i(\cdot,\cdot)$ indicates no treatment received and $\Delta_i(a)$ is the ICE indicator (0 for no ICE and 1 for ICE occurring).

While the definition of such an estimand is simple, the estimation of the PO without any treatment may be challenging, especially in active-comparator studies. We will discuss more about the estimation under this hypothetical strategy in Section 3.1. We call this hypothetical strategy the *no treatment hypothetical (NTH) strategy*.

The third type of hypothetical strategy is to define the PO as the outcome if the patient takes the medication until the ICE and then stops taking the medication. Using the PO language, the estimand is defined as

$$E[\{Y_i(1, g_i(T_i(1)))\Delta_i(1) + Y_i(1,1)(1-\Delta_i(1))\} - \{Y_i(0, g_i(T_i(0)))\Delta_i(0) + Y_i(0,0)(1-\Delta_i(0))\}],$$

where $T_i(a)$ is the time to the ICE under treatment a and $g_i(T_i(a))$ is the treatment regimen: taking treatment a until the occurrence of the ICE and then having no access to treatment until a specified assessment time. This strategy assumes patients may still benefit from or be harmed by the treatment even though they discontinue the treatment earlier. This strategy may be suitable for handling ICEs due to AE at a "normal time" (not for AE related to the COVID-19 pandemic), especially for treatment with potential long-term or disease-modification effect. We call this strategy *partial treatment hypothetical (PTH) strategy*.

The fourth type of hypothetical strategy is interested in the PO assuming patients who discontinue treatment (an ICE) due to an AE would have "null" efficacy compared to the control treatment (Lipkovich et al., 2020; Qu et al., 2020b). For the control group, the ICE due to AE is handled by the CDH strategy. For the experimental treatment group, we assume the PO for a patient in the experimental treatment group with ICE due to AE is what the PO would be under the control treatment from the beginning of the clinical trial, i.e.

$$Y_i(1, \text{No Treatment}) = Y_i(0,0) + \delta,$$

where δ is the average treatment difference under the null hypothesis. This leads to an estimand

$$E[\{(Y_i(0,0) + \delta)\Delta_i(1) + Y_i(1,1)(1-\Delta_i(1))\} - \{Y_i(0,0)\Delta_i(0) + Y_i(0,0)(1-\Delta_i(0))\}]$$

For superiority studies, δ=0, and for non-inferiority studies, δ is the non-inferiority margin (assuming the smaller the outcome, the better). This approach can only be applied to estimands with a hypothesis, which is most often the case in clinical trials. We call this hypothetical strategy the *null hypothesis hypothetical (NHH) strategy*.

In a placebo-control study, the NHH strategy is the same as the NTH strategy if there is no placebo effect (e.g., patients taking blinded placebo would have the same outcome as not taking any study medication, $Y_i(a,-1) \cong Y_i(a,0)$). The last three hypothetical strategies all shrink the estimand towards the treatment effect under the null hypothesis (by essentially nullifying the

treatment effect for patients experiencing the ICE) with certain reasonable assumptions. The estimand based on the NHH strategy is always equal to the treatment effect under the null hypothesis. For superiority studies with no placebo effect, the NTH and PTH strategies provide an estimand equal to the treatment effect under the null hypothesis.

The NTH, PTH, and NHH strategies may be appropriate for handling discontinuations of the study treatment due to tolerability or other AEs occurring under normal circumstances, where such patients are assumed to have no or partial benefits from the treatment.

3.2. Treatment policy strategy

ICH E9 (R1) describes the treatment policy strategy as "the occurrence of the intercurrent event is considered irrelevant in defining the treatment effect of interest: the value for the variable of interest is used regardless of whether or not the intercurrent event occurs." Following the notation in Section 2, let $A_i^* = \{A_i, g_i(Z_i(A_i))\}$ be the treatment regimen (policy) patient i takes (which generally is not precisely defined in the protocol), where Z_i (possibly multidimensional) is post-baseline intermediate outcomes that affect treatment changes captured by patient-specific function $g_i(\cdot)$ that maps the evolving patient's outcomes $Z_i(A_i)$ into treatment decision 0 or 1. Using the causal framework, the estimand using this treatment policy strategy is defined by

$$E\left\{Y_i\left(1, g_i(Z_i(1))\right) - Y_i\left(0, g_i(Z_i(0))\right)\right\}. \quad (3)$$

Treatment policy strategy essentially compares outcomes (across randomized groups) associated with whatever actual treatment regimen (or "policy") is used for every patient no matter whether a patient followed protocol or was even receiving treatment at the time of the outcome measurements.

The treatment policy strategy is different from the *dynamic treatment regimens* (DTR) (Murphy et al., 2001; Moodie et al., 2007) in which the time-varying treatment regimens are clearly defined based on evolving patients' outcomes (e.g. adding additional concomitant or rescue medication if the intermediate efficacy outcome is poor). Let $g(\cdot)$ (which does not have the subscript i) be a function that maps Z_i to the choice of subsequent treatments. The estimand for DTR is defined as

$$E\left\{Y_i\left(1, g(Z_i(1))\right) - Y_i\left(0, g(Z_i(0))\right)\right\}. \quad (4)$$

In DTR, the rules for treatment changes ($g(\cdot)$ without the subscript i) are clearly defined (e.g. add medication X if the Z_i is greater than δ), while treatment policy strategy allows patients to have their own individualized rules ($g_i(\cdot)$ with the subscript i).

With the treatment policy strategy for handling ICEs, the treatment regimen of interest is the randomized treatment with all possible deviations (such as use of rescue medication or stopping the study medication). One argument for using the treatment policy strategy is that the treatment policy regarding the use of a drug may be anticipated in real clinical practice, and it guarantees no bias by preserving randomization. While treatment policy strategy may appear an optimal combination of RCTs and observational studies, in fact, there is reason to believe it combines the weaknesses rather than strengths of both.

First, without a careful description and requirement for certain requirements of compliance for the treatment regimen, it is very unlikely the resulting treatment effect estimate can be applied to a real clinical setting. Except for pragmatic studies (Tunis et al., 2003), the difference in the settings (e.g. frequency of visit, diligence of follow-up, allowed concomitant medications, etc.) between clinical studies and the real-world situation is generally large. In addition, sponsors may be able to manipulate the estimand (e.g. increasing the treatment difference) by specifying in the protocol the use of rescue medications with suboptimal efficacy which may not generalize to clinical practice. For an estimand using a hypothetical strategy, the treatment effect for the real-world setting can be projected by discounting potential outcomes for non-compliant patients, but it will be difficult to do so for an estimand based on the treatment policy strategy.

Secondly, to obtain generalizable inference for a treatment effect, one must assume sensible requirements for the use of rescue medication and duration of drug interruptions during the study. In a long-term study, if a patient takes a treatment for one week and stops the treatment, it is hard to argue that the long-term outcomes for this patient is reflective of the treatment regimen expected in a real-world setting. One may argue many successful studies did use all data collected regardless of treatment compliance (e.g. CV outcome studies), but this may only be the case when a small proportion of patients were severe non-adherers, which has a limited impact on the overall results. In a

disaster or pandemic, the proportion of severely non-compliant patients could become very large, and the problem cannot be relegated to having a negligible impact. Therefore, the argument for the treatment policy strategy to reflect the appropriate treatment regimen becomes increasingly vulnerable in the COVID-19 pandemic. This is exactly the reason why studies using treatment policy strategies potentially require amending the definition of their primary estimand (Meyer et al., 2020) to handle the COVID-19 related ICEs. ICEs due to COVID-19 illness and related controlled measures are clearly not part of the treatment regimen of interest, so the treatment policy strategy should not be applied in this situation.

In summary, in most cases, the vaguely defined treatment regimen (policy) may lead to challenges in understanding the "true" treatment effect and unsuccessful extrapolation of the clinical trial study results to the real-world setting. Treatment policy strategies should generally be avoided except for two situations: (1) ICEs are explicitly included in treatments of interest under a rigorous treatment regimen (e.g. a DTR) – of course one can argue these events are part of treatment regimens and should not be considered ICEs, and (2) selected ICEs in a pragmatic study in which the study setting is similar to the real-world setting.

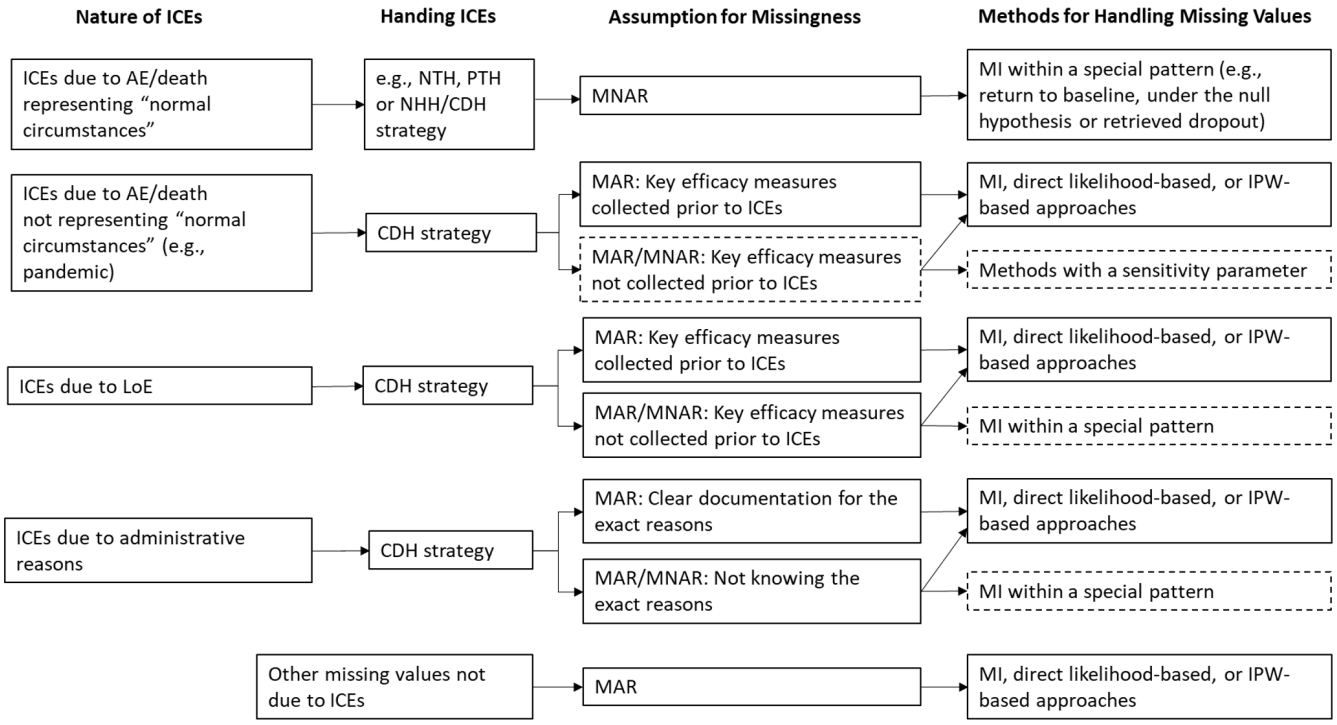
4. Handing ICEs and missing values

As argued here and in many other sources, ICEs should be distinguished from missing data (which often results from ICEs). Therefore, handling missing data belongs to the estimation procedure that should be considered after the strategies for dealing with ICEs are defined. Outcomes considered missing for one strategy may be non-missing for another. Outcomes measured after ICEs may not be the POs of interest under a hypothetical strategy, thus they cannot be used. In this section, we treat the observed outcomes that are not the desired POs as missing values. Therefore, in the rest of this article, values referred to as "missing" can be the result of using a hypothetical strategy to handling ICEs (forcing us to discard outcomes collected after ICEs) or true missing values due to the fact that the outcome measurements are not collected. In the context of a longitudinal clinical trial, missingness can be classified into four categories (Rubin, 1987; Little, 1995):

- Missing not at random (MNAR). Conditional on the observed values, the probability of missingness is dependent of unobserved (missing) outcomes.
- Missing at random (MAR). Conditional on the observed values, the probability of missingness is independent of any unobserved outcomes.
- Covariate dependent MAR (Cov-MAR). Conditional on the baseline covariates, the probability of missingness is independent of any observed or unobserved outcomes.
- Missing completely at random (MCAR). The probability of missingness is independent of any observed and unobserved outcomes.

Based on the above definitions, MCAR and Cov-MAR are special cases of MAR. The National Research Council (2010) and Little (2012) clearly point out the importance of collecting reasons for missingness and using such information to better ascertain the assumptions of missingness. Surprisingly, it is still a common practice to use one assumption across the board for all missing values in developing estimation strategies. Information on reasons for missingness and ICEs may enable us to make the most appropriate assumptions for missing values and to use the most appropriate statistical methods. Figure 1 provides an illustration of handling ICEs in defining estimands and handing missing values in estimation.

A couple of cautionary remarks should be made for interpreting the diagram. First, Figure 1 only includes ICEs that are not part of treatment regimens. For those ICEs that are part of treatment regimens, we can either not consider them ICEs or use the treatment policy strategy for handling these ICEs with the same result as ignoring them altogether. Secondly, Figure 1 only serves to illustrate (with selected statistical procedures) some general principles the reader may apply to their specific needs. There are many imputation and analysis strategies for handling missing values (see Mallinckrodt and Lipkovich, 2016 and references therein). In real clinical trials, a variety of appropriate imputation methods may be selected and often sensitivity analyses need to be performed (see Cro et al., 2020). However, we believe the proposed framework should generally be followed. Specifically: (1) the classification of ICEs and imputation methods should be based on the nature of ICEs, and (2) the PO language should be used to define the estimand, which means some kind of hypothetical strategies should mostly be used in addressing ICEs.



* Only ICEs that are not part of treatment regimens are included in this diagram.

Figure 1. Handling missing values based on the nature of ICEs. ICEs that are part of treatment regimens are not included in this diagram. Missingness (or missing values) includes missing data as a result of handling ICEs by a hypothetical strategy and missing measurements of the outcome. The solid boxes are used for the primary strategy and dashed boxes are used for alternative strategies or sensitivity analyses. Abbreviations: AE, adverse events; CDH, controlled direct hypothetical; ICEs, intercurrent events; IPW, inverse probability weighting; LoE, lack of efficacy; MAR, missing at random; MI, multiple imputation, MNAR, missing not at random; NHH, null hypothesis hypothetical; NTH, no treatment hypothetical; PTH, partial treatment hypothetical.

4.1. ICEs due to AEs

We suggest ICEs due to AEs be classified into two subcategories: (1) due to AEs representing the normal circumstances and (2) due to AEs not representing normal circumstances (pandemic and other region or nationwide crises). For example, the AE of COVID-19 illness that does not occur before the onset of pandemic or after it ends should be considered in the second subcategory.

AEs in the first sub-category reflect the expected environment for patients. It is not plausible to assume patients may still gain the full benefit of the drug if patients are not able to complete the treatment due to an AE. Therefore, we may use the NTH strategy (assuming patients have no benefit from the treatment), the PTH strategy (assuming patients have partial benefit), or the NHH strategy (assuming patients have "null" treatment effect) to handle such ICEs.

When using the NTH strategy to handle such ICEs, missing data due to ICEs can be imputed using baseline values if assuming no change from baseline. Examples of such methods include multiple imputation (Rubin, 1987) by baseline values and likelihood-based return-to-baseline method (Zhang et al., 2020). Note the imputation by baseline values, which assumes the PO has the same mean as the baseline while accounting for variability, is not equivalent to the baseline value carried forward method, assuming the PO takes the exact same baseline value (without imputation variability). Alternatively, missing values due to AE related ICEs can be imputed using retrieved dropout multiple imputation (CHMP, 2010), i.e. using the outcome for patients who do not use any treatment other than the standard care after the AEs but stay in the study (assuming the treatment effect before AEs is washed out). Note if patients take additional non standard care medications after ICEs, the outcomes for these patients should not be used to inform the imputation model.

If the PTH strategy is used for handling such ICEs, the outcomes measured after the ICEs are the POs of interest and therefore can be used for estimation. For those patients with truly missing outcome after ICEs, the data observed for other patients in the same treatment group and with similar ICEs may be used to impute the missing values (e.g. using "retrieved dropout" imputation). Note even though retrieved dropout imputation can be used for both NTH and PHT, the assumptions behind it are different: for the NTH strategy, retrieved dropout imputation assumes the treatment is washed off starting

from the occurrence ICEs to the time of outcome measurement; for the PTH strategy, there is no such assumption. Another plausible approach is to use copy-reference imputation (Carpenter et al., 2013) which assumes the efficacy retained up to the ICE but also assumes no additional treatment benefit after the ICE. This approach may not be appropriate if the treatment effect can be washed out quickly after treatment discontinuation.

If using the NHH strategy to handle these ICEs, missing values can be imputed using a reference-based imputation (e.g. jump-to-reference approach, Carpenter et al., 2013). In non inferiority studies, the non-inferiority margin can be added/subtracted after the imputation. We call this method of imputation multiple imputation under the null hypothesis. As a note, we call the multiple imputation methods using a special population or with a special assumption, multiple imputation with a special pattern for the remainder of the document.

For the second category of ICEs due to an AE not representing "normal circumstances" (e.g., COVID-19 illness), we may still be interested in the outcome that would have been observed had the patients taken the assigned medication without the AE. Therefore, the PO as if patients would have completed the study under assigned treatment regimens is of interest. The CDH strategy should be used to handle these ICEs, and methods based on the MAR assumption are recommended for handling the resulting missing values. If there is a strong reason to believe the MAR assumption is not valid, an alternative method may be used. For example, when there is a biological plausibility the occurrence of severe COVID-19 illness or death that generally causes treatment discontinuation may be correlated with the intermediate efficacy outcome of interest and the intermediate outcome prior to the AE is not collected. For such scenarios, the imputation for the PO under a special pattern (e.g. assuming the missing efficacy outcome prior to the AE is poor) may be performed. Since this type of methods requires selection of a special pattern (based on subjective judgment), it is generically recommended as a sensitivity analysis. In addition, when all patients with similar patterns have missing values (e.g. missingness due to death), a sensitivity parameter may be introduced for imputation (Mehrotra et al., 2017; Zhang et al., 2020; Cro et al., 2020).

4.2. ICE due to LoE

As mentioned in Section 3, ICEs due to LoE include treatment discontinuation due to LoE and use of rescue medications that are not part of the treatment regimen. A medication belonging to standard care is generally considered part of the treatment regimen, while a rescue medication, for ethical reasons, is generally not considered part of the treatment regimen.

The CDH strategy is often used to handle these ICEs due to LoE. This is because the scientific questions addressed by the study is what outcomes patients would get if taking treatment as directed provided there is no counterindication. If the efficacy measurements prior to ICEs are collected, the assumption of MAR is plausible and statistical methods based on the MAR assumption should be used. We can use a variety of methods for estimating treatment effects under MAR. These include methods based on likelihood such as direct likelihood (e.g. linear mixed models for repeated measures) and multiple imputation. Non-likelihood based methods are often preferred when likelihood is not readily available (such as for repeated measures binary data) and include weighted generalized estimating equations (WGEE) and doubly robust methods combining modeling ICE (treatment discontinuation) and outcome process (Robins and Rotnitzky, 1995; Robins et al., 1995; Bang and Robins, 2005). Methods based on multiple imputation are popular in that they can be easily coupled with sensitivity analyses under various departures from MAR, e.g. using delta-adjusted sensitivity analyses (Cro et al., 2020). See Mallinckrodt and Lipkovich (2016) and references therein. In cases when important efficacy measurements predictive of the ICEs are not collected, multiple imputation under a special pattern may be used. For example, we may assume these patients have similar intermediate efficacy outcomes as other patients within their treatment arm with similar ICEs. Methods based on MAR essentially assume patients with an ICE have observed "counterparts" who did not experience an ICE having similar outcome history as the patient with the ICE prior to the ICE. For situations when no such data is available by design (for example, strict rescue conditions preclude observing outcomes for any patients meeting non-response condition), sensitivity analyses such as multiple imputation under special pattern can be used.

4.3. ICEs due to administrative reasons

ICEs due to administrative reasons are ICEs generally considered unrelated to efficacy or safety. Treatment discontinuations or unacceptable duration of drug interruptions due to COVID-19 control measures during the pandemic belong in this category. POs under the designed treatment regimen should be considered, and the resulting missing value can be reasonably considered as MAR. For ICEs due to administrative reasons

but lacking clear documentation on the exact reasons of ICEs, such as loss to follow-up, the resulting missing values may not be considered MAR and can be imputed under a special pattern in the primary analysis or in a sensitivity analysis. For example, we may assume the efficacy outcomes prior to loss to follow-up are similar to those who have ICEs due to LoE.

4.4. Missing values not due to ICEs

Missingness not due to ICEs is generally considered to be purely due to randomness and a MAR based statistical method should be used to handle such missing values. Examples include invalid procedures in handling blood samples and performing the laboratory analytics, missing visits due to COVID-19 control measures, etc.

5. Summary and discussion

The novel COVID-19 pandemic may serve as a testing ground for the existing research and guidelines on estimands and handling of missing values, including the newly released ICH E9 (R1). It spurs discussion in the clinical trial community regarding revising protocols and statistical analysis plans to address the ICEs and missing values related to the pandemic (Meyer et al., 2020). Fortunately, the amendments to the current protocols and SAPs to address these issues can be done relatively quickly (although amending protocols is generally a long process) based on the framework provided by ICH E9 (R1); however, the very fact that so many studies impacted by the pandemic require amending, the protocol and SAPs is a sign the framework can be further improved.

In this paper, we start with reviewing the causal inference and POs framework, arguing it should be used for defining causal estimands, as the purpose of most clinical trials is to draw causal inference on the efficacy and safety of a new treatment. Using the causal framework, we discuss each strategy proposed in the ICH E9 (R1) for handling ICEs in detail. Our main conclusion is most clinically meaningful causal estimands should be defined based on POs and are necessarily "hypothetical." We provide a few examples of different hypothetical strategies for addressing different clinical questions; our intention was not to provide a complete list of possible scenarios for hypothetical strategies, rather aiming to define a general direction.

The treatment policy strategy, which intends to combine the advantages of controlled randomized clinical trials (following a rigid protocol and randomization) and observational studies (real world setting after discontinuation from initial treatment), seems to assume the drawbacks of these two types of studies. Accepting outcomes for whatever treatments were taken makes the description of treatment(s) of interest difficult and hardly generalizable to real-world settings, as treatment regimens in randomized clinical trials are unlikely to be close to those in real clinical practice.

We emphasize the value of hybrid strategies (handling ICEs differently according to the nature of the ICE) formed by classifying ICEs into 3 categories (due to AE, due to LoE, and due to administrative reasons) based on the nature of ICEs, and discuss their further classification into potential subcategories (Figure 1). We provide recommendations and various options on handling different ICEs and the resulting missing values.

We recommend the CDH strategy should generally be used for handling COVID-19 related ICEs as the purpose of studies (for non-COVID treatment) is to understand the treatment effect under normal circumstances (without this pandemic). The resulting missing values due to the CDH strategy of handling COVID-19 related ICEs or missing visits due to COVID-19 controlled measures may be dealt with under the MAR assumption. With the CDH strategy, for patients with the ICEs of treatment discontinuation due to AE or death, the hypothetical outcome may be imputed or inferred using an imputation under a special pattern if it is reasonable to assume an interaction between the COVID-19 disease and the intermediate (efficacy) outcomes prior to the ICE.

To reduce the ambiguity of the current definitions of ICEs, we suggest the following considerations in the application of the ICH E9 (R1) guidance. First, prior to discussing ICEs, treatment regimens of interest need to be defined precisely. Secondly, to be considered an ICE, this event should be a deviation from the treatment regimens of interest. Then, we can focus on handling the "true" ICEs that are not part of the treatments of interest. In addition, there are few situations when we need composite and WOT strategies. The principal stratum is not a method for handling ICEs. Therefore, we suggest hypothetical strategies should be predominately used to define causal estimands. More discussion on various hypothetical strategies are needed, and the selection of hypothetical strategies should depend on the nature of ICEs and study objectives.

ICH E9 (R1) lists three types of populations: all randomized study patients, a subset of patients based on baseline covariates, and a principal stratum by occurrence of a specific ICE. While we generally agree on the three categories, the principal stratum may also be defined by any post baseline outcome variable.

The success of implementing the conceptual and analytic framework proposed or discussed in this article critically depends on accurate and diligent collection of underlying reasons for ICEs and missingness. This should not be an extra burden for data collection, as good data collection is important, especially for the data elements (ICEs or missing values) that impact the key analyses. A number of COVID-19 related guidance documents (EMA, 2020b; FDA, 2020) also emphasize the importance of collecting such information.

This paper does not cover the situation when a single ICE may be caused by multiple reasons (Qu et al., 2020a). For example, one patient who feels the efficacy is not improved as expected while experiencing a mild AE may choose to discontinue the study medication. Although we do not cover such situations in this article, the framework provided can be used to select the appropriate hypothetical strategy and imputation methods to handle the ICE and the resulting missing values under such "trade off" scenarios.

In conclusion, as the pandemic may have prompted greater attention on estimands and missing data, the framework in this article advances the field beyond the current crisis. It may help streamline the process of choosing estimands and handling missing values in protocols and statistical analysis plans.

Acknowledgements

We are grateful to Stephen Ruberg and Bohdana Ratitch for numerous discussions on estimands which motivated us to write this article. We would like to thank Yu Du, Linda Shurzinske, and Dana Schamberger for a careful review of this article.

References

1. ACRO (2020), "Considerations to Support Clinical Trial Monitoring Oversight During COVID-19," *Association of Clinical Research Organizations* [online], Available at https://www.acrohealth.org/wp-content/uploads/2020/03/ACRO-Statement-on-Monitoring-Oversight-FINAL-3.13.20.pdf.
2. Akacha, M., Bretz, F., and Ruberg, S. (2017), "Estimands in Clinical Trials Broadening the Perspective," *Statistics in Medicine*, 36, 5–19.
3. Bang, H., and Robins, J. M. (2005). "Doubly Robust Estimation in Missing Data and Causal Inference Models". *Biometrics*, 61(4), 962-973.
4. Bornkamp, B., and Bermann, G. (2020), "Estimating the Treatment Effect in a Subgroup Defined by an Early Post-Baseline Biomarker Measurement in Randomized Clinical Trials With Time-to-event Endpoint," *Statistics in Biopharmaceutical Research*, 12, 19-28.
5. Carpenter, J. R., Roger, J. H., and Kenward, M. G. (2013). "Analysis of longitudinal Trials with Protocol Deviation: A Framework for Relevant, Accessible Assumptions, and Inference via Multiple Imputation," *Journal of Biopharmaceutical Statistics*, 23(6), 1352-1371.
6. CHMP (2010). "Guideline on Missing Data in Confirmatory Clinical Trials," European Medicines Agency Committee for Medicinal Products for Human Use (CHMP), EMA/CPMP/EWP/1776/99 Rev. 1. [online], Available at https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-missing-data-confirmatory-clinical-trials_en.pdf.
7. Collins, S. H., and Levenson, M. S. (2020) Comment on "Statistical Issues and Recommendations for Clinical Trials Conducted During the COVID-19 Pandemic", Statistics in Biopharmaceutical Research, DOI: 10.1080/19466315.2020.1779123
8. Cro, S., Morris, T. P., Kenward, M. G., & Carpenter, J. R. (2020). "Sensitivity Analysis for Clinical Trials with Missing Continuous Outcome Data Using Controlled Multiple Imputation: A Practical Guide," *Statistics in Medicine*. https://doi.org/10.1002/sim.8569
9. EMA (2020a), "Guidance on the Management of Clinical Trials during the COVID-19 (Coronavirus) pandemic," *European Commission* [online], Available at https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-10/guidanceclinicaltrials_covid19_en.pdf.
10. EMA (2020b), "Points to Consider on Implications of Coronavirus Disease 4 (COVID-19) on Methodological Aspects of Ongoing Clinical Trials," *European Commission* [online], Available at https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-10/guidanceclinicaltrials_covid19_en.pdf.
11. FDA (2020), "Guidance on Conduct of Clinical Trials of Medical Products during COVID-19 Pandemic," *U.S. Food and Drug Administration* [online], Available at https://www.fda.gov/media/136238/download.

12. Frangakis, C. E., and Rubin, D. B. (2002), "Principal Stratification in Causal Inference," *Biometrics*, 58, 21–29.

13. Hernán M. A., and Robins J. M. (2020). *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.

14. Hernán, M. A. and Scharfstein, D. (2018), "Cautions as Regulators Move to End Exclusive Reliance on Intention to Treat," *Annals of Internal Medicine*, 168, 515–516.

15. ICH E9 (R1) (2019). "ICH Harmonised Guideline Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statisical Principles for Clinical Trials E9(R1) ," *International Conference on Harmonisation* [online], Available at https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf.

16. Imbens, G. W., and Rubin, D. B. (1997), "Bayesian Inference for Causal Effects in Randomized Experiments With Noncompliance," *Annals of Statistics*, 25, 305–327.

17. Lipkovich, I., and Dmitrienko, A. (2014), "Biomarker Identification in Clinical Trials," in Clinical and Statistical Considerations in Personalized Medicine, eds. Carini, C., Menon, S., and Chang, M., New York, NY: Chapman and Hall/CRC Press.

18. Lipkovich, I., Ratitch, B., and Mallinckrodt, C. H. (2020), "Causal Inference and Estimands in Clinical Trials," *Statistics in Biopharmaceutical Research* 12:1, 54-67, DOI: 10.1080/19466315.2019.1697739.

19. Little, R. J. A. (1995), "Modeling the Drop-Out Mechanism in Repeated Measures Studies," *Journal of the American Statistical Association*, 90, 1112–1121.

20. Little, R. J., D'Agostino R., Cohen M.L., Dickersin K., Emerson S. S., Farrar J. T., Frangakis C., Hogan J. W., Molenberghs G., Murphy S. A., Neaton J. D., Rotnitzky A., Scharfstein D., Shih W. J., Siegel J. P., and Stern H. (2012), "The prevention and Treatment of Missing Data in Clinical Trials," *New England Journal of Medicine*, 367, 1355-1360.

21. Little, R. J., and Rubin, D. B. (2000), "Causal Effects in Clinical and Epidemiological Studies via Potential Outcomes," *Annual Review of Public Health*, 21, 121–145.

22. Mallinckrodt, C. H., Bell, J., Liu, G., Ratitch, B., O'Kelly, M., Lipkovich, I., Singh, P., Xu, L. and Molenberghs, G. (2020), "Aligning Estimators with Estimands in Clinical Trials: Putting the ICH E9 (R1) Guidelines into Practice," *Therapeutic Innovation & Regulatory Science*, 54, 353-364.

23. Mallinckrodt, C., and Lipkovich, I. (2016), "Analyzing Longitudinal Clinical Trial Data: A Practical Guide," Boca Raton, FL: Chapman & Hall/CRC Press.

24. Meyer, R.D., Ratitch, B., Wolbers, M., Marchenko, O., Quan, H., Li, D., Fletcher, C., Li, X., Wright, D., Shentu, Y. and Englert, S. (2020), "Statistical Issues and Recommendations for Clinical Trials Conducted During the COVID-19 Pandemic," Submitted to the COVID-19 special issue of Statistics in Biopharmaceutical Research. Statistics in Biopharmaceutical Research, DOI: a10.1080/19466315.2020.1779122.

25. McDermott, M. M., and Newman, A. B. (2020), "Preserving Clinical Trial Integrity During the Coronavirus Pandemic," *JAMA* [online], DOI:10.1001/jama.2020.4689.

26. Mehrotra, D. V., Liu, F., and Permutt, T. (2017), "Missing Data in Clinical Trials: Control□Based Mean Imputation and Sensitivity Analysis," *Pharmaceutical Statistics*, 16(5), 378-392.

27. Moodie, E. E. M., Richardson, T. S., and Stephens, D. A. (2007), "Demystifying Optimal Dynamic Treatment Regimes," *Biometrics*, 63, 447–455.

28. Murphy, S. A., van der Laan, M. J., Robins, J. M., and CPPRG (2001), "Marginal Mean Models for Dynamic Regimes," *Journal of the American Statistical Association*, 96, 1410-1423.

29. Pearl, J. (2001), "Causal Inference in the Health Sciences: A Conceptual Introduction," *Health Services and Outcomes Research Methodology*, 2, 189–220.

30. Pearl, J. (2009), "Causal Inference in Statistics: An Overview," *Statistics Surveys*, 3, 96–146.

31. Permutt, T. (2018), "Effects in Adherent Subjects," *Statistics in Biopharmaceutical Research*, 10, 233-235.

32. Permutt, T. (2020), "Treatment Effects, Comparisons, and Randomization," *Statistics in Biopharmaceutical Research*, 12:137-141.

33. Qu, Y., Luo, J., and Ruberg, S. J. (2020a), "Implementation of Tripartite Estimands Using Adherence Causal Estimators Under the Causal Inference Framework," *Pharmaceutical Statistics*. In press. https://onlinelibrary.wiley.com/doi/abs/10.1002/pst.2054.

34. Qu, Y., Shurzinske, L. and Sethuraman, S. (2020b), "Defining Estimands Using a Mix of Strategies to Handle Intercurrent Events in Clinical Trials," Submitted journal and under review. arXiv preprint arXiv:2006.03105.

35. Ratitch, B., Bell, J., Mallinckrodt, C., Bartlett, J.W., Goel, N., Molenberghs, G., O'Kelly, M., Singh, P., and Lipkovich, I. (2020b), "Choosing Estimands in Clinical Trials: Putting the ICH E9(R1) Into Practice," *Therapeutic Innovation & Regulatory Science*, 54, 324-341.

36. Ratitch, B., Goel, N., Mallinckrodt, C., Bell, J., Bartlett, J. W., Molenberghs, G., Singh, P., Lipkovich, I., and O'Kelly, M. (2020a), "Defining Efficacy Estimands in Clinical Trials: Examples Illustrating ICH E9 (R1) Guidelines," *Therapeutic Innovation & Regulatory Science*, 54, 370-384.

37. Robins, J.M. (1986), "A New Approach to Causal Inference in Mortality Studies With a Sustained Exposure Period—Application to Control of the Healthy Worker Survivor Effect," *Mathematical Modelling*, 7,1393–1512.

38. Robins, J. M. (1998), "Correction for Non-Compliance in Equivalence Trials," *Statistics in Medicine*, 17, 269–302.

39. Robins, J. M., and Rotnitzky, A. (1995), "Semiparametric Efficiency in Multivariate Regression Models with Missing Data," *Journal of the American Statistical Association*, 90, 122–129.

40. Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995), "Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data," *Journal of the American Statistical Association*, 90, 106–121.

41. Rodbard, H. W., Rosenstock, J., Canani, L. H., Deerochanawong, C., Gumprecht, J., Lindberg, S. Ø., Lingvay I., Søndergaard A.L., Treppendahl M.B., Montanya E. (2019), "Oral Semaglutide Versus Empagliflozin in Patients with Type 2 Diabetes Uncontrolled on Metformin: The PIONEER 2 Trial," *Diabetes Care*, 42, 2272-2281.

42. Rubin, D. B. (1978), "Bayesian Inference for Causal Effects: The Role of Randomization," *Annals of Statistics*, 6, 34–58.

43. Rubin, D. B. (1987), "Multiple Imputation for Nonresponse in Surveys", New York: Wiley.

44. Rubin, D. B. (2005), "Causal Inference Using Potential Outcomes," *Journal of American Statistical Association*, 100, 322–331.

45. Scharfstein, D. O. (2019), "A Constructive Critique of the Draft ICH E9 Addendum", *Clinical Trials*, 16, 375–380.

46. Shepherd, B. E., Gilbert, P. B, and Dupont, C. T. (2011), "Sensitivity Analyses Comparing Time-To-Event Outcomes Only Existing in a Subset Selected Postrandomization and Relaxing Monotonicity," *Biometrics*, 67, 1100–1110.

47. Tunis, S. R., Stryer, D. B., Clancy, C. M. (2003), "Practical Clinical Trials: Increasing the Value of Clinical Research for Decision Making in Clinical and Health Policy," *JAMA*, 290, 1624–1632.

48. Zhang, J. L., and Rubin, D. B. (2003), "Estimation of Causal Effects via Principal Stratification when Some Outcomes are Truncated by 'Death'," *Journal of Educational and Behavioral Statistics*, 28, 353–368.

49. Zhang, Y., Golm, G., and Liu, G. (2020), "A Likelihood-Based Approach for the Analysis of Longitudinal Clinical Trials with Return-to-Baseline Imputation," *Statistics in Biosciences*, 12, 23-36.

AN INTERVIEW WITH DR. MARGARET GAMALO, THE EDITOR-IN-CHIEF OF *THE JOURNAL OF BIOPHARMACEUTICAL STATISTICS*

BIOP Report: Congratulations on your recent appointment as the editor-in-chief of *Journal of Biopharmaceutical Statistics* (*JBS*). What is your vision on its mission, history and current status?

MG: *The Journal of Biopharmaceutical Statistics* (*JBS*) has been around for about three decades. It started back in early 1990s and since that time you can see how the industry has evolved over the years told through the collective perspective of statisticians. It is fascinating to study the progression of scientific research published in the journal reflecting both the scientific and evolving landscape of drug development. For example, the first two decades exhibited statistical issues ranging from proper use of one-sided vs two-sided tests in hypothesis testing, appropriate analysis for stability and expiry, bioequivalence, non-inferiority and superiority, subgroups and multi-center trials, interim analysis and adaptive designs, multiplicity, linear models analysis of endpoints, meta-analysis, among others for which the journal has provided the stage for consensus and application.

Since then, the Journal has seen a transformation in scientific topics influenced by shifts in research & development and the growing number of stakeholders. For example, a big trend has been the move to a more patient-centric health-care model, closely followed by the impact that technology has on all areas of the life sciences and the changing business models. Most recently, many of the published manuscripts revolve on topics such as efficient trial designs including the use of external data, estimation and uncertainty for go-no-go decisions, finding optimal individualized treatment rules or biomarker-guided treatments, treatment heterogeneity and considerations for regional and payer evaluation of multi-regional clinical trials, among others. Some proposed solutions apply an agile, iterative test-and-learn approach, rather than running long and expensive development processes to concoct the perfect solution. Other solutions often require modern technological innovation such as Artificial Intelligence (AI) and Machine Learning (ML) which provide significant opportunities to enhance drug discovery, clinical development, and commercialization.

When I wrote my editorial in December 2019 in time for the first issue under my editorship, I pondered the role of scientific publishing in an evolving industry. I realized that while our problems and insights today are much different from when the Journal was first envisioned, not much has changed in terms of its vision and mission. The *Journal* remains committed to the principle that better education of statisticians enables informed debate and decision-making about the valid application of new methodologies and aids in addressing misuse of statistical concepts, e.g. p-values, for scientific discoveries. The *Journal* will continue to strive to achieve excellence by publishing articles that present important new advances in an everchanging field of statistics within the pharmaceutical industry. *JBS* will always be a platform for education and dissemination of statistical research and innovation, repository for statistical solutions and choices, and a forum for scientific opinion on issues impacting methodology and applications in the pharmaceutical industry.

BIOP Report: There are several statistical journals with a focus on statistics issues in drug development and many more with a broad focus on medical

statistics and biostatistics. In your mind, what are the competing journals for *JBS*? What is your view on the relationship between *JBS* and these competing journals? What kind of journal that you would like to see JBS become in the next few years?

MG: There are two ways of thinking of the existence of other journals, i.e., whether to look at them as competitors or whether their existence complements the journal and enlarges the collective influence of statistics in the biopharmaceutical industry by broadening readership and impact of scientific publication. I prefer the latter embracing the perspective of *abundance*, i.e., there is abundance of scientific thought and research material for everyone. This, I think, is the better long-term strategy for scientific publication and consistent with how science has evolved throughout history. While Galileo Galilei[1] may have stated that "In questions of science, the authority of a thousand is not worth the humble reasoning of a single individual" we all know that science is also strengthened through constant validation - the reproducibility of results, the conduct of peer-reviewed, open literature research. Hence, the existence of other journals is also essential for quality scientific and statistical research. Furthermore, science and statistics moves and grows with collaboration. There is more and more to know in the world, and one can only have so much in our heads. In fact, the share of stuff we know as individuals is declining in any field. Inevitably, we need to collaborate as most research problems require multiple kinds of expertise.

Collaboration in science and statistics is good for making bold advances, i.e., innovation. Interaction with people with different perspectives or approaches prevents us from getting tunnel vision. Hence, one objective in scientific publishing should be to collectively grow the biopharmaceutical statistics field together through a feedback loop of continuous scientific and statistical innovation. Of note, innovation is key because it is multiplicative, meaning that the same input generates greater output far beyond the biopharmaceutical statistics field. The real problem then in scientific publication is sustaining statistical innovation which, I believe, is anchored on drivers of innovation, i.e., problems, constraints and opportunities. While the biopharmaceutical industry may be a mature field with established and highly regulated scientific and statistical problems, our problems can be unique given the interface of different stakeholders e.g., payor, regulator, patients, physicians, pharmaceutical companies. Capitalizing on this uniqueness to formulate innovative scientific and statistical solutions could have far reaching implications on other fields.

With this paradigm in mind, the *Journal* will strive to promote multidisciplinary collaborative research as a logical response to the expanse and pace of scientific revolution and transformation of the field. Not one discipline will be able to integrate all aspects of any problem or issue of interest alone with proper applicability and sufficient viability or competitiveness across a multi-stake holder industry. Increased focus on statistical research that has good cross-functional appeal apart from just pure statistical interest is important. The Journal believes that effective science is achieved not just by one discipline but by the collaborative and multi-disciplinary effort within the entire biopharmaceutical field.

BIOP Report: What are the new measures that you are taking to continue improving the quality and the visibility of the journal and make it truly impactful?

MG: One of the initiatives I instituted when I assumed the role of Editor-in-Chief was to ensure diversity in the editorial board. We have increased the number of women associate editors (AEs) and currently there is a diverse pool of AEs representing pharmaceutical companies, academia and regulatory agencies from various geographies. I realized that a journal relies on multiple and varied voices having a wide range of experiences. In fact, a diverse and inclusive editorial board brings the different perspectives that a journal needs to ensure quality and unleash value-driving insights, methods and practices. We screen papers for appropriateness to be published in the journal, i.e., papers that fits the aims and goal of the journal. We identify novel methods and applications and aspire to ensure relevance of topics while maintaining scientific

1 Misner, C. (1973). Wo Thome, KS & Wheeler, JA. Gravitation, 549.

integrity. We also plan special issues to address questions or bring up more discussions in some new and hot topics and weaves manuscripts coherently through different viewpoints or perspectives. Another consequence of a diverse pool of AEs is having a network to a broader pool of peer reviewers who are much more engaged. Finally, we rely on our reviewers to provide quality scientific and statistical reviews.

Diversity and inclusion must happen even in scientific publishing. Science and statistics will not meet its potential until the research culture enables and supports contributors from all backgrounds and circumstances and contributions of all kinds based on the interests, skills, and resources available. Failure to achieve diversity and inclusion of all stakeholders in science and statistics will slow progress in discovery and translation of knowledge to solving humanity's most pressing problems.

Another initiative that we are accessing currently is to harness the power of the crowd by highlighting key innovations or discussions in social media. I think as a leader in scientific publication, we need to be more proactive in disseminating information than merely passive repositories of scientific thinking brought to use through a Google search. It is estimated that ~68% of Americans get their news from social media[2]. The ease of use of social media platforms for communicating and disseminating information also makes them attractive to scientists. Furthermore, I believe that social media gives us the opportunity to engage directly with a wide range of audiences and helps us understand our readers.

Within the next 5 years we have several other initiatives planned as well e.g., partnering with a statistics professional organization on proceedings or narrowing the proceedings on hot topics or breakthroughs. There will be a few more changes along the way as we are trying to think about what the role of scientific publishing in a personalized information age means. How can we adapt to that environment and perhaps influence it as well particularly in biopharmaceutical statistics? I am open to suggestions and anyone is free to reach out with ideas on how we can improve.

2 *https://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/*

BIOP Report: *JBS* has successfully published several special issues in the past and attracted much attention. Do you plan to publish more special issues in the coming years and what are the topics of these special issues?

MG: We recently launched a call for paper for special issues in Real World Data/Evidence (RWD/RWE) and in the implementation of Estimands in clinical research. RWD/RWE is a growing area which still requires thought and consensus on how it can be applied given the range of possibilities. In the future, there will be a blurring of how evidence of effectiveness of investigational new drugs will be established. Moreover, this field will continue to grow in the next decade finding interconnections with clinical trials, clinical practice through electronic health records and digital health. Estimands, on the other hand, is a complex issue that is quite difficult to explore upfront and implement in the planning stage of a clinical trial. However, currently, it is changing the way we design trials, write the objectives, collect the data, conduct the trial, and perform analysis because the framework requires us to be more unambiguous about the questions we would like to answer. The complexity is also due to the presence of multiple scientific questions of interest about relevant treatment effects, interpretation of study results, and added value of drugs to different stakeholders (i.e. regulatory, prescribers, patients and payers).

Our guest editors are busy connecting with key scientific leaders in this field as well as disseminating the effort through different social media venues. What we believe to be important is to have a balanced and informative issue that will serve as a definitive reference for these topics to many scientists and statisticians in our field. For these special issues, the reviews will be rolling, i.e., once the reviews are completed the accepted manuscript will be posted in the webpage right away for access. Once a sufficient number of manuscripts are collected and reviewed, we will close the issue and publish it in print. I urge those who are interested to reach out to Junjing Lin (Takeda), Helen Qi (Bristol Myers Squibb), Yodit Seifu (Merck), or Bill Wang (Merck).

We will also launch another special issue on pediatrics very soon. Of note, majority of the investigational drugs

being studied in adults will study pediatric patients as well either through a requirement or in pursuit of an incentive. The challenges of running trials in children are accelerating efforts in innovative trial design and analysis. We hope that this special issue will provide a simple but comprehensive guide for statisticians/ clinical research scientists to determine the extent of development in a pediatric trial in accordance with the principles of extrapolation and how these trials can be streamlined to be as lean as possible to ensure that it provides the maximum information with the minimal number of pediatric patients exposed to research risk and to ensure timely completion of pediatric studies. We are trying to have this special issue coincide with a global virtual workshop on extrapolation in pediatric drug development. This workshop and subsequent special is still in the planning stages but will focus on statistician's and clinician's experience on pediatric drug development.

BIOP Report: As far as we know, many statistical journals are facing challenges to find highly qualified reviewers to complete review in a short timeframe, say one and half month, what do you think that *JBS* can do to address the issues such as delayed manuscript review?

MG: Indeed, this is a big problem in scientific publishing. Our Managing Editor, Victoria Chang (BeiGene), has been excellent in reminding AEs when the reviews are needed. We have been able to manage asking the reviewers to actively turn in their reviews. Of course, there are some review slips here and there as this is all voluntary work. For special issues, we ask the guest editors to have their own system of ensuring expedient reviews by having a standby review committee. As I have mentioned earlier, having a diverse set of associate editors has been very helpful. We still have plans to expand our editorial board with folks from the European Union and Asia. In fact, if you have an interest in serving as an AE and you have the passion to provide service to biopharmaceutical statistics and the society, please reach out to us.

We also encourage young scientists to take part of this endeavor actively. It does not require that one must be well experienced to serve as a reviewer. I think the major criteria to be a good reviewer are curiosity, critical thinking and the ability to ask good questions. I am aware that many of us will say that we are eyeballs deep with work. On the other hand, I argue that we need to take care of ourselves as well. One way of doing that is ensuring we retain our scientific and statistical thinking and keeping up to date on innovation and new statistical techniques. I always am reminded that we may need to keep learning if we want to be relevant and as we live longer. Perhaps Mahatma Gandhi was right in saying that we need to "*Learn as if you were to live forever*".

BIOP Report: Do you have any advice for the statisticians who would like to submit manuscripts to *JBS*?

MG: My general advice to all statisticians and not just to those who would like to submit manuscripts is -*Be curious and tell a good story*. Most breakthrough discoveries started with curiosity - the impulse to seek new information and experiences and explore novel possibilities. Curiosity is beneficial for all because it cultivates many levels whether it is one's organization or more broadly - the society. In fact, curiosity helps society make better decisions. When we are curious, we think deeply and rationally about decisions and come up with more creative solutions.

Curiosity does not necessarily have to result in a monumental breakthrough and certainly publishing in a scientific journal does not necessarily mean only novel solutions are entertained. Science moves by increments and not by leaps and bounds. Sometimes the insight to the problem is enough. What I also see as a problem is that there are times when statisticians hold back on their idea fearing that they may be too obvious. I think great ideas may sometimes seem obvious because the solution has all the parts of the question lining up and shedding light on a solution. Therefore, 'obvious' answers are not visible to most people, partly because most people are not thinking about the question. Ideas only come to those who recognize a problem and look for innovative solutions. I posit that even Einstein could possibly not find a solution if he had the wrong question.

Inside every scientific discovery, there is a good story. I think it is important to share that story as I am sure there are a lot of insights that went through, e.g., how the problem came about, why is it an important problem to pursue, how was the solution discovered, why other solutions failed. These experiences are actually very

informative and could help many researchers out there know or understand what works and what does not. Brilliant statisticians may sometimes be dissuaded by writing not because they do not know how to write but by not trying. We are our own limit. We need believe that something different can happen in order to break old patterns and we can choose that new outlook at any time. Part of being an effective statistician is not only to develop or apply sophisticated numerical calculations but also being an effective communicator in writing and in speaking.

BIOP Report: As we know, you had extensive working experience in government and pharmaceutical Industry. What is the impact of this unique experience on your perspectives on the role of statistics and statisticians in clinical trials research that demands a close collaboration between biopharmaceutical industry, government and academic?

MG: I learned the principles of drug development at the FDA and I learned how to apply them while understanding the challenges of drug development in the industry. I realized, many of the regulations in clinical trials are common sense and centered on ensuring safety of patients. Having reviewed hundreds of investigational new drugs (INDs) and new drug applications (NDAs), what is the right thing to do is sometimes very easy to spot because it is rational. I can also see the difficulty with implementing mitigations from the industry side for some of the concerns raised by regulatory agencies. I learned to assess what is ideal and what is applicable. In the case of the latter, there is no perfect solution most of the time, but the quality of the medicinal product and patient safety is paramount. Hence, when I think about what my job responsibilities and the role of a statistician in the biopharmaceutical industry, it may just be encapsulated by the provisions of Title 21 of the Code of Federal Regulations which is consistent with ensuring good clinical practice. Of note, good clinical practice recognizes that protecting data integrity is part and parcel of *ensuring safety of patients*.

More broadly, I think *statisticians need to be involved as key decision makers*. When we can understand and interpret data correctly, our ability to identify crucial areas requiring attention in drug development are enhanced, and our proposals for mitigating these key areas are likely to respond to the needs of our organization or the industry. In an age where data is essential for making big decisions whether in business or government, statisticians need to be `at the table' so we can assist and encourage informed decision making. However, this also entails that we need to be able to communicate in the language that is understandable by non-statisticians. Statisticians may need to be comfortable communicating about the problem not just in terms of numbers. We need to understand the whole problem and not just numerical ramifications, e.g., scientific, clinical, regulatory, payor, etc. Having a holistic view is what we need so we can provide more valuable and insightful feedback.

My experience on both sides of the industry (regulatory and pharma) has also shaped my collaboration with many statisticians in the industry. I have been more discerning on what topics are most impactful and so I think that *statisticians need to influence scientific thinking and progress* in the biopharmaceutical industry. Particularly, I learned how to *think big, start small, and learn fast* - our role in the industry is to have a broad vision while being mindful of how we act on it. For over 5 years I have been using a great deal of Bayesian methodology. However, I realized that most sample sizes are driven by the number of exposures needed to have sufficient data to establish safety. Hence the value of Bayesian methodology in terms of efficiency in late phases of development may not be apparent to encourage a strong push for change. However, in areas of unmet need and in pediatrics, the use of Bayesian methodology is clear because of feasibility and because of ethical principles of not having duplicative information to warrant translating conclusions from on population to another. That situation gave me a better perspective to focus on what innovative advances can bring meaningful change in policy. In fact, most of these innovative methods have been expanded to applications of propensity scoring methods to augment clinical trials particularly in pediatrics, orphan diseases and unmet medical need indications. So even in scientific research and policy, the words of Justice Ruth Bader-Ginsburg reverberate that "*Real change, enduring change, happens one step at a time.*"

BIOP Report: COVID-19 pandemic is having significant and long-lasting impact on how clinical trials are conducted. Meanwhile government, universities and many pharmaceutical companies are working together to find vaccines and new treatments for this disease. Do you have a plan to use the journal as a venue to promote the discussion on challenging issues arising in clinical trial designs and analyses?

MG: *The Statistics in Biopharmaceutical Research* (*SBR*) is already having a series of special issues on impact of COVID-19 on clinical trials and in COVID-19 related research. Some of my friends and colleagues are already working hard on that area and I am amazed at the speed of coordination and implementation. Two of the manuscripts I have been involved in writing will be published in that endeavor. Consistent with the spirit of collaboration I mentioned previously, I decided not to go with another special issue on COVID-19 in *JBS* because it would then be in competition with the *SBR* effort. Instead, any COVID-19 related research identified as helpful to the scientific community or related research activities will be given priority and expedited review. This allows us to publish any research and findings with greater speed and agility.

The main issue with developing drugs for COVID-19 is speed of innovation. However, many of the tools for acceleration have been discussed extensively in literature, e.g., adaptive design, data sharing, etc. What is lacking, from a statistical perspective, is on knowledge of appropriate endpoints in relation to patient population. Hence, COVID-19 disease progression models are needed to learn about how to conduct COVID-19 treatment clinical trials is important absent Phase 2 trials. Ensuring that clinical trials have a common set of data that can help inform other development is also important. However, currently is there is little data available to assess how this can impact speed of development. The best would be when data is already out there, what can we learn from it so that we can be better prepared should there be another catastrophic event of similar nature in the future.

I do encourage statisticians to contribute to the scientific efforts for developing treatments for COVID-19. I think it is a worthwhile endeavor and reminds us how interconnected we are and if we do not collaborate, we will be in this situation for a long time. I believe this is the best time for science and statistics. In fact, as I have mentioned previously - with uncertainty comes great innovation. Problems and constraints are backdrops for opportunity. As a cheery reminder, Sir Isaac Newton produced an unbelievable number of exceptional results including seminal experiments on the law of universal gravitation while quarantined during the London plague of 1665-1666[3] – though, I believe, he must have been curious and persistent even before that.

Acknowledgement

Margaret Gamalo would like to thank Yodit Seifu, Victoria Chang, Junjing Lin for their helpful comments to her responses.

Margaret (Meg) Gamalo, PhD *is Senior Director – Biostatistics, Global Product Development – Inflammation and Immunology at Pfizer Innovative Health. She combines expertise in biostatistics, regulatory and adult and pediatric drug development. She recently was a Research Advisor, Global Statistical Sciences at Eli Lilly where she helped secure the approval of baricitinib as the first modern treatment for atopic dermatitis and position baricitinib as a first-in-indication treatment for alopecia areata. She also helped in the clinical trials for baricitinib as a treatment for COVID-19. Prior to working at Lilly, she was a Mathematical Statistician at the Center for Drug Evaluation and Research in the Food and Drug Administration supporting regulatory reviews of drugs in the infectious diseases and ophthalmology therapeutic areas. Meg is an expert in pediatric extrapolation and leads the Pediatric Innovation Task Force at the Biotechnology Innovation Organization. In this task-force, she is leading a cross-functional think tank on why extrapolation needs to be a default strategy in pediatric drug development. She is also an active member of the European Forum for Good Clinical Practice (EFGCP) – Children's Medicine Working Party (CMWP) to establish decision criteria for the inclusion of adolescents in adult research. In the statistics profession, she is a member of the Executive Committee of the Biopharmaceutical Section and has served multiple administrative and scientific responsibilities within the section since 2014. She enjoys writing and mentors a group of statisticians in research activities on topics related to Bayesian methods, evidence synthesis, causal estimation for RWD/RWE, and in policy-oriented work on pediatric drug development.*

3 https://www.nationaltrust.org.uk/woolsthorpe-manor/features/year-of-wonders

SUMMARY OF AMERICAN STATISTICAL ASSOCIATION BIOPHARMACEUTICAL SECTION'S VIRTUAL DISCUSSION WITH REGULATORS ON TYPE I ERROR CONSIDERATIONS IN MASTER PROTOCOLS WITH COMMON CONTROL IN ONCOLOGY CLINICAL TRIALS

Rajeshwari Sridhara (FDA), Olga Marchenko (Bayer), Qi Jiang (Seagen), Richard Pazdur (FDA), and Bruce Binkowitz (Shionogi)

Master protocols are generally identified as protocols that try to answer multiple questions with respect to multiple diseases and/or multiple treatments. These protocols have been implemented as different types of trials such as basket trials, umbrella trials and platform trials depending on the Master protocol objectives. Master protocols have the potential to accelerate and save resources, particularly patient resource, with centralized governance structure, data sharing and use of a common control in the development of innovative treatments for cancer patients.

The example that follows exemplifies the necessity for a master protocol with a common control which can save precious patient resource, offer higher probability for a patient of being assigned to an innovative, experimental treatment and accelerate cancer drug development. For example, to evaluate treatment for patients with advanced renal cell cancer, five concurrent studies with sunitinib as the control treatment were conducted ((1) Checkmate 214, (2) Keynote 426, (3) Javelin Renal 001, (4) NCT02420821, and (5) NCT02811861 evaluating nivolumab, pembrolizumab, avelumab, atezolizumab, and lenvatinib, respectively). A single trial using a master protocol evaluating all the 5 treatments with a common control could have saved patient and other resources in this example with shared costs and potentially accelerated time.

Generally, it is understood that in a Master protocol the use of a common control can be efficient. However, there are differing views regarding adjustment of Type I error for the multiple comparisons of different treatments to the same control arm in a randomized controlled trial utilizing a master protocol. The American Statistical Association (ASA) Biopharmaceutical Section (BIOP) worked with Oncology Center of Excellence (OCE), US Food and Drug Administration (FDA) and hosted open forum discussion held on October 8, 2020 focusing on point and counterpoint regarding the necessity to adjust Type I error for multiple hypotheses testing when a common control is used in a randomized study under a master protocol. While there are both single arm and randomized trials that can be run under a master protocol, this discussion on Type I error consideration was limited to randomized control trials.

Oncology drug development is going through revolutionary changes both in terms of type of indications and type of drugs and with these changes there are increased numbers of smaller molecularly defined subset of patients or rare disease groups, unique indications, which pose challenges. Many drugs have been approved based on single arm trials and smaller number of patients based on tumor response. However, randomized studies are the key to ensure that the observed clinical benefit

(example: overall survival) and risk are attributable to the treatment under consideration. Use of Master protocols with a common control can allow the conduct of randomized studies in such situations.

The twenty panelists* for the discussion included members of the BIOP Statistical Methods in Oncology Scientific Working Group representing pharmaceutical companies, representatives from international regulatory Agencies (FDA, EMA, HC, PMDA, TGA, and SMC), academicians and expert statistical consultants. In addition, over 80 members attended the virtual meeting. The discussions were moderated by the BIOP Statistical Methods in Oncology Scientific Working Group co-chairs, Dr. Olga Marchenko from Bayer and Dr. Qi Jiang from Seagen, and Dr. Rajeshwari Sridhara, contractor from Oncology Center of Excellence, FDA.

The two hours discussion was productive and covered different scenarios where Type I error adjustment may be considered in a master protocol with a common control. While there were concerns in specific situations where type I error adjustment may be necessary, the panelists agreed that adjustment of type I error for multiplicity when a common control is used may not be necessary if the hypotheses are inferentially independent. However, when some of the hypotheses are inferentially dependent such as comparing different doses of the same drug or drug combinations with the same components, Type I error adjustment could be necessary. A detailed summary of the discussions will be published in *Statistics in Biopharmaceutical Research* (SBR) in the near future.

This was the inaugural open forum hosted by ASA BIOP to have scientific discussions among diverse stakeholder group – academicians, international regulators, and pharmaceutical companies focused on emerging statistical issues in cancer drug development. We look forward to similar open forum discussions in the future on a variety of important topics that include statistical aspects in cancer drug development involving different stakeholders and a multi-disciplinary approach.

*Panelists

- **Scott Berry** (Berry Consultants)
- **Bruce Binkowitz** (Shionogi)
- **Michael Coory** (TGA, Australia)
- **Thomas Gwise** (FDA)
- **Lorenzo Hess** (SMC, Switzerland)
- **Qi Jiang** (Seagen)
- **Filip Josephson** (EMA)
- **Naoto Kotani** (PMDA, Japan)
- **Nicole Li** (Merck)
- **Olga Marchenko** (Bayer)
- **Richard Pazdur** (Oncology Center of Excellence, FDA)
- **Martin Posch** (Medical Statistics at the Medical University of Vienna)
- **Andrew Raven** (HC, Canada)
- **Mary Redman** (Clinical Research Division, Fred Hutch)
- **Kit Roes** (EMA)
- **Yuan Li Shen** (FDA)
- **Richard Simon** (Consultant)
- **Rajeshwari Sridhara** (Contractor, Oncology Center of Excellence, FDA)
- **Marc Theoret** (Oncology Center of Excellence, FDA)
- **Yevgen Tymofyeyev** (Statistics and Decision Sciences, Janssen RD).

Disclaimer: *The views expressed are the personal views of the panelists*

13 BIOP MEMBERS WERE ELECTED TO BE ASA FELLOWS IN 2020

Name	Affiliation
Chung-Chou H. Chang	University of Pittsburgh
William Scott Clark	Eli Lilly and Company
Anastasia Ivanova	University of North Carolina-Chapel Hill
Yuan Ji	University of Chicago
Laura Lee Johnson	FDA
Pandurang M. Kulkarni	Eli Lilly and Company
Sergei Leonov	CSL Behring
Liang Li	University of Texas MD Anderson Cancer Center
Jason Jinzhong Liao	Merck & Co., Inc.
Robert Alan Oster	University of Alabama at Birmingham
Yongming Qu	Eli Lilly and Company
Michael Rosenblum	Johns Hopkins University, Bloomberg School of Public Health
Richard Conrad Zink	TARGET PharmaSolutions Inc.

CHUNG-CHOU H. CHANG

University of Pittsburgh

For demonstrated ability to advance the field of biostatistics; for being an invaluable collaborator and co-investigator for a large number of clinical researchers; and for being a superb teacher and mentor of students throughout the health sciences.

WILLIAM SCOTT CLARK

Eli Lilly and Company

For developing and implementing an innovative leadership-based strategy at Eli Lilly; for championing diversity and inclusion across many dimensions; and for coaching, mentoring, collaborating, and partnering in ways that empowered hundreds of research scientists and statisticians.

ANASTASIA IVANOVA

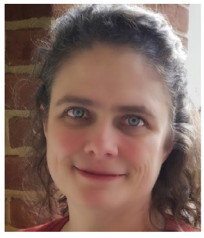
The University of North Carolina at Chapel Hill

For outstanding contributions in clinical trial design to advance clinical research and expedite the discovery of treatments that address crucial clinical needs.

YUAN JI

The University of Chicago

For pioneering contributions to statistical methodology and application, including interval-based dose finding, big data cancer genomics, and bioinformatics using robust Bayes approaches; for providing exemplary public user interfaces; and for exemplary mentorship and service.



LAURA LEE JOHNSON

US Food and Drug Administration

For advancing scientific principles on the use of patient experience data in clinical trials for regulatory and policy decisions; for contributions to the design and analysis of trials to evaluate treatments for rare diseases and in integrative health; and for outstanding service to the profession.

JASON JINZHONG LIAO

Merck & Co., Inc.

For innovative statistical methods and applications in the pharmaceutical industry; for advancing comparability/biosimilars studies, agreement studies, and assay development; for bringing life-saving drugs to market; and for extensive service to the profession.

PANDURANG M. KULKARNI

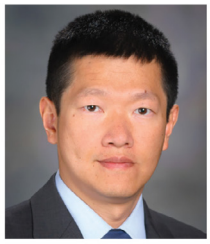
Eli Lilly and Company

For exceptional and innovative leadership of large statistical organizations; for significant contributions to bringing new medical treatments from discovery to market; for steadfast service to the ASA and the profession; and for outstanding mentoring and people development.

SERGEI LEONOV

CSL Behring

For methodological contributions to optimal experimental design, adaptive designs, and pharmacokinetics in the context of drug development; for leading interdisciplinary project teams in the pharmaceutical industry; and for service to the profession.



LIANG LI

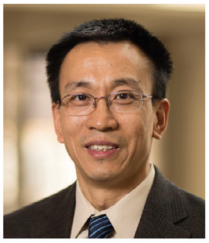
MD Anderson Cancer Center

For excellent and sustained statistical research and collaboration in the analysis of observational longitudinal cohort studies and chronic disease research and for outstanding service to the profession.

ROBERT A. OSTER

University of Alabama at Birmingham

For notable contributions to the health science research community in the field of statistics through diverse collaborations; for furthering statistics education in the health sciences; and for service and exceptional leadership to the ASA and other professional societies.



YONGMING QU

Eli Lilly and Company

For substantial and sustained contributions to statistical methodology for clinical trials and drug development; for significant influence and impact on bringing new drugs from research to market; and for excellent service to the profession.

MICHAEL ROSENBLUM

Johns Hopkins University Bloomberg School of Public Health

For outstanding contributions to statistical methodology and applications, especially with respect to the adaptive design and optimal analysis of randomized trials.

RICHARD CONRAD ZINK

TARGET PharmaSolutions, Inc.

For leadership and dedication to the Biopharmaceutical Section, including pioneering novel channels of communication; for exemplary contributions to the ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop; and for commitment to students through the scholarship award.

MINUTES OF EXECUTIVE COMMITTEE MEETING (SPRING METING) HELD VIRTUALLY ON MARCH 16, 2020

Janelle K Charles (BIOP Secretary, PPD)

Attendees: *Amit Bhattacharyya, Bruce Binkowitz, Thomas Birkner, Emily Bulter, Janelle Charles, Kun Chen, Freda Cooner, Alex Dmitrienko, Abie Ekangaki, Jennifer Gauvin, Xin Huang, Weili He, Stephine Keeton, Rakhi Kilaru, Judy Li, Lisa Lupinacci, Ilya Lipkovich, Jingyi Liu, Ted Lystig, Elena Polverejan, Veronica Powell, Erik Pulkstenis, Jane Qian, Yongming Qu, Lei Nie, Steve Novick, Yabing Mai, Peter Mesenbrink, Brian Millen, Jonathan Moscovici, Melvin Munsaka, Mark Rothman, Kyle Wathen, Xiaofei Wang, Steve Wilson, Wei Zhang, Richard Zink*

1. Introductions and Welcome
Bruce Binkowitz

Bruce provided the opening remarks, attendance was recorded, and the meeting was called to order. He encouraged the EC to provide feedback about this second virtual EC meeting. If agreed, the Manual of Operations and Charter will be updated to formalize the annual spring virtual meetings.

He led discussions about potential leadership and outreach activities for the Section. Additionally, he presented excerpts from the 2020 Section Treasurer Guidelines with possible initiatives to benefit members when a Section has a large balance for over 3 years (see *Attachment*). He proposed the creation of a temporary committee to work with the funding committee to generate ideas for spending money. The committee, preliminarily named of the "We need to spend our money" committee, should report out at the JSM EC Meeting. Brian Millen volunteered to serve on this committee.

Action Item: EC members interested in serving on this committee to email Bruce and Janelle.

2. Secretary's Report
Janelle Charles

Janelle reviewed the pending action items from the Transition Meeting minutes. There were no requested changes to the minutes.

A motion was made and seconded to approve the EC Meeting minutes.

3. Treasurer's Report
Jane Qian

Jane presented the 2019 year-end report and the 2020 budget; see Attachment. She noted that the planned spending for the 40th Anniversary Committee was not included in the 2020 budget.

A motion was made and seconded to approve the budget.

4. Chair-Elect Report
Weili He

Weili informed that Bo Huang (Pfizer) and Gene Pennello (FDA) will be the 2021 Workshop Co-chairs and presented the other 2021 appointments to date; see *Attachment*. She noted that Lisa Lupanacci will be the Chair of the Leadership in Practice Committee and not Abie Ekangaki as reported on the *Attachment*. All the remaining appointments are to be confirmed during the JSM EC Meeting.

5. Past-Chair Report
Richard Zink

Richard welcomed the newly elected EC members and thanked the outgoing members for their service. He presented the open positions for the 2021 elections as follows: Chair-Elect, Publication Officer, Program Chair Elect, and Council of Section Representative. Members were asked to contact Richard if interested or have recommendations for the election slate.

6. Funding Committee
Richard Zink

Richard informed that the Funding Committee has provided funding in amount of $200 for the Meeting within a Meeting that provides statistical training at JSM for high school and middle school teachers. He also

informed that there is a Beyond AP Statistics Workshop also held at JSM.

He proposed that annual funding be provided to support registration for these meetings. The EC discussed that the funding would be offered in the amount of $2500 with the option that the organizers may submit additional request if more funds are needed.

A motion was made and seconded to approve funding in the amount of $2500 annually.

Action Items:

1. Jane to add this funding as a line item to the budget.
2. Richard will contact Anna Nevius, organizer for the two meetings to inform of this funding and any follow-up actions.

7. 40th Anniversary Committee Jennifer Gauvin

Jennifer presented the members of the Committee and activities being planned for the 40th anniversary celebrations; see *Attachment*. The logo competition, leg by Meg Gamalo-Siebers, is expected to start by Fall 2020. Jennifer also summarized plans for industry sponsorship and advertisement. The budget proposal will be presented at the JSM EC Meeting.

The EC discussed ideas for engaging the Program Chair, Workshop Co-chairs, and Section Chair planning or executing the 40th anniversary activities in 2021. Jennifer noted that these are still being discussed within the Committee. The EC also discussed that logistical arrangements, including space reservations and contracts, should be confirmed in advance; particularly, for the Workshop.

Action Item: Jennifer to follow-up with Kristin Mohebbi (or Kathleen Wert) at ASA for logistical planning for the Workshop.

8. Membership Committee Richard Zink for Kun Chen

The Membership Committee is developing the survey with the plan to disseminate in Summer 2021. The Committee plans to publish the survey results in the BIOP Report in Fall 2021. The survey is usually administered every three years.

The EC discussed the possibility of having the survey administered earlier than summer in 2021 to capture results prior to the 40th anniversary activities.

Action Item: Richard to follow-up with the Membership Committee about the feasibility for earlier administration of the survey.

9. Fellows Committee Ilya Lipkovich

Ilya presented the current membership of the Fellows Committee and summarized the goals of the committee; see *Attachment*. The Committee will prepare an article about the ASA Fellows nomination process and testimonials from past ASA Fellows. The article will be published in the Spring Issue of the BIOP Report. A follow-up webinar on the "best practices" for the ASA Fellows nomination process is tentatively planned in August-September 2020. He informed that Stephen Ruberg and Christy Chuang-Stein were invited speakers.

The EC discussed the possibility establishing a "hot-line" for potential nominees to seek assistance in preparing their dossiers or reviewer pool for sponsors to gain feedback in preparing sponsorship packages. Illya noted that the Committee is continuing to investigate the potential challenges of a "hot-line", including limiting the number of dossiers sent for review as well as maintaining confidentiality of documents.

10. ASA BIOP Regulatory-Industry Statistics Workshop

2019 Workshop Updates

Judy Li and Renee Rees

Renee provided highlights and lessons learnt from the 2019 Workshop: "From Small Data to Big Data, From RCT to RWE, the Impact of Statistics"; see *Attachment 1*. She presented recommendations to the future planning committees, which included to explore opportunities to increase check-in efficiency and to find optimum time for taking photos of the student travel grant winners to avoid overlap with the poster session.

The EC discussed the advertising, process, and review of the submissions for the Statistics in Biopharmaceutical Research (SBR) Special Issue. Judy and Renee informed they use an objective evaluation process and that the process is time consuming to evaluate the submissions.

Action Item: Renee and Judy to contact Weili and Bruce after reviews are completed if they would like to provide updates for process improvements or lessons learnt at the JSM EC meeting.

2020 Workshop Updates
Yabing Mai and Thomas Birkner

Yabing provided an update on that the planning is on track for the 2020. They are working closely with ASA for any potency contingency planning due to the coronavirus. He summarized the planned sessions to be included in the 2020 program; see *Attachment 1*. The program will include BIOP involvement from the Leadership in Practice Committee, Mentoring Committee, and the Real World Evidence Scientific Working Group Committee.

11. Best Contributed Paper Award Freda Cooner

Freda presented the voting results and six award winners in 2019; see *Attachment 1*. Four out of the six winners accepted invitations to present a BIOP webinar. To date, Devan Mehrotra has won 9 awards and Brian Wiens 3 awards.

The EC discussed the current Hall of Fame process, which stipulates that once an individual wins 10 awards they enter into the Hall of Fame and are not eligible for future best paper awards. EC members shared concerns that ineligibility to win future awards may discourage participation. It was noted that, in general, JSM presentations are not solely given with intent of winning an award.

Action Item: Freda and Bruce to discuss the Hall of Fame process and update the EC at a future meeting; particularly whether awards may be won after an individual enters the Hall of Fame.

Post-meeting Note: Freda and Ted met with the Core EC following the meeting. The following decisions were made:

- There will be no "Hall of Fame" award based on a cumulative amount of awards. There will be no exclusion from winning based on the number of previous awards that a presenter has won.
- There should be recognition of best Contributed Paper winners, ideally annually in the JSM program. The Best Contributed Paper Award Committee to discuss how to do this for BIOP as a first step, and report back to EC (at JSM meeting or transition meeting). If a method of recognition settled and agreed, this can be included in the Manual of Operations revision in 2020.

12. Leadership in Practice Committee (LipCom) - Lisa Lupanacci

Lisa introduced members of the Committee and reminded the EC of the goals of the LipCom. She informed of the planned 2020 activities, which included co-hosting a leadership panel discussion session and a half-day leadership workshop. Additionally, plans are underway to collaborate with Mentoring Committee for efforts like podcasts about mentoring, etc.

The EC discussed encouraged LipCom to consider possibility of hosting a separate leadership workshop in the future.

13. Non-Clinical Biostatistics Working Group Xin Huang

See *Attachment 1* for details.

14. Program Chair Report Stephine Keeton

Stephine reported that there has been an increase in the number of session proposals. For the invited program there were 30 invited session proposals, of which, 5 were selected for JSM 2020. She noted that one of the invited sessions will be hosted by the Real World Evidence Scientific Working Group. See *Attachment 1* for more details.

The EC discussed opportunities to engage students in the round table sessions, e.g. advertising through academic email lists. It was noted that students may need funding to pay for round table attendance. Additionally, EC members suggested that a mentoring roundtable could be a topic of interest to students.

Action Item: LipCom and the Mentoring Committee to collaborate on round table discussions or activities to engage students.

15. Outreach and Collaboration Committee Amit Bhattacharyya

Amit shared an opportunity for BIOP to collaborate with the Section on Teaching Statistics. The content may be tailored to suit topics of interest the BIOP members. The Committee to meet with the Section on Teaching and follow-up with the EC with more information at a future meeting.

He informed of ongoing discussions with PSI for hosting joint PSI/BIOP webinars. One such webinar that is being developed focuses on regulatory perspectives applicable to United States and Europe.

He also presented a proposal by the Bay area Biotech-pharma Statistical Workshop (BBSW) to organize and co-host a mini-symposium on November 4 2020 with BIOP. The proposal is for participation from BIOP and not for any financial support for this event. BBSW will provide the financial and logistical management on this event. See *Attachment 1* for more details. The BBSW would like to have two representatives from BIOP to join a four-member organizing committee.

Amit will be passing leadership of this committee to Matilde and other members in the next year.

A motion was made and seconded to organize and co-host the mini-symposium.

Action Item: EC members to contact Amit if interested in participating in the mini-symposium. Amit to follow-up with Matilde for possible FDA participant to represent BIOP on the organizing committee for this mini-symposium.

16. Mentoring Committee Bruce Binkowitz for Juliet Ndukum

Bruce presented the recent and future activities of the Mentoring Committee; see *Attachment 1*. The EC discussed opportunities for providing more technical mentoring and possibility for scientific working group co-chairs or members to serve as mentors.

Action Item: Wei Zhang to discuss with Jennifer Gauvin and Jerry Wang, co-chairs of the Scientific Working Group (SWG) Committee, the idea of engaging the SWGs in the mentoring activities and formalize a proposal for discussion at a future EC meeting.

17. Best Student Paper Award Haoda Fu

There were no updates presented.

18. Scholarship Award Brian Millen

Brian summarized the applications received to date; see *Attachment 1* for details. He noted that the number of applicants was reduced from last year. The Scholarship Award Committee will investigate methods to improve application numbers in the next year.

19. Contributed Posted Award Jingyi Liu

Jingyi informed that he will be rotating off the committee and Bo Huang will chair this committee. New members will be identified at a future EC members. From JSM 2019, there are 3 awardees and 2 honorable members. The committee is awaiting confirmations if all winners will present at JSM 2020. Poster submissions for this year are ongoing.

20. Scientific Working Groups Jennifer Gauvin and Jerry Wang

Jennifer summarized the SWG Health Check reports; see *Attachment 1* and *Attachment 2*. She informed there were no updates progress updates provided from the Re-randomization SWG. The EC discussed process for disbanding inactive SWGs and whether this information should be added to the Manual of Operations. The EC also discussed that possibility to have term limits for SWGs wherein a member with the technical expertise may be eligible to become chair of an SWG.

Action Items:

1. Jennifer and Jerry to update at a future EC meeting proposed guidelines for disbanding an SWG that is inactive.
2. Bill Wang to discuss with Jennifer and Jerry the topic of term limits for update at a future EC meeting.

21. Publications/Communications Update Yongming Qu, Wei Zhang, Xiaofei Wang, Richard Zink, Jonathan Moscovici.

See *Attachment 1* for details.

22. Closing Janelle Charles, Bruce Binkowitz

Due the technical issues, the meeting was prematurely ended, and the review of action items was completed by email.

MINUTES OF EXECUTIVE COMMITTEE MEETING (JSM MEETING) HELD VIRTUALLY ON AUGUST 3, 2020

Janelle K Charles (BIOP Secretary, PPD)

Attendees: *Amit Bhattacharyya, Hiya Banerjee, Bruce Binkowitz, Thomas Birkner, Emily Butler, Janelle Charles, Kun Chen, Freda Cooner, Alex Dmitrienko, Abie Ekangaki, Jennifer Gauvin, Margaret Gamalo-Siebers, Weili He, Stephine Keeton, Rakhi Kilaru, Lisa Lupinacci, Ilya Lipkovich, Jingyi Liu, Ted Lystig, Elena Polverejan, Veronica Powell, Jane Qian, Yongming Qu, Lei Nie, Steve Novick, Yabing Mai, Peter Mesenbrink, Brian Millen, Jonathan Moscovici, Melvin Munsaka, Matilde Sanchez-Kam, Kyle Wathen, Xiaofei Wang, Lanju Zhang, Wei Zhang, Richard Zink*

1. Introduction and Welcome
Bruce Binkowitz

Bruce welcomed the EC to the first ever virtual 2020 JSM EC meeting. He reminded the EC of the upcoming virtual Open JSM Business Meeting and virtual Transition meeting in September. Attendance was recorded and the meeting was called to order.

2. Secretary's Report
Janelle Charles

Janelle reviewed the pending action items from the 2020 Spring EC Meeting minutes. The EC noted a correction to the minutes was needed to remove the named Chair of the Scholarship Committee as that decision had not been made at the Spring meeting.

A motion was made and seconded to approve the EC Meeting minutes with the correction.

Post-Meeting Note: The final 2020 Spring EC Meeting minutes was circulated to the EC after the meeting and saved to the Google Drive.

3. Treasurer's Report
Jane Qian

Jane presented the budget for the period ending June 30, 2020 showing a balance of approximately $390,000; see *Attachment*. She noted that more information of the funds needed for the 40th Anniversary Committee was to be obtained for inclusion in the 2021 budget proposal.

4. Chair-Elect Report
Weili He

Weili presented the incoming BIOP Elected Officers from the 2020 ASA elections: Alan Hartford (Chair-Elect), Freda Cooner (Program-Chair Elect), Inna Perevozskaya (Secretary), and Mark Levenson (Council of Sections Representative). She also presented the committee appointees for 2021; see *Attachment*.

5. Past-Chair Report
Richard Zink

Richard presented the open positions for the 2021 elections as follows: Chair-Elect, Publication Officer, Program-Chair Elect, and Council of Section (CoS) Representative. Members were asked to contact Richard if interested or have recommendations for the election slate.

He informed EC that the Manual of Operations was available review and comments; see Attachment for Google Drive link to document. EC members may provide comments as tracked changes directly on the document or emailed to Richard at *richard.c.zink@gmail.com*.

6. Funding Committee
Richard Zink

Richard informed that there have been no new funding requests since the 2020 Spring EC Meeting.

7. Best Student Paper Award
Lanju Zhang

Lanju presented the four award winners (3 monetary awardees and one honarable mention); see *Attachment*. Richard reminded that the Student Award page on the BIOP website needs to be updated in the fall for the upcoming paper submissions period.

Action Item: Lanju to send all updates for the website to Steve Novick, webmaster, for posting.

8. Non-clinical Biostatistics (NCB) Working Group - Steve Novick

Steve informed the EC of the five workstreams within the NCB Working Group; see *Attachment*. Among these workstreams, the non-clinical p-value workstream is preparing a manuscript focused on p-values as pertains to non-clinical research. This workstream is chaired by Stan Altan and the manuscript is expected to be published in the BIOP Report in Q4 2020. He also summarized the activities of the non-clinical Bayesian workstream. Steve suggested that the incoming Program Chair solicit participation from the NCB for next year's JSM Program.

He reminded that the 2021 NCB Conference will be held at Rutgers University in 21-23 June 2021 with a theme of Nonclinical Statistics in the Age of Data Science.

The EC discussed that Peter Mesenbrink is also exploring the *p*-value topic, so there is a potential to collaborate with the NCB p-value workstream for upcoming manuscripts in the BIOP Report.

Action Item: Steve to connect Peter Mesenbrink with Stan Altan.

9. Mentoring Committee
Sourav Santra

Sourav presented the feedback from the mentoring committee survey and the upcoming activities in 2020-2021; see *Attachment*. He noted that there is a need for 5 more mentors to participate in the 2020-2021 Mentoring Program.

He informed that the Mentoring Committee will collaborate with the Leadership in Practice Committee (LipCom) to host a podcast series targeted to raise the awareness of mentoring.

Action Items:

1. EC members who are interested in mentoring are encouraged to reach out to Sourav.
2. Bruce to announce opportunity for mentors at the JSM Open Business Meeting.

10. Fellows Committee
Ilya Lipkovich

Ilya presented the current membership of the Fellows Committee and summarized the activities of the committee; see *Attachment*. Brenda Crowe will serve as the Fellows Committee Chair. He presented the BIOP members that will be awarded as ASA Fellows this year.

The Committee published an article comprising a collection of reflections by members of the ASA Fellows Committee, nominators, and recent nominees in the Summer Issue of the BIOP Report and June issue of Amstat news. A follow-up webinar on the "best practices" for the ASA Fellows nomination process will be held on 17 September 2020. He informed that Stephen Ruberg, Paul Gallo, Ivan Chan and Christy Chuang-Stein were invited speakers.

Ilya informed that the Committee is planning to initiate a working group of experienced sponsors to perform brief reviews of nomination dossiers.

Bruce congratulated Richard Zink and Yongming Qu new ASA Fellows.

Action Item: EC members interested in serving on the Fellows Committee to contact Ilya Lipkovich.

11. Outreach and Collaboration Committee
Michelle Zhang and Matilde Sanchez-Kam

Michelle and Matilde summarized the upcoming activities of the Committee; see *Attachment*. She informed of ongoing collaborations with PSI for hosting joint PSI/BIOP webinars. One such webinar will be held in November 2020 on the topic of estimands.

12. 40th Anniversary Committee Richard Zink and Lisa Lupinacci

Lisa summarized the activities being planned for the 40th anniversary celebrations; see *Attachment*. The Committee has developed a postcard that with 'Save the Date' that will be included in the 2020 Regulatory Industry Statistics Workshop package. She noted that planning activities at upcoming conferences is difficult considering the coronavirus situation. She mentioned that the Committee is considering session proposals for JSM and the Workshop.

Richard informed that after the Workshop he will meet Amy Farris from ASA to discuss sponsorship opportunities for the 40th anniversary celebrations.

The EC discussed whether a professional designer could be contracted to update the BIOP logo as part of the 40th anniversary activities. Bruce noted that this may be incorporated into the budget proposal from the 40th Anniversary Committee. Yabing Mai informed that a spot has been reserved for an activity associated with 40th anniversary celebration at the virtual happy hour at this year's Workshop.

Action Item:

1. EC members interested in volunteering for the 40th anniversary to contact Meg Gamala-Siebers. Weili will provide Meg with the BIOP volunteer list.
2. Yabing will provide details to the 40th Anniversary Committee for hosting an activity at this year's virtual happy hour.
3. Lisa will send the postcard to Bruce to share at the JSM Open Business Meeting.

13. 2020 ASA BIOP Regulatory-Industry Statistics Workshop Yabing Mai and Thomas Birkner

Yabing and Thomas updated the EC on the status of planning for the 2020 Workshop, which will be held virtually for the first time this year. The theme for this year's Workshop is Lead and Impact: Turning Innovation into Practice. They presented the highlights from this year's program, which will include a virtual happy hour; see *Attachment*. There will be free short courses: one for causal inference and one for complex innovative designs.

There were approximately 650 registrants as of 31 July 2020 with a break-even target of 800 registrants. All registrants will receive a care package delivered by mail. Additionally, Yabing and Thomas informed that approximately $40,000 had been received from sponsorships, including six principal sponsorships from AbbVie, Amgen, Covance, Merck, Penfield and Novartis. The EC discussed that there might be some surplus from sponsorships; however, the virtual platform used for hosting the Workshop costed about $70,000. Any surplus from the Workshop will be split between ASA and BIOP.

Thomas and Yabing thanked the EC for all the support for the Workshop.

14. Best Contributed Paper Award Committee Freda Cooner

Freda presented the link to the online evaluation form for this year's voting. There are 17 topic-contributed and 14 paper-contributed sessions; see *Attachment*. She informed the EC of the six award winners from 2019.

She also presented ideas being explored by the committee, including proposal for roundtable sessions with the award winners, footnotes in JSM program book to recognize past winners, 1-1 mentorship for presentation skills, short courses or sessions to improve communication.

Action Item: The Committee to formulate proposal for engaging past winners to present at a future EC meeting.

15. Council of Section Representative Ted Lystig and Brian Millen

Ted presented updates from the ASA Council of Sections governing board meetings and upcoming activities that may be of interest to BIOP; see *Attachment*. The BIOP Charter is up for revision in 2021. A few key points for revision include the timing of meetings linked to ENAR and voting by ad hoc versus elected members. The revised Charter will first be reviewed and approved by the EC before voted on by the BIOP membership during the 2021 ASA election period. Additionally,

BIOP is eligible to present the ASA Award in 2021. The EC discussed that a committee needs to be formed to establish criteria for selecting the award winner.

Ted and Brian informed that the ASA COSGB is collecting feedback on improvements for remote tools, ASA website.

Action Items:

1. CoS representatives to provide proposed updates to the BIOP Charter as well as nomination process for ASA award outstanding section service for EC review at the Transition EC meeting. Aim to provide 1-2 weeks prior to meeting for advance review.
2. EC members to provide feedback to CoS representatives of any challenges accessing remote tools (e.g. WebEx) or with the ASA website.
3. Bruce Binkowitz to add agenda item for EC Transition Meeting for detailed discussion about modernizing the ASA/BIOP webpages. BIOP CoS representatives can share this feedback from this discussion to ASA COSGB.

16. Scholarship Award Committee Brian Millen

Brian presented the committee members and scholarship award winners; see *Attachment*.

17. Committee to Address Excess BIOP Cash on Hand - Brian Millen

Brian presented the initial membership and important elements of the committee's responsibilities; see *Attachment*. The proposed name is the Next Generation Stewardship Committee. The Committee is expected to bring forward spending proposals at least annually to the EC for vote.

A motion was moved and seconded for the creation of the committee. The Committee details to be added to the Manual of Operations.

Action Item: Weili to add the Next Generation Stewardship Committee to the committee list and coordinate with Steve for updates to the webpage.

18. Leadership in Practice Committee (LipCom) - Abie Ekangaki

Abie presented the committee members and provided updates on the committee activities; see *Attachment*. The activities include a half day interactive leadership workshop to be held at the 2020 Regulatory Industry Statistics Workshop. He informed that LipCom is collaborating with the Mentoring Committee to host a mentoring podcast series that will include interviews with mentee-mentor pairs. The Committee is exploring other podcasts ideas; Peter Mesenbrink informed that he has a podcast on soft skills for statisticians and suggested this may be added to the LipCom podcast series. Additionally, he informed that the LipCom website was being developed.

The EC discuss that the incoming Chair-Elect, Alan Hartford, will need to appoint a new LipCom member.

Action Item: Abie Ekangaki and Peter Mesenbrink to discuss potential collaboration
on leadership podcast and update at future EC meeting.

19. Program Chair Report Stephine Keeton

Stephine reported that there were 30 invited session proposals, of which 5 were sponsored by BIOP and 9 co-sponsored by BIOP. She informed that with JSM being virtual the speed sessions had been removed from the program. She presented the 2020 virtual JSM updates; see *Attachment*.

20. Scientific Working Groups Jennifer Gauvin and Jerry Wang

Jennifer informed there have been no new updates or proposals for new scientific working groups. She reported by Yeh-Fong Chen from the FDA had interest

in revamping the re-randomization working group. The EC discussed that if there is an option to revamp a SWG then the criteria for decommissioning an inactive working group also needs to be established. This topic will be revisited at a future EC meeting.

Action Items:

1. Scientific Working Group (SWG) Committee to work with Yeh-Fong Chen to formalize proposal for Re-randomization SWG and provide proposal to EC for review in advance of Transition EC meeting. SWG Committee will also work on process for decommissioning inactive SWGs for review at the Transition EC Meeting. Aim to provide 1-2 weeks prior to meeting for advance review.
2. Bruce Binkowitz to follow-up with Bill Wang regarding the term limit for SWG members.

21. Membership Committee Kun Chen

Kun presented the members of the Membership Committee and the plans for accelerating the timeline to provide results from the membership survey; see *Attachment*. This accelerated timeline means that the survey results will be disseminated in April-May 2021 and available for the 40th anniversary celebrations.

A motion was moved and seconded to approve the accelerated timeline for the membership survey activities.

22. Contributed Posted Award Jingyi Liu

See *Attachment*t for details.

23. Publications/Communications Update Yongming Qu, Wei Zhang, Xiaofei Wang, Richard Zink, Jonathan Moscovici.

See *Attachment* for details.

Wei informed that there were no webinars planned from July to September.

Richard summarized the podcasts planned for 2020. The EC discussed the requests to have podcasts transcribed. Richard to explore these costs and update at a later EC meeting.

Steve informed of the major updates to the webpage, including details for the 2021 NCB Conference and content for the Statistical Methods in Oncology Scientific Working Group. He raised concerns about use of outdated web-building tools. The EC discussed whether the upgrades for the web tools could be funded by BIOP since the website is a micro-site housed within the ASA webpage. The EC discussed that an approach may involve BIOP providing the specifications to ASA for the web improvements.

Action Item: COS representatives to share feedback on the outdated web tools to the ASA COSGB.

24. Any Other Business and Action Item Review

Bruce informed the EC that Jiajun Liu has been appointed as the BIOP Survey Monkey manger to oversee the license and maintenance of the Survey Monkey account.

Janelle reviewed the major action items and informed that the complete list of action items will be emailed after the meeting.

25. Closing Bruce Binkowitz

Bruce thanked EC members for their participation and the meeting was adjourned.

CONFERENCES

2021 Joint Statistical Meeting

The Joint Statistical Meetings (JSM) is the largest gathering of statisticians and data scientists in North America, and will be held at Seattle, Washington in August 7-12, 2021. JSM 2021 covers topics ranging from statistical applications to methodology and theory to the expanding boundaries of statistics, such as analytics and data science. JSM also offers a unique opportunity for statisticians in academia, industry, and government to exchange ideas and explore opportunities for collaboration. Beginning statisticians (including current students) can learn from and interact with senior members of the profession. More information can be found at *https://ww2.amstat.org/meetings/jsm/2021*.

2021 ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop

The ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop is sponsored by the ASA Biopharmaceutical Section in cooperation with the FDA Statistical Association. The conference lasts two days, with invited sessions co-chaired by statisticians from industry, academia, and the FDA. In addition, short courses on related topics are offered the day prior to the workshop. The workshop is planned to hold in September 21–23, 2021, Rockville, MD. More information can be found at *https://ww2.amstat.org/meetings/biop/2021/*.

Bay area Biotech-pharma Statistical Workshop on COVID-19

Organizer: Bay area Biotech-pharma Statistical Workshop

Date: November 5, 2020

Theme: Balancing Speed and Evidence in Developing COVID-19 Therapies

More information can be found at *https://www.bbsw.org/symposium-2-nov-5*.