Volume 29, No. 2 • Summer 2022

BIOPHARMACEUTICAL REPORT

Real World Evidence

Chair: Alan Hartford **Editors:** Herbert Pang, Ling Wang, Kristi L. Griffiths

CONTENTS

Notes from the editors

Time flies as we are well into the second half of 2022, we continue to adapt to the new normal with the evolving COVID-19 pandemic. From what we hear, many of us are able to get back to more in-person activities and travel. This summer will mark the return of the in person JSM after two years in hiatus and in September there will be the Regulatory Industry Statistics Workshop (RISW) in Maryland. In the past year or so, we have seen a growing number of "real-world evidence" (RWE) guidances from the FDA and other regulatory agencies and research institutions. Our second issue of 2022 will be dedicated to the theme of RWE with featured articles from industry, government, and academia.

Under the RWE theme, we open with an update from the ASA BIOP working group on RWE, written by **Mark Levenson** (FDA) and **Weili He** (Abbvie). Next, we feature an article by **Thomas Brown** (Syapse) and members of the RWE Alliance. They introduce the coalition and elaborate on their mission which is to harness RWE to improve the lives of patients. This is followed by an article by a cross-industry collaboration. **Binbing Yu** (AstraZeneca), **Qing Li** (MorphoSys), and **Harry Yang** (Fate Therapeutics) provide a brief overview of real-world data and RWE in drug development from industry perspectives. Next up is a feature article contribution from **Xiang Zhang** (CSL Behring) and **Douglas Faries** (Eli Lilly), on the topic of the role of statisticians in the RWE era. They provide a vision towards improving RWE to inform healthcare decision making. Our fifth and sixth feature articles are from NIH grantees of RWE projects. **Shu Yang** (NCSU) and **Xiaofei Wang** (Duke) outline approaches for real-world data-integrated randomized clinical trial analysis, which is the primary focus on their NIHR01 project funded by the NIA. This is followed by an article written by **Chenqi Fu** (Penn State), **Herbert Pang** (Genentech) and **Jiawen Zhu** (Genentech), on evaluating the impact of different randomization ratios in designing hybrid control trials on their NIHU01 project funded by the FDA. After the featured articles, we have a contribution from non-clinical statistics written by **Aili Cheng** (Pfizer) and her Pfizer colleagues. They discuss chemistry, manufacturing, and control (CMC) statistical support for COVID-19 vaccine development. We would like to highlight ASA BIOP's effort to facilitate the career development of statisticians, data scientists and quantitative researchers, and with that we had a fireside chat with **Sandeep Menon** (Pfizer) on leadership development. Sandeep shares a lot of great experiences and advice for us, especially in the area of going beyond statistics in your career. Later in this issue, you will find a summary report from a virtual discussion organized by the ASA BIOP Statistical Methods in Oncology Scientific Working Group, the FDA Oncology Center of Excellence, and LUNGevity Foundation. The topic of discussion is "Time-to-event Endpoints in Cancer Trials in the Presence of Non-Proportional Hazards". The final article is from the CSM/QTL Special Interest Group on "Central Statistical Monitoring – why we need to know more" by **Tim Rolfe** (GSK), **Susan Talbot** (Amgen), **Rakhi Kilaru** (PPD), and **Sharon Love** (UCL). The 40th year anniversary of the BioP section will be celebrated in person at the 40+1 events taking place at the JSM and the RISW, you may check out the flyer prepared by the organizing committee. In the last section, We also share an update of upcoming conferences which may be of interest to the BIOP community. The editors would like to thank all the authors and interviewee of the articles for their time and contributions, and wish that everyone enjoys this second issue of the BIOP Report in 2022.

ASA BIOP REAL-WORLD EVIDENCE SCIENTIFIC WORKING GROUP UPDATE

Mark Levenson (FDA CDER) and Weili He (Abbvie)

As reported in the BIOP report in the summer of 2019, a Scientific Working Group (SWG) on real-world evidence (RWE) was formed in 2018. The group's purpose is to ensure that statisticians are well-versed in this rapidly developing field and the field benefits from statistical research and participation. Since the 2019 report, there has been many developments in RWE and the SWG has made substantial progress. In this issue, we provide some updates on recent regulatory guidances in RWE and the progress of the SWG. For the background and rationale in setting up the SWG along with the SWG specific objectives, please see the first report.

Since the last update of the SWG in the summer of 2019, we have seen the release of several new draft or final guidances from both FDA and EMA, as listed below.

FDA:

- Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products (September 2021)
- Data Standards for Drug and Biological Product Submissions Containing Real-World Data (October 2021)
- Real-World Data: Assessing Registries to Support Regulatory Decision-Making for Drug and Biological Products (November 2021)
- Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products (December 2021)

EMA:

- Guideline on Registry-Based Studies (October 2021)

These guidelines provide additional specificity on how to assess fit-for-use RWD sources, data standards, and general considerations on the use of RWE and RWD for regulatory decisions.

Our phase 1 effort started in early 2018 and completed in the summary of 2020. The SWG submitted three manuscripts, as shown below, which were all accepted and published in the Statistics in Biopharmaceutical Research (SBR) journal in early 2021.

1. Biostatistical Considerations When Using RWD and RWE in Clinical Studies for Regulatory Purposes: A Landscape Assessment
2. The Current Landscape in Biostatistics of Real-World Data and Evidence: Clinical Study Design and Analysis
3. The Current Landscape in Biostatistics of Real-World Data and Evidence: Causal Inference Frameworks for Study Design and Analysis.

Our phase 2 effort started after the conclusion of the phase 1 effort in the summer of 2020. We divided the SWG into three subteams, working on the following three topics:

- Team 1: Estimands - From Concepts to Applications in Real-World Setting
- Team 2: Statistical Consideration for Fit-For-Use Real-World Data to Support Regulatory Decision Making in Drug Development
- Team 3: Examples of Applying Causal Inference Roadmap to RWE Clinical Studies

The membership in phase 2 consisted of representatives from regulatory agencies, industry, and universities. Additional information on the SWG can be found here. The approach in phase 2 followed the same approach in phase 1: Each subteam conducted a focused literature review to address the following four questions for a given topic as appropriate: its regulatory context, a precise problem statement, a summary of current approaches, and a gap analysis. Each team took turns to report their findings during monthly tele-conferences. After nearly 2 years of work by the SWG, we are happy to report that the SWG subteams have prepared three manuscripts, submitted or resubmitted to SBR in Q2 2022.

We are currently planning on the phase 3 effort and will target to start that phase of the work in the coming months. We want to thank the lead Editor of the BIOP Report, Dr. Herb Pang from Genentech, for his invitation for the update. ■

INTRODUCING THE REAL-WORLD EVIDENCE ALLIANCE:

A COALITION DEDICATED TO HARNESSING REAL-WORLD EVIDENCE TO IMPROVE THE LIVES OF PATIENTS

Thomas Brown (Syapse), Marni Hall (IQVIA), Tara Isherwood (Syneos Health), Michelle Leavy (OM1), Irene Nunes (Flatiron Health), Lowell Schiller (Aetion), Lauren Silvis (Tempus), Aracelis Torres (Verana Health)

Real-world data (RWD) and real-world evidence (RWE) have advanced a number of therapeutic options for patients over the past several years. RWD can be collected from a variety of different sources: electronic health records; administrative claims and billing data; product and disease registries; and personal devices, wearables, and health applications. The curation, transformation, and analysis of these data produce RWE—a new source of clinical evidence about a medical product's use, including its benefits and risks, to inform decisions made by regulators about these products.

Regulators and policymakers have recognized the need for new RWE policies to advance innovation crucial for continued improvements in patient care. Leading RWD and analytics organizations have come together to form a new coalition—the RWE Alliance—to serve as a unified, expert voice to ongoing policy conversations. This article introduces the RWE Alliance and provides an overview of our policy priorities.

I. Who We Are and What We Do

The RWE Alliance formed in 2021 in response to the need to harness the collective insights of leaders in the RWD/RWE space to inform policymaking in the rapidly evolving area of healthcare and technology related to the generation and use of RWE. Congress included important language about the role of RWD/RWE in regulatory decision making in the 21st Century Cures Act (2016) and in the Food and Drug Administration Reauthorization Act (FDARA) (2017). The COVID-19 pandemic further highlighted to the U.S. Food and Drug Administration (FDA), Congress, and stakeholders across the healthcare sector that RWE has the potential to increase knowledge about vaccines, drugs, devices, and diagnostics and make a real difference in the lives of patients.

Although various groups and trade associations have advocated for policies to advance the use of RWE in regulatory decision making, the RWD and analytics organizations that generate, transform, and analyze RWD had no coordinated expert voice in these policy discussions. Thus, five RWD and analytics organizations—Aetion, Flatiron Health, IQVIA, Syapse, and Tempus—joined forces to create the RWE Alliance in May 2021. In April 2022, we announced the addition of five more organizations—ConcertAI, OM1, Syneos Health, Verana Health, and Verily—to the Alliance. The RWE Alliance engages with FDA and Congress to advocate for policies that benefit patients by advancing the use of RWE in regulatory decision making.

II. Objectives of the RWE Alliance

The Alliance envisions a future in which data collected in everyday clinical practice will be used to generate evidence that informs regulatory decision making. To advance this goal, we have identified four policy priorities to guide our policy work.

1. Advancing FDA's RWE Framework

We support FDA's efforts to develop policies to advance its RWE Framework. As part of the 21st Century Cures Act, Congress instructed FDA to publish a framework on the use of RWE for regulatory decision making. Since then, FDA has published a draft

framework and issued guidance on key RWD/RWE topics—including, most recently, on data derived from electronic health records, medical claims databases, and registries; data standards for submitting RWD for FDA's review; and regulatory considerations for using RWD/RWE in FDA's decision making.

We support the ongoing work at FDA to advance the RWE framework and have shared our expertise by providing substantial comments on four recent FDA RWE guidance publications. We aim to ensure that FDA's RWE policies promote RWE across therapeutic areas and provide clear recommendations on the generation and use of high-quality RWD/RWE. We also believe it is crucial for regulators to remain flexible to accommodate technological and methodological advancements, given the rapid pace of innovation with respect to RWD/RWE.

2. Encouraging Use of RWE to Better Understand Treatment Effects in Underrepresented Populations

We support policies that promote the use of RWE to better understand how to treat underrepresented populations. Clinical trials do not always fully represent relevant patient populations in real-world settings, which can exacerbate disparities in healthcare access and treatment for underrepresented groups. RWE is well suited to provide FDA, healthcare providers, and patients with information on how treatments work for populations that clinical trial data do not capture.

3. Enhancing Opportunities for RWE Organizations to Consult with FDA

We seek to establish opportunities for RWE organizations to consult with FDA on issues relevant to the potential uses of RWE for regulatory decision making. As part of that work, we will share insights from the RWE industry with FDA and work with the Agency to establish opportunities for RWE organizations to engage the Agency on ways to improve RWD/RWE methodology and applications.

4. Increasing Communication on the Generation and Use of RWE

Finally, we aim to solidify best practices for the use of RWE for regulatory purposes to encourage widespread understanding of RWE's benefits and applications to patient care. We also support FDA's efforts to provide appropriate transparency about the Agency's review of RWE in marketing applications, as these public communications will also help advance best practices with respect to RWE for regulatory purposes.

III. To Learn More About Our Work

The RWE Alliance is excited about the potential that RWE has to transform the healthcare ecosystem and improve the lives of patients. If you would like to learn more about our work, we invite you to visit our website at *www.rwealliance.org*.

References

1. U.S. Food & Drug Administration, Framework for FDA's Real-World Evidence Program, *https://www.fda.gov/media/120060/download.*

2 RWE Alliance, Policy Developments, *https://rwealliance.org/rwe-policy-developments/*. ■

A BRIEF OVERVIEW OF REAL-WORLD DATA AND REAL-WORLD EVIDENCE IN DRUG DEVELOPMENT FROM INDUSTRY PERSPECTIVES

Binbing Yu (AstraZeneca), Qing Li (MorphoSys), Harry Yang (Fate Therapeutics, Inc.)

1. Introduction

Real world data (RWD) are data pertaining to patient health status and/or the delivery of health care collected from a variety of sources such as electronic health records (EHRs), claims and billing activities, pragmatic clinical trials, product and disease registries, patient-generated data including in home-use settings, mobile or wearable devices. Real-world evidence (RWE) is the clinical evidence regarding the usage and potential benefits or risks of a medical product derived from the rigorous analysis of RWD with proper analytical methodology (FDA, 2016). Spurred by the 21st Century Cures Act in the United States and similar policy efforts in other countries, RWD and RWE are transforming the drug development process towards a new patient-centric paradigm. From an industry perspective, we review how biopharmaceutical companies leverage RWD/RWE to make informed decisions, and expedite the drug development process. Following the drug development life-cycles, we present how RWD/RWE can complement the randomized control trials (RCTs) from drug discovery, through exploratory and confirmatory clinical trials, regulatory approval, to post-marketing phases. Some successful examples are provided. We also discuss the challenges and future directions of maximizing the value of RWD/RWE in pharmaceutical R&D.

RWE consists of information collected during routine clinical practice whereas randomized control trials (RCTs) are conducted in highly selective populations in well-controlled settings. RCTs can provide evidence on the efficacy and safety of a drug, and have been the gold standard for evidence generation supporting regulatory approval (Kim et al. 2018).

Table 1. Comparison of RCT and RWE

	RCTs	RWE
Outcome of interest	Efficacy/safety	Effectiveness/safety
Setting	Research	Real world
Patient population	Homogeneous	Heterogeneous
Population size	Small – moderate	Large - huge
Patient follow-up	Fixed	Variable
Treatment	Fixed	Variable
Attending physician	Investigator	Practitioner
Costs	High	Low
Generalizability	Low – moderate	Moderate - high
Control for bias	Design and conduct	Analysis

There are several major differences between RCT and RWE as shown in Table 1. For example, the primary objective of the RCT is to demonstrate efficacy and safety in a controlled environment. In contrast, the RWE is principally used to show the effectiveness of the treatment in diversified situations in the real-world setting. Although RCTs are considered as the gold standard for drug approval, there are several disadvantages of RCTs (Kim et al. 2018; Naidoo et al. 2021). While

the RCTs are often conducted in rigorously controlled conditions, thus can reduce bias and improve the internal validity of the analysis results, they also come with the drawbacks of high financial costs and long execution time. Often times, because of restrictive inclusion and exclusion criteria, RCTs do not account for the broader patient population encountered in routine clinical practice and other specificities, e.g. vulnerable populations, ethnic differences, comorbid conditions, concomitant drugs, and differences in lifestyles. Furthermore, RCTs are often of limited study duration and unable to assess long-term safety and effectiveness and the regular follow-up and close monitoring in most clinical trials do not reflect routine clinical practice. It is clear that RWE would complement traditional clinical trial data, especially in the assessment of safety and efficacy in the real-world settings.

2. Use of RWD/RWE in Drug Development

2.1.Preclinical and Drug Discovery Phase

In the preclinical and drug discovery phase, RWD can be used to depict the burden and epidemiology of diseases under study, to characterize the patients with unmet medical needs, and to understand the natural disease history of the target population. For example, the population-based cancer incidence, mortality, survival and prevalence data from the population-based cancer registries, e.g., the Surveillance, Epidemiology, and End Results (SEER) Program (NCI, 2022) and European Network of Cancer Registries (ENCR), provide authoritative and representative information about the burden and trends of cancer in the USA and Europe.

RWD can be used to evaluate the biomarker prevalence and discover the target for the development of personalized medicine. For example, the whole genomic sequencing (WGS) data allows researchers to identify variants that differ between the reference population and may indicate a higher risk of disease and/or likelihood of responding to a specific treatment. This can help prioritize early target discovery and select investigational drugs to be tested in the first human trials.

2.2. Clinical Trial Design, Operating and Analysis

Historically, RWD has been used to help determine the treatment effect and sample sizes for powering RCTs. With careful design and patient selection, the RWD can be used to emulate the target clinical trials (Hernan and Robins, 2016). RWD can be utilized in the planning and execution of clinical trials, including accelerated patient recruitment by applying trials inclusion/exclusion criteria against de-identified patient data from EHR databases to determine eligible patients, using analytics and selecting fast enrolling sites based on past performance such as the number of violations. Recently, RWD has been used to assess and enhance the inclusion and diversity of the under-represented patient population in clinical trials (FDA, 2020), and risk-based monitoring to mitigate data quality issues. RWD can also help assess the impact of inclusion and exclusion criteria on trial feasibility and inform the selection of site/country.

Pragmatic clinical trials (PCTs), conducted in real-world clinical practice settings, with typical patients and by qualified clinicians, can serve as a bridge between RWE and RCTs. In PCTs, investigators often relax inclusion criteria requirements and accept a broader and more representative patient population. However, the patients are still randomized to treatment and control groups. When properly designed and conducted, PCTs can both test the treatment effect and understand the differences of treatment effects in different health care settings. They can generate evidence to inform both regulatory and payer decision-making. For example, the DAPA-MI trial was a pioneering registry-based PCT that combines the RCT elements with innovative, real-world trial elements (Usman et al. 2022). The unique design features led to a higher recruitment rate and lower overall costs in comparison to conventional clinical trials.

Borrowing from the external data in the real-world setting and historical clinical trials has received increasing interest in drug development. Various statistical methods, including propensity-score matching, Bayesian dynamic borrowing to form a synthetic control arm in single-arm trials or to augment control arms of RCTs have been proposed and implemented (Ho et al. 2021). For example, the Medical Device Innovation Consortium (MDIC) published an External Evidence Methods (EEM) Framework which highlights the potential for incorporating data external to a clinical trial into the analysis of a medical device. (MDIC, 2022).

2.3.Registration and Market Application

RWE can be used to provide critical evidence for drug approval (Purpura et al. 2021). For example, FDA recently approved Prograf (tacrolimus) in combination

with other immunosuppressant drugs for preventing organ rejection in adult and pediatric patients receiving lung transplantation. The approval demonstrates that a well-designed, non-interventional (observational) study with reliable and relevant RWD, when compared to a suitable control, can be considered adequate for regulatory approval. As more drugs are approved by regulatory authorities either through the FDA orphan drug and breakthrough therapy designations or EMA Conditional Approval, using RWE to supplement the findings in RCTs helps avoid costly post-marketing trials and ensures early access. For example, Blinatumomab received accelerated approval for the treatment of Philadelphia chromosome-negative relapsed or refractory B-cell precursor acute lymphocytic leukemia based on a single arm trial. For this effort, historical control data for 694 patients were extracted from RWD (FDA, 2018).

2.4 Post Approval and Product Life-Cycle Management

RWE was traditionally used in the regulatory process through pharmacovigilance programs to understand the long-term safety of a drug or device in the real-world use. In 2008, the FDA launched the Sentinel Initiative to create a national electronic system, the Sentinel System, for medical product safety surveillance (Ball et al. 2016). The Sentinel System has been used to evaluate the risk of stroke after using antipsychotics along with other indications.

From the product life cycle management perspective, effective insights gleaned from RWD bring about payer value propositions. RWE can bridge the gaps in evidence to guide payer decisions. RWD can provide evidence not addressed by RCTs, such as long-term effectiveness and safety, head-to-head drug comparisons, cost-effectiveness analyses, medication use and adherence patterns, identification of relevant responder and non-responder patient subpopulations, and patient-reported outcomes (PROs) (Roberts and Ferguson, 2021). For example, the RWD from a retrospective cohort study showed promising results in reductions in HbA1c, weight, and insulin requirements for patients with type 1 diabetes who initiated a SGLT2 inhibitor adjunct to insulin. Individuals with higher baseline HbA1c and BMI demonstrated higher benefit (Palanca et al. 2022)

Last, RWE can be used to aid in the benefit-risk assessment of populations that are historically not included in RCTs. There have been several successful regulatory approvals for label expansion such as broadening the label to include a pediatric population or updating approval for chemotherapeutic agents that are used in combination with other treatments. Based on RWE, the FDA approved Ibrance (palbociclib) for the treatment of men with HR+, HER2 metastatic breast cancer and the Sapien 3 device for Transcatheter Aortic Valve Replacement (TAVR).

3. Discussion

With the successes of RWD/RWE in various phases of pharmaceutical development, there is increasing demand and tremendous enthusiasm to revolutionize drug development. However, opportunities of using RWE come with a multitude of challenges. If not properly addressed, these challenges may compromise the validity of conclusions drawn from the RWD and the confidence in the RWE.

First, it is challenging to frame the clinical questions and identify relevant RWD. A team of statisticians, clinicians, epidemiologists and data scientists should work closely to clearly formulate the research questions and utilize appropriate statistical methods for the RWD. Second, difficulty getting access to patient level data due to platform or privacy restrictions also constrains the use of RWD. The legal and ethical requirements for data sharing vary widely from region to region. There is no clear regulatory and legal framework for integrating data from multiple sources and maintaining patient privacy and information security. Third, the lack of data quality standards and common data models also hamper the broad use of RWD.; there are still great barriers in technology and capabilities before the RWD can be fully used. Advances in informatics technology, data capturing and analytics are critical for the real-time and efficient use of RWD. Fourthly, robust statistical methods and data analytical tools are much needed for generating reliable RWE. Furthermore, the explosion of RWD calls for the increasing demand of qualified statisticians and data scientists who have the technical knowledge of statistics and programming as well as the medical background of the RWD. Last, because of the constantly changing landscape, there is not a clear regulatory pathway regarding marketing approval based on RWE. Recently, FDA (2018) published a framework for the RWE Program, which serves as the roadmap for more fully incorporating RWD/RWE in the regulatory paradigm. In spite of daunting challenges, RWE has made many successful strides in the pharmaceutical industry and will continue to be the driving force of

medical innovations along with the advances in information technology, advanced technology, effective collations and open regulatory environment.

Disclosures: The authors are employees and holding stocks of respective companies. This article reflects the views of the authors and should not be construed to represent AstraZeneca, MorphoSys and Fate Therapeutics' views or opinions.

References

Ball, R., Robb, M., Anderson, S.A., and Dal Pan, G. (2016). The FDA's Sentinel Initiative—A Comprehensive Approach to Medical Product Surveillance, Clinical Pharmacology & Therapeutics, 99(3):265-268. doi:10.1002/cpt.320.

FDA (2016) 21st Century Cures Act https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence

FDA (2018) Framework for FDA's Real-World Evidence Program *https://www.fda.gov/media/120060/download*

Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016 Apr 15;183(8):758-64. doi: 10.1093/aje/kwv254.

Kim HS, Lee S, Kim JH. Real-world Evidence versus Randomized Controlled Trial: Clinical Research Based on Electronic Medical Records. J Korean Med Sci. 2018;33(34):e213. Published 2018 Jun 26. doi:10.3346/jkms.2018.33.e213

Ho M, van der Laan M, Lee H, Chen J., Lee K, Fang Y, He W, Irony T, Jiang Q, Lin X, Meng Z, Mishra-Kalyani P, Rockhold F, Song Y, Wang H & White R (2021) The Current Landscape in Biostatistics of Real-World Data and Evidence: Causal Inference Frameworks for Study Design and Analysis, Statistics in Biopharmaceutical Research, DOI: 10.1080/19466315.2021.1883475

Medical Device Innovation Consortium (MDIC) (2022) External Evidence Methods Framework. *https://mdic.org/wp-content/uploads/2022/04/MDIC-EEM-Framework.pdf*

Naidoo P, Bouharati C, Rambiritch V, Jose N, Karamchand S, Chilton R, Leisegang R. Real-world evidence and product development: Opportunities, challenges and risk mitigation. Wien Klin Wochenschr. 2021 Aug;133(15-16):840-846. doi: 10.1007/s00508-021-01851-w.

National Cancer Institute (NCI) (2022) Surveillance, Epidemiology, and End Results (SEER) Program Populations (1969-2020) (www.seer.cancer.gov/popdata), National Cancer Institute, DCCPS, Surveillance Research Program, released February 2022.

Palanca A, van Nes F, Pardo F, Ampudia Blasco FJ, Mathieu C. Real-world Evidence of Efficacy and Safety of SGLT2 Inhibitors as Adjunctive Therapy in Adults With Type 1 Diabetes: A European Two-Center Experience. Diabetes Care. 2022 Mar 1;45(3):650-658. doi: 10.2337/dc21-1584.

Purpura CA, Garry EM, Honig N, Case A, Rassen JA. The Role of Real-World Evidence in FDA-Approved New Drug and Biologics License Applications. Clin Pharmacol Ther. 2022 Jan;111(1):135-144. doi: 10.1002/cpt.2474.

Roberts MH, Ferguson GT. Real-World Evidence: Bridging Gaps in Evidence to Guide Payer Decisions. Pharmacoecon Open. 2021 Mar;5(1):3-11.

US FDA (2002) Enhancing the Diversity of Clinical Trial Populations — Eligibility Criteria, Enrollment Practices, and Trial Designs Guidance for Industry *https://www.fda.gov/regulatory-information/search-fda-guidance-documents/enhancing-diversity-clinical-trial-populations-eligibility-criteria-enrollment-practices-and-trial*

Usman, M.S., Van Spall, H.G.C., Greene, S.J. et al. The need for increased pragmatism in cardiovascular clinical trials. Nat Rev Cardiol (2022). *https://doi.org/10.1038/s41569-022-00705* ■

STATISTICIANS IN THE RWE ERA – A VISION TOWARDS IMPROVING RWE TO INFORM HEALTHCARE DECISION MAKING

Xiang Zhang (CSL Behring) and Douglas Faries (Eli Lilly)

Introduction

What's real-world data (RWD)? According to US Food and Drug Administration (FDA; 2018), *"Real-world data are the data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources."* Each of us likely contributes to RWD on a regular basis: we go to the doctor's office for an annual exam or we pick up prescriptions from a pharmacy store. Both activities will generate medical claims records that will eventually become a portion of RWD data sources. Historically, RWD sources were either based on a local healthcare system (e.g., a disease registry managed by an academic institution) with a limited number of subjects or a larger insurance claims database that has not collected sufficient information necessary for broad based research questions. Those limitations – along with the lack of randomization in RWD that introduces potential confounding and bias - have restricted the use of RWD for scientific research, and evidence generated from RWD have been considered low quality in the evidence-based medicine paradigm.

With advanced data collection technology, linking of health-care claims data with EMRs and survey data, large disease registries with the collection of key clinical data, etc., the promise of real-world evidence as a key part of medical research is now becoming a reality. The fast-growing availability of high-quality RWD – at a fraction of the cost of RCTs - presents a promising opportunity for health-care researchers to generate timely and relevant real-world evidence (RWE) to inform physicians, health policy makers, other health-care decision makers, and to ultimately benefit patients. The use of RWE now expands across the whole drug development spectrum, from drug discovery and helping in the design and augmentation of clinical trials, to phase IV research and value-based agreements with health-care payers. An important milestone in this process was The 21st Century Cures Act and the efforts that followed in using RWE to inform regulatory decisions. The FDA released its framework for RWE and also published several draft guidance documents regarding the use of RWE in drug development (FDA 2021a, 2021b, 2021c, 2021d); EMA published a vision for use of real-world evidence (RWE) in EU medicines regulation (Arlett et al. 2022); MHRA published a guidance on the use of real-world data in clinical studies to support regulatory decisions (MHRA 2021); Health Canada announced the intention to optimize the use of RWE for regulatory decisions in order to improve the extent and rate of access to prescription drugs in Canada (Health Canada 2019).

As Charles Dickens once wrote, *"It was the best of times, it was the worst of times"* (A tale of Two Cities). With the big data revolution – the spotlight is now shining on analytics to help realize the value expected from RWD. However, the analytical challenges are not trivial even with high-quality RWD, and they demand new innovation, influence, and strong communications. We view this as a great opportunity for statisticians to help lead the way to more clear and interpretable evidence from RWD.

We are delighted to see some great steps in this direction. There are growing interests and initiatives inside statistical and other professional societies regarding RWD/RWE, which include establishment of cross-industry statistical working groups (e.g. American Statistical Association RWE Working Group) that aim to develop and educate the field regarding best practices and novel statistical methods for analyzing RWD. We have also seen a dramatic increase in the number of RWD/RWE related sessions in professional statistical conferences such as the Joint Statistical Meetings. There are more funding opportunities provided by government agencies such as FDA. As statisticians working in the RWE field, we feel it is our responsibility to relentlessly pursue statistical solutions to meet the RWE challenges – and we describe below some areas where further work is needed. Specifically, we discuss RWD quality and the importance of assessing the assumptions behind our comparative effectiveness analyses for causal inference.

Challenges and Opportunities: Quality of Real World Data

RWD is not really RWE. RWE is produced by a cross functional effort requiring a relevant research question, appropriate research design, fit for purpose real world data, and analytics aligned with the research question (estimand). For such efforts, one big challenge is the reliability of the data as RWD records are typically collected for purposes other than research and thus without the rigor and quality checks involved with prospective RCTs. In fact, the focus of the initial RWE guidance documents from the FDA is largely on ensuring one establishes that the data is fit for purpose. For instance, in health-care claims data we rely on algorithms based on diagnostic coding to generate cohorts for research – but coding is not necessarily an exact science. We observe patients with diabetes diagnosis codes but no pharmacy records of anti-diabetic prescriptions - so do those patients really have diabetes? Patients may visit their doctor because of painful, swollen joints, and the doctor was not able to make the deterministic diagnosis at this initial visit ("is it rheumatoid arthritis or osteoarthritis, or other medical conditions? "). The doctor ordered lab tests and based on those results he/she correctly diagnosed the condition as rheumatoid arthritis (RA). However, the patient's medical record of the 1st visit may have several diagnoses including rheumatoid arthritis and osteoarthritis (OsA). Therefore, if the data were analyzed, OsA could be viewed as part of this patient's medical history but in fact this patient never had OsA. Statisticians could help improve data reliability by improving research on algorithms to detect recording errors and improve cohort building in large RWD databases as well as providing greater understanding of the robustness of the outcome analyses to the potential data quality issues.

Challenges and Opportunities: Causal Inference

One critical analytical challenge is providing causal inference from (non-randomized) real world studies and hybrid designs (i.e. clinical trials with complete or partial real-world control arms). The lack of randomization leads to potential bias if factors driving treatment choices are related to the outcomes under analysis. Statistical methods under Rubin's Causal Model (RCM) such as propensity score matching/stratification and Pearl's Causal Model (PCM) such as direct acyclic graph (DAG) approach, have been widely utilized in non-randomized studies/hybrid design to infer causality between interventions of interest and comparison groups. For causal inference, such analyses rely on assumptions such as strong treatment ignorability, positivity, and correct statistical modeling. We will address each of these assumptions one at a time – with the notion that statisticians could play a key role in understanding the robustness of any RWE claims through strategic and thorough assessment of these analytic assumptions.

In both RCM and PCM, strong treatment ignorability, i.e., the treatment assignment is independent of potential outcomes conditioning on a set of measured confounders, is a key assumption and violating this assumption could cause significant bias in the estimated treatment effect. However, if any unmeasured confounder exists, the strong treatment ignorability assumption may no longer hold, as the treatment assignment is not independent of potential outcomes given all confounders. To address this issue, various methods have been proposed and applied. For instance, instrumental variable methods (Angrist et al. 1996) and regression discontinuity designs (Cook 2008) are widely applied in economic research as the interest there is to investigate the impact of a particular policy on economic outcomes; negative controls (Lipsitch et al. 2010) have been used in medical research to test the robustness of estimated effect against possible unmeasured confounding. Over the past two decades, unmeasured confounding has received more and more attention and a plethora of novel methods have been developed, such as E-value (VanderWeele and Ding 2017) and Bayesian hierarchical models (McCandless et al. 2007). Several review papers (Uddin et al. 2016; Streeter et al. 2017; Zhang et al 2018) provide good references for these methods. Unfortunately, despite the availability of various methods, the quantitative evaluation of the potential impact of unmeasured confounding in comparative observational studies or in hybrid designs are still underutilized. This is likely due to lack of familiarity of methods, complexity, and the challenges that the applicability of many methods depends on the availability of specific information on the unmeasured confounders. For instance, methods such as propensity score calibration require the identification of specific unmeasured confounders and the existence of at least a subset of patients with data on the confounders. In addition, many sensitivity analyses are conducted post-hoc, based on

observed data, but not pre-specified in the analysis plan. Pre-specification of analysis is accepted as an important statistical principle to prevent the potential bias due to data dredging – and this should include planning of the sensitivity analysis. Understanding the potential threat of unmeasured confounding is best addressed in the design stage of research – as it is possible that the impact of unmeasured confounding could be larger than the expected effect size of the intervention under study (Girman et al. 2014).

We believe that a structured, pre-specified sensitivity analysis plan for unmeasured confounding could lead to improved understanding of the quality and strength of the observed RWE and provide greater confidence in the use of RWE. Zhang et.al. (2020) proposed a structured approach to assess the impact caused by unmeasured confounding, including the recommendation of beginning with sensitivity analysis that are broadly applicable such as the E-value or rule-out method. These methods do not require external information on unmeasured confounding or even identification of what unmeasured confounders may exist – and thus can be a starting point across all causal inference research.

Rosenbaum and Rubin (1983) established the foundation for analytics for comparative real-world evidence with the development of the propensity score. The propensity score for a given patient is simply the probability that the patient would receive Treatment A, given their set of baseline values (variables available at the time of the decision to use Treatment A or Treatment B). They demonstrated how a scalar value like the propensity score can provide balance between the treatment groups for multiple baseline covariates and thus reduce bias from confounding variables. Further work demonstrated how the propensity score could be utilized via regression, matching, stratification, and weighting. In addition to the assumption of 'no unmeasured confounding' discussed above, correct inference requires that correct models are used. In the case of propensity score-based analyses, this means two correct models: one model describing the treatment selection mechanism and the other model for the outcome measure. Over the past 20 years, researchers have proposed enhanced methods that provide greater robustness against model misspecification. This includes a variety of 'doubly robust' methods – including approaches that combine inverse weighting and regression (Lunceford and Davidian 2004) through double score matching (Yang and Zhang 2022). The attractiveness of such methods is in their robustness - they produce causal treatment effect estimates if one gets either the treatment selection model OR the outcome model correct, but does not require both to be correct.

More recently, researchers have proposed incorporating the use of machine learning (ML) techniques into comparative effectiveness analyses. For instance, model selection can be improved by the use of ML techniques for selection of covariates for the propensity or outcome model. van der Laan and colleagues (van der Laan and Rubin 2006; van der Laan et al. 2007) proposed a Super Learner approach to estimate potential outcomes for each treatment group within the targeted maximum likelihood estimation (TMLE) method. Rather than selecting a single estimation model, the Super Learner approach uses a set of potential estimation algorithms and through cross validation, arrives at a weighted average of the individual model estimates that is more robust than using a single method. This approach can be implemented in the R-package (Gruber and van der Laan 2012). Zagar et al. (2022) proposed the concept of model averaging as a tool for comparative real-world analyses. As one does not know the true data generating mechanism, they proposed incorporating many methods/models (e.g. stratification, matching, penalized regression) and used cross validation techniques to either identify the best method for the particular dataset (based on minimizing MSPE) or weigh each method according to their ability to predict outcomes. The weighted average of the treatment effect estimates across all methods entered into the process, with weights based on the MSPE from cross validation, is the model averaged treatment effect estimate. Simulations suggest incorporating ML tools in this manner produces a more robust estimate of the treatment effect. Thus, while one can never know or expect to arrive at a perfect model, recent advances in methodology applied to the comparative effectiveness space is allowing for more robust approaches to model building and selection.

A third assumption supporting causal inference is positivity: the assumption that each patient has a positive probability of receiving any treatment under consideration in the analysis. Current best practices begin by examining the overlap between the propensity score distributions from each of the two treatment groups. Imbens and Rubin (2015) note that differences in covariate distribution between treatment groups will manifest themselves in the propensity score distributions. Thus, examining the overlapping areas of the

propensity distributions, the area of 'common support', has become standard practice. Trimming approaches to protect the positivity assumption include selecting the largest interval of overlap between the two propensity score distributions. While good practice for ensuring causal inference, trimming and even matching methods can affect the population of inference for the analyses that are conducted.

A quality process at the design stage of the study includes a discussion about the estimand of interest. Per the ICH E9 guidance (2017), estimand is a critical component of the design and it includes 1) the population of interest, 2) the outcome from each patient used to measure the question of interest, 3) details of how intercurrent events will be addressed, 4) the statistical summary/approach used to compare the treatment groups. While traction has grown for the concept of the estimand in RCTs, we believe that such discussions are even more critical for real world research. Real world data often contain greater level of intercurrent events such as medication switching, concomitant medication, non-adherence, missing data, along with the need for trimming the population to preserve positivity. A discussion of the estimand of interest will guide how each of these real-world data issues is addressed in a more strategic fashion. In prospective observational research this process is further challenged by the fact that one does not have data on positivity until baseline data is gathered. Thus, the feasibility of the study data to address the estimand of interest may need to be re-examined after baseline data is gathered. However, the concept of 'outcome free' evaluation of the feasibility assessment – as proposed by Rubin (2007) is important to protect the integrity of the study. For further discussion of estimands in real world research see Lipkovich et al. (2020). A related concept, the target trial proposed by Hernan et al. (2016) for designing observational research, also guides researchers through such decisions during the design phase of the research. Similarly, following good practice frameworks, as outlined by researchers including the American Statistical Association RWE Workgroup (Fang et al. 2020), points researchers toward the use of the estimand thinking in planning real world designs and analysis.

Discussion

Analytics for RWE, such as causal inference methodology, has come a long way over the past 40 years. However, the field is demanding more. Big data revolution has brought many changes that we are witnessing. First and foremost, data science as a discipline has grown tremendously to meet the gap in the analysis world, with expertise focusing on big data issues. Next, we see the potential that RWE combined with ML/artificial intelligence techniques can help drive better outcomes through personalized medicine. The third is that we need to increase the quality and confidence from RWE analysis for use of the RWE in regulatory decision making. Lastly, there is more and more use of RWE/RWD across the whole drug development process.

In this brief article we have focused on the area of comparative analyses based on RWD and establishing a strategic quantitative assessment of the core assumptions behind comparative RWE (unmeasured confounding, modeling, positivity). While it is easy to limit the use of RWE by just noting the inherent potential bias with such research, we feel statisticians are well positioned to inform decision makers when and why RWE is more or less relevant to the decision at hand. With clearer information about the robustness regarding an RWE study finding, healthcare decision makers will be better equipped to make optimal use of RWE – with more informed decisions regarding approvals, formulary access decisions, and decisions by physicians and patients on individual care. Our vision is to see the statistical field championing in the RWE analytics arena and providing the needed methodology and leadership driving optimal use of RWE.

References:

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association,* 91(434), 444-455.

Arlett, P., Kjær, J., Broich, K., & Cooke, E. (2022). Real-World Evidence in EU Medicines Regulation: Enabling Use and Establishing Value. *Clinical Pharmacology & Therapeutics,* 111(1), 21-23.

Considerations for the Use of Real-World Data and RealWorld Evidence to Support Regulatory Decision-Making for Drug and Biological Products (2021). https://www.fda.gov/media/154714/download. Accessed June 5th, 2022.

Cook, T. D. (2008). "Waiting for life to arrive": a history of the regression-discontinuity design in psychology, statistics and economics. *Journal of Econometrics,* 142(2), 636-654.

Data Standards for Drug and Biological Product Submissions Containing Real-World Data Guidance for Industry (2021). https://www.fda.gov/media/153341/download. Accessed June 5th, 2022.

European Medicines Agency: ICH E9 (R1) addendum on estimands and Sensitivity Analysis in Clinical Trials to the guideline on statistical principles for clinical trials EMA/CHMP/ICH/436221/2017; published 30 Aug 2017.

Fang, Y., Wang, H., & He, W. (2020). A statistical roadmap for journey from real-world data to real-world evidence. *Therapeutic Innovation & Regulatory Science,* 54(4), 749-757.

Framework for FDA's Real-World Evidence Program (2018): https://www.fda.gov/downloads/ScienceResearch/SpecialTopics/RealWorldEvidence/UCM627769.pdf. Accessed June 5th, 2022.

Girman, C. J., Faries, D., Ryan, P., Rotelli, M., Belger, M., Binkowitz, B., ... & Drug Information Association CER Scientific Working Group. (2014). Pre-study feasibility and identifying sensitivity analyses for protocol pre-specification in comparative effectiveness research. *Journal of Comparative Effectiveness Research,* 3(3), 259-270.

Gruber, S., & van der Laan, M. (2012). tmle: An R package for targeted maximum likelihood estimation. *Journal of Statistical Software,* 51, 1-35.

Health Canada: Optimizing the Use of Real World Evidence to Inform Regulatory Decision-Making (2019). https://www.canada.ca/en/health-canada/services/drugs-health-products/drug-products/announcements/optimizing-real-world-evidence-regulatory-decisions.html. Accessed June 5th, 2022.

Li, H., Wang, C., Chen, W. C., Lu, N., Song, C., Tiwari, R., ... & Yue, L. Q. (2022). Estimands in observational studies: Some considerations beyond ICH E9 (R1). *Pharmaceutical Statistics*.

Hernán, M. A., Sauer, B. C., Hernández-Díaz, S., Platt, R., & Shrier, I. (2016). Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *Journal of clinical epidemiology,* 79, 70-75.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Lipkovich, I., Ratitch, B., & Mallinckrodt, C. H. (2020). Causal inference and estimands in clinical trials. *Statistics in Biopharmaceutical Research,* 12(1), 54-67.

Lipsitch, M., Tchetgen, E. T., & Cohen, T. (2010). Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology (Cambridge, Mass.)*, 21(3), 383.

Lunceford, J. K., & Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in medicine,* 23(19), 2937-2960.

McCandless, L. C., Gustafson, P., & Levy, A. (2007). Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Statistics in medicine,* 26(11), 2331-2347.

MHRA guidance on the use of real-world data in clinical studies to support regulatory decisions (2021). *https://www.gov.uk/government/publications/mhra-guidance-on-the-use-of-real-world-data-in-clinical-studies-to-support-regulatory-decisions/mhra-guidance-on-the-use-of-real-world-data-in-clinical-studies-to-support-regulatory-decisions.* Accessed June 5th, 2022.

Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory DecisionMaking for Drug and Biological Products (2021). *https://www.fda.gov/media/152503/download*. Accessed June 5th, 2022.

Real-World Data: Assessing Registries to Support Regulatory Decision-Making for Drug and Biological Products Guidance for Industry (2021). *https://www.fda.gov/media/154449/download*. Accessed June 5th, 2022.

Rosenbaum, P.R. and Rubin, D.B. (1983) The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrica, 70, 41-55.

Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. Statistics in medicine, 26(1), 20-36.

Streeter, A. J., Lin, N. X., Crathorne, L., Haasova, M., Hyde, C., Melzer, D., & Henley, W. E. (2017). Adjusting for unmeasured confounding in nonrandomized longitudinal studies: a methodological review. *Journal of clinical epidemiology*, 87, 23-34.

Uddin, M., Groenwold, R. H., Ali, M. S., de Boer, A., Roes, K. C., Chowdhury, M. A., & Klungel, O. H. (2016). Methods to control for unmeasured confounding in pharmacoepidemiology: an overview. *International journal of clinical pharmacy,* 38(3), 714-723.

van der Laan M.J., Polley E.C., Hubbard. (2007). *Statistical Applications in Genetics and Molecular Biology*. V 6:25.

Van Der Laan, M. J., & Rubin, D. (2006). Targeted maximum likelihood learning. *The international journal of biostatistics,* 2(1).

VanderWeele, T. J., & Ding, P. (2017). Sensitivity analysis in observational research: introducing the E-value. *Annals of internal medicine,* 167(4), 268-274.

Yang, S., & Zhang, Y. (2022). Multiply robust matching estimators of average and quantile treatment effects. *Scandinavian Journal of Statistics*.

Zagar, A., Kadziola, Z., Lipkovich, I., Madigan, D., & Faries, D. (2022). Evaluating bias control strategies in observational studies using frequentist model averaging. *Journal of Biopharmaceutical Statistics,* 32(2), 247-276.

Zhang, X., Faries, D. E., Li, H., Stamey, J. D., & Imbens, G. W. (2018). Addressing unmeasured confounding in comparative observational research. *Pharmacoepidemiology and drug safety,* 27(4), 373-382.

Zhang, X., Stamey, J. D., & Mathur, M. B. (2020). Assessing the impact of unmeasured confounders for credible and reliable real-world evidence. *Pharmacoepidemiology and Drug Safety,* 29(10), 1219-1227. ■

RWD-INTEGRATED RANDOMIZED CLINICAL TRIAL ANALYSIS

Shu Yang (NCSU), Xiaofei Wang (Duke)

Introduction

Approval of pharmaceutical products has almost always relied on positive evidence from well-designed and well-conducted phase III randomized trials. More recently, massive real-world data from routine health care delivery are becoming increasingly available. They include electronic medical records, claim and billing databases, population-based disease and product registries, data collected from wearable devices and smartphone applications, and many others. The 21st Century Cures Act (2015) encourages using real-world data and real-world evidence in drug evaluation and approval to address the questions of the therapeutic and safety of new treatments faster and less costly. The enaction of the Cures Act, however, flared heated debates. Proponents of the Cures Act argued that "Patients cannot wait," and using real-world data would speed up the process of drug development. On the other hand, real-world data were not collected for research purposes and thus may be subject to various biases due to confounding (Yang and Zhang, 2022), omitted variables (Yang and Ding, 2020), missing values (Yang, Wang, and Ding, 2019), irregular data patterns (Yang, 2021), and so on. As a result, opponents argued that the Cures Act would greenlight ineffective drugs entering the market, endangering patients.

With the advances in information technologies, parallel data of randomized trials and observational studies on the same treatment exist. Therefore, there is a great interest and need to integrate the data on the same treatment observed from these multiple data sources with complementary features, reconciling the intense debates regarding the Cures Act. On the one hand, randomized clinical trials (RCTs) offer the highest level of evidence of treatment safety and efficacy as randomization eliminates both measured and unmeasured confounders. However, patients enrolled in randomized trials are conveniently ascertained and represent a more restrictive patient group of the target real-world patient population to which the new treatment will be given. Therefore, the treatment effects estimated by standard methods lack external validity for the target population. On the other hand, real-world or observational studies often contain a much larger number of patients of the same disease and represent either a random sample of the target population. However, due to lack of treatment randomization, there are always concerns over measured and unmeasured confounders. Given the complementarity of RCTs and observational studies, integrated analysis approaches are called for to efficiently exploit the relative strengths of the data from both RCTs and observational studies and carefully address their potential drawbacks (Colnet et al., 2022).

In this article, we first discuss the common questions that investigators often have when planning an integrated analysis with parallel data from RCTs and observational studies. We then review several recent methods that exploit the complementary features of RCTs and observational studies. These methods allow us to answer these questions by providing robust and efficient estimates of the average treatment effect (ATE) with external validity, offering a pretesting approach for elastic poolability of RCT and observational studies for better efficiency in estimating the heterogeneity of treatment effect over a set of treatment modifiers, and learning targeted, optimal, and interpretable individualized treatment regimes.

Questions that can be answered by integrated analysis

The motivation of our research can be illustrated by a project that evaluates the effect of adjuvant chemotherapy in stage 1B resected non-small cell lung cancer (NSCLC). Stage 1B NSCLC is the T2N0M0 tumor that is greater than 3cm and has not spread to the lymph nodes and/or other parts of the body. CALGB 9633 is the only phase III randomized trial conducted to evaluate the effect of adjuvant chemotherapy in this patient

Table 1: Covariate and outcome means comparison of the CALGB 9633 trial sample and the NCDB sample.

	Gender (X_1)	Age (X_2)	Histology(X_3)	Tumor size(X_4)	Alive at 3 years (Y)
CALGB 9633	0.64	60.83	0.40	4.60	0.25
NCDB	0.55	67.87	0.39	4.94	0.33

population (Strauss et al., 2008). Approximately 344 patients were randomized to adjuvant chemotherapy versus observation with equal allocation. In the final analysis, the median follow-up was 74 months, and 155 deaths were observed. The primary endpoint overall survival was not significantly different (HR 0.83, 90%CI 0.64-1.08, p=0.125). Supplementary analysis demonstrated a significant survival difference in favor of adjuvant chemotherapy for patients with tumors ≥4 cm (HR 0.69, 90%CI 0.48-0.99, p=0.043). The trial had been criticized for being underpowered for trial design(Katz and Sadd, 2009). The small trial size also forbids a meaningful evaluation about how risk factors, e.g., tumor size and age, modify treatment effect. While the current guideline for treating stage 1B NSCLC is mainly based on its findings, there is a persistent scientific interest to know if the benefit of adjuvant chemotherapy is generalizable to the general population of stage 1B NSCLC patients. The concern over a lack of external validity arises from the observation that, much like other randomized trials, the patients recruited to CALGB 9633 were under-represented in both female and the elderly (70 years or older) and in patients with larger tumors. These are strong prognostic factors for survival, when compared to the general target patient population in North America. This is nicely illustrated by the distribution of the demographic and clinical characteristics of stage 1B patients from the National Cancer Database (NCDB, Morgenstern et al., 2016). The NCDB is a clinical oncology database sourced from hospital registry data, and it contains more than 70% of newly diagnosed cancer patients in North America. Of the 15379 patients from the NCDB reported between 2004–2016, 4324 and 11055 stage 1B NSCLC patients received adjuvant chemotherapy and observation care, respectively. Besides the sheer size difference between the two data sources, the baseline covariates of the CALGB patients differ from those of the NCDB patients, as seen in Table 1. The CALG 9633 sample has disproportionately high percentages of male and younger patients with smaller tumor sizes.

Given that the RCT sample represents a healthier sample than the NCDB sample, the estimates based on CALGB 9633 sample would be biased for the true effect of adjuvant chemotherapy on the target population of stage 1B NSCLC patients. One important clinical question is whether adjuvant chemotherapy benefits the general stage IB NSCLC population. For clinicians, a reasonable representation for the general stage IB NSCLC population can be the NCDB patients who meet the same eligibility criteria of CALGB 9633. Besides lack of generalizability, RCTs often face another limitation: the sample size is too small to provide sufficient power to assess the heterogeneity of treatment effects. An accurate estimation of conditional average treatment effect varying over one or more treatment modifiers, e.g., tumor size and age in our motivation example, will allow clinicians to individualize treatment according to the patient's characteristics. On the other hand, the observational NCDB study collection mechanism provides a large and diverse sample typically representative of the target real-world population. In summary, below are three questions that statisticians often want to answer through an integrated analysis of RCTs and observational studies.

- **Generalizability:** Whether the findings of an RCT can be generalized to a target population represented by one or more observational studies or a population-based registry.
- **Heterogeneity of treatment effect and poolability:** Can the observational studies data be used to improve estimation efficiency of treatment effect heterogeneity, which is defined as conditional average treatment effect as a function of a treatment effect modifier, e.g., tumor size?

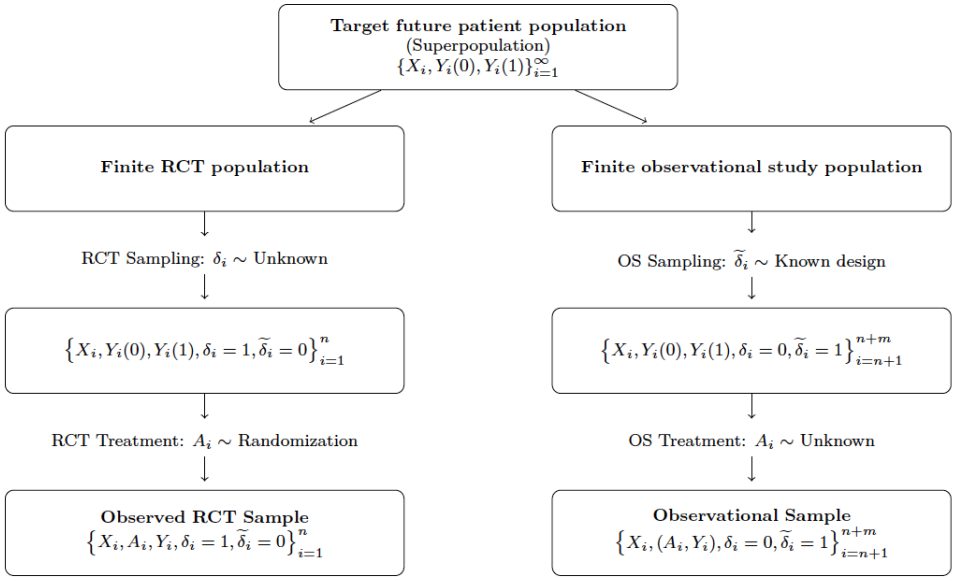


Figure 1: Illustrative data structure for integrated analysis of data from RCTs and large observational studies.

- **Heterogeneity of treatment effect and confounding function**: Whether the data from the BIASED observational studies can be used to improve the assessment of treatment heterogeneity?
- **Targeted, optimal individualized treatment regimens:** How to assign treatment based on a patient's characteristics to maximize the average benefit over a real-world patient population?

While the new approach for integrating data from large observational studies to estimate treatment effects is very appealing, unfortunately it is also confounded due to lack of treatment randomization. Methods that effectively integrate RCT and observational studies are needed but are largely under-developed. Our research group has received an National Institute of Aging R01 grant (NIA 1R01AG066883) to develop new statistical methods in this area. These methods utilize complementary features of RCTs and large population-based observational studies to leverage the advantages of both data sources to develop accurate and robust treatment effect evaluations for a target patient population routinely seen in large population-based observational studies or an underrepresented population (e.g., elderly patients or minority groups). Our framework channels an integrative analysis toolkit to empower clinical trial analysis by harnessing large real-world data and providing more generalizable (Lee et al., 2021; Lee et al., 2022, Wu and Yang, 2022a, Chu et al., 2022), efficient (Yang, Kim, and Song, 2020; Yang and Ding, 2020), and robust (Yang, Zeng, and Wang, 2020a; Yang, Zeng, and Wang, 2020b; Wu and Yang, 2022) inference of treatment effects for a target patient population. The rest of the article summarizes some progress we made in developing these new methods. We will present more details on methods for generalizability but summaries for other methods.

Generalizability

Data structure

Let X be the p-dimensional vector of covariates, A be the treatment assignment with two levels {0,1}, and Y be the outcome of interest. We adopt the potential outcomes framework to formulate the generalizability problem. Following the Stable Unit Treatment Value Assumption (SUTVA) (see, Imbens and Rubin, 2015), we assume that each subject in the target population has a potential outcome $Y(a), a \in \{0,1\}$, representing the outcome had the subject been given the treatment a. The conditional average treatment effect (CATE) is defined as $\tau(X) = E\{Y(1) - Y(0) \mid X\}$. We are interested in estimating the population ATE $\tau_0 = E\{\tau(X)\}$, where the expectation is taken with respect to the distribution of the target population. Let $\delta = 1$ denote RCT participation, and let $\tilde{\delta} = 1$ denote the observational study participation. Also, define the sampling score as $\pi_{\delta}(X) = \text{pr}(\delta = 1 \mid X)$, the design weight for the observational sample as $d = 1/P(\tilde{\delta} = 1 \mid X)$, and the conditional outcome mean function as $\mu_{a,\delta}(X) = E(Y \mid X, A = a, \delta)$ for $a, \delta \in \{0,1\}$. As seen in Figure 1, to generalize findings to the future patient population, one may consider a super-population framework that describes the distribution of all patients with a certain disease to whom the new treatment is intended to be given. The RCT is a sample from the target population with an unknown sampling mechanism, and the observational sample is a sample from the target population with a known sampling mechanism.

Methods

The problems of extending findings from RCTs to a target population has been termed as generalizability or transportability (e.g., Cole and Stuart, 2010; Pearl and Bareinboim, 2011, Dahabreh et al., 2019). Most existing methods rely on direct modeling of the sampling score π_δ (X), the sampling analog of the propensity score. The subsequent sampling score adjustments include inverse probability of sampling weighting (IPSW, Cole and Stuart, 2010, Dahabreh et al., 2019). Most sampling score adjustment approaches require the sampling score model to be correctly specified. Moreover, weighting estimators are unstable if the sampling score is too extreme. In addition, these methods often assume the observational study sample to be a simple random sample from the target population and implicitly require either the population size or all the baseline information of the population to be available.

In contrast to the approaches that focus on predicting sample selection probabilities, Lee et al. (2021) proposed to estimate the sampling score weights directly by calibrating covariates balance between the RCT sample and the design-weighted observational sample to address the selection bias of the RCT sample. A similar method has been studied by Hainmueller (2012) for causal inference with treatment selection bias. In particular, we estimate the calibration weights $\{\omega_i : \delta_i = 1\}$ by solving $\min \sum_{i=1}^{n} \omega_i \log \omega_i$ subject to the balancing constraint and $\omega_i \geq 0$ for all i, $\sum_{i=1}^{n} \omega_i = 1$. The balancing constraint is, $\sum_{i=1}^{N} \delta_i \omega_i g(X_i) = \sum_{i=1}^{N} \tilde{\delta}_i d_i g(X_i) / \sum_{i=1}^{N} \tilde{\delta}_i d_i$, where $g(X)$ is vector-valued function and often chosen to be the moment functions of X, i.e. $\{X, X^2, X^3, \cdots\}$. The calibration weighting (CW) estimator is given by

$$\hat{\tau}^{\mathrm{CW}} = \sum_{i=1}^{n} \widehat{\omega}_i \left\{ \frac{A_i Y_i}{\pi_{A_i}} - \frac{(1 - A_i) Y_i}{1 - \pi_{A_i}} \right\},$$

where π_{Ai} is the known treatment assignment propensity in the RCT. Under the standard identification assumptions in causal inference, we show that when either of the following two assumptions holds:

(a) $\tau(X) = \gamma_0^T g(X)$, or (b) $\pi_\delta(X) = P(\delta = 1|X) = \exp\{\eta_0^T g(X)\}$

for some η_0, the CW estimator is consistent and follows an asymptotic normal distribution. We further proposed an augmented CW (ACW) estimator that is doubly robust and also achieves the semiparametric efficiency bound when both nuisance models are correctly specified. It is known that the parametric approach is prone to model misspecification, especially when $\tau(X)$ and $\pi_\delta(X)$ are complex. To cope with model misspecification, we adopted the method of sieves, which allows flexible data-adaptive estimation of the nuisance functions, while the ACW estimator retains the usual root-n consistency under regularity conditions. We conducted extensive simulation studies to evaluate the finite sample performances of the proposed estimators. It is concluded that the ACW estimator is shown to be doubly robust and more efficient than the IPSW estimator and the CW estimator. When both outcome and sampling score models are misspecified, the ACW(S) estimator, a variant of ACW estimator using the sieve method, is still unbiased and efficient. The variance estimators can be calculated empirically using bootstrap. The empirical coverage rates for the unbiased ACW estimators are close to the nominal level.

Data application

We apply the proposed estimators to evaluate the effect of adjuvant chemotherapy for stage 1B NSCLC. The outcome is the indicator of remaining alive within three years after the surgery. i.e., Y=1 if alive at 3 years after surgery and Y=0 otherwise. As seen in Table 1, the CALGB 9633 sample has a significantly higher percentage of male and younger (< 70 years old) patients with smaller tumor size. It remains an important question whether adjuvant chemotherapy benefits the general NSCLC patient population represented by NCDB, with a higher percentage of female and older age and larger tumor size. Table 2 gives a summary of the proposed methods and other methods to generalize the estimated treatment effect to the target population represented by the NCDB sample. ACW-t is the ACW estimator with the nuisance functions $\mu_a(X,1)$ are estimated based on the

Table 2: Point estimate, standard error and 95% Wald confidence interval of the causal risk difference between adjuvant chemotherapy and observation based on the CALGB 9633 sample and the NCDB sample.

	Est.	S.E.	95% Wald C.I.		Est.	S.E.	95% Wald C.I.
Naive	−0.083	0.048	(−0.177, 0.011)	ACW-t	−0.104	0.060	(−0.221, 0.013)
IPSW	−0.088	0.052	(−0.190, 0.014)	ACW-t(S)	−0.138	0.286	(−0.699, 0.422)
CW	−0.106	0.065	(−0.234, 0.021)	ACW-b	−0.106	0.069	(−0.241, 0.029)
AIPSW	−0.088	0.052	(−0.191, 0.014)	ACW-b(S)	−0.153	0.057	(−0.265, −0.041)

trial sample. ACW-t(S) is the sieve variant of the ACW-t estimator using the method of sieves for sampling score and outcome models. ACT-b is the ACW estimator with the nuisance functions μ_a (X,1) is estimated based on both RCT and observational study samples and ACW-b(S) is its sieve variant. The results indicate that in the RCT sample there is an 8.3% decrease in the risk of death within 3 years for adjuvant chemotherapy over observation. The IPSW, AIPSW (an augmented variant of IPSW), ACW-t and ACW-t(S) estimators, which utilized the covariate information of the NCDB sample, show an 8.8–13.8% decrease in the risk of death within 3 years. However, the causal effect is not significant according to the 95% confidence interval. By leveraging the predictive power of the NCDB sample, the ACW-b(S) estimator give an estimate of 15.3% risk decrease, which is significant at the 0.05 level.

Elastic Poolability based on Pretesting

The heterogeneity of treatment effect lies at the heart of precision medicine. Randomized controlled trials are gold-standard for treatment effect estimation but are typically underpowered for heterogeneous effects. In contrast, large observational studies have high predictive power but are often confounded due to a lack of randomization of treatment. To tackle the second question for integrated analysis of RCTs and observational studies, Yang et al. (2020a) proposed a test-based elastic integrative analysis of the RCT and large observational studies to estimate treatment effect heterogeneity with a vector of known treatment effect modifiers. When the observational studies data are not biased, this approach combines the trial and observational studies data for efficient estimation. Due to the possible incomparability of the observational study with the RCT (e.g., unmeasured confounding, time concurrency, and measurement errors), direct integration may lead to biases. Utilizing the trial design, we constructed a test to decide whether or not to use the observational study in an integrative analysis with the RCT. Post-selection is notoriously difficult. We characterized the asymptotic distribution of the test-based estimator under local alternatives. We provided a data-adaptive procedure to select the test threshold that promises the smallest mean square error and an elastic confidence interval with an excellent finite-sample coverage property.

Confounding function modeling

To address the third question, Yang et al. (2020b) showed that the observational study, even subject to hidden confounding, may empower trials in estimating the heterogeneity of treatment effect using the notion of confounding function. The confounding function summarizes the impact of unmeasured confounders on the difference in the potential outcomes between the treated and untreated groups accounting for the observed covariates, which is unidentifiable based only on the observational study. Coupling the RCT and observational studies, we showed that the heterogeneity of treatment effect and confounding function are nonparametrically identifiable. We derived the semiparametric efficient scores and the rate-doubly robust integrative estimators of the heterogeneity of treatment effect and confounding function under parametric structural models. Furthermore, we clarified the conditions under which the integrative estimator of the treatment effect heterogeneity is strictly more efficient than the RCT estimator. Building upon these concepts, Wu and Yang (2022) further proposed an integrative R-learner that accommodates modern, flexible machine learning methods for the heterogeneity of treatment effect and confounding function.

Targeted, Optimal, and Interpretable Individualized Treatment Regimes

Personalized decision-making, aiming to derive optimal individualized treatment rules (ITRs) based on individual characteristics, has attracted increasing attention. Interpretable ITRs are desirable for clinicians or policymakers due to their intuitive appeal and transparency. The gold-standard approach to estimating the ITRs is conducting an RCT, where subjects are randomized to different treatment groups, and the bias is minimized to the extent possible. However, RCTs are limited in external validity because of their selection restrictions and therefore are not representative of the target real-world population. Conventional learning methods of optimal interpretable ITRs for a target population based only on RCTs are biased. To learn the generalizable optimal interpretable ITRs, Wu and Yang (2020) proposed an integrative transfer learning method based on weighting schemes to calibrate the covariate distribution of the experiment to that of the large observational studies. Moreover, due to privacy and confidentiality

concerns, comprehensive individual-level data is often prohibited from sharing with researchers. In contrast, summary statistics of patient characteristics of the target population are often available and can be easily shared for research purposes. Chu et al. (2022) proposed a calibrated AIPW estimator of the value function using summary statistics from the target population and then searched for the optimal ITR for the target population by maximizing the calibrated AIPW value estimator over a pre-specified class of ITRs. The resulting ITRs are targeted, optimal, and interpretable.

Conclusion

RCTs have been regarded as the gold standard for treatment effect evaluation due to randomization of treatment, which may lack external validity and are underpowered to detect treatment effect heterogeneity due to sampling bias and sample size limitations. Large observational studies contain rich information on how patients respond to treatment in real-world settings, but standard treatment effect estimates may be confounded. We have reviewed several new methods for robust and efficient estimation of average treatment effects, conditional average treatment effects, and individualized treatment rules, including calibration, test-based integrative analysis, and confounding function modeling. These methods exploit the complementing features of RCTs and observational studies. The outcome can be general, including binary, continuous, or survival outcomes. For example, Lee et al. (2022) extended the method of generalizability for time-to-event endpoint, for which the average treatment effect is defined as a function of treatment-specific survival curves, and this estimand has the difference in survival rates at a landmark time and the difference or ratio of restricted mean survival times as special cases.

Interestingly, combining probability and nonprobability samples has received much attention in the survey methodology (Yang, Kim, and Song, 2020; Yang and Kim, 2021; Yang, Kim, Hwang 2021). Probability samples are considered the gold standard approach for finite population inference, selected under known sampling designs, representing the target population. However, many practical challenges arise in collecting and analyzing probability sample data, such as data collection costs, timely issues, and increasing non-response rates. On the other hand, with advances in technology, nonprobability samples have become increasingly available for research purposes, such as remote sensing data and web-based volunteer samples. Nonprobability samples provide rich information about the target population and can be potentially helpful for finite population inference; however, they may not represent the target population due to the unknown sampling mechanisms. The complementarity of probability and nonprobability samples makes combining the information gathered from these data sources a promising avenue for finite population inference. Given the similarity, it would be helpful to exchange ideas between different fields to spark new ideas.

Lastly, we would like to acknowledge the NIA R01 grant support and our students' and collaborators' tremendous contribution to developing these new methods. We are excited about the new paradigm of evidence-based medicine and look forward to its bright future.

References

21st Century Cures Act, H.R. 34, 114th Cong. (2015). *https://www.congress.gov/bill/114th-congress/house-bill/34/*

Chu, J., Lu, W., & Yang, S. (2022). Targeted Optimal Treatment Regime Learning Using Summary Statistics. *arXiv preprint* arXiv:2201.06229.

Colnet, B., Mayer, I., Chen, G., Dieng, A., Li, R., Varoquaux, G., Vert, J.P., Josse, J. & Yang, S. (2020). Causal inference methods for combining randomized trials and observational studies: a review. *arXiv preprint* arXiv:2011.08047.

Dahabreh, I. J., Robertson, S. E., Tchetgen, E. J., Stuart, E. A., & Hernán, M. A. (2019). Generalizing causal inferences from individuals in randomized trials to all trial-eligible individuals. *Biometrics,* 75(2), 685-694.

Dahabreh, I. J., & Hernán, M. A. (2019). Extending inferences from a randomized trial to a target population. *European Journal of Epidemiology,* 34(8), 719-722.

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political analysis,* 20(1), 25-46.

Imbens, Guido W, and Donald B Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* Cambridge University Press.

Katz, A., & Saad, E. D. (2009). CALGB 9633: an underpowered trial with a methodologically questionable conclusion. *Journal of clinical oncology,* 27(13), 2300-1.

Lee, D., Yang, S., Dong, L., Wang, X., Zeng, D., & Cai, J. (2021). Improving trial generalizability using observational studies. *Biometrics*. *https://doi.org/10.1111/biom.13609*

Lee, D., Yang, S., & Wang, X. (2022). Generalizable survival analysis of randomized controlled trials with observational studies. *arXiv preprint* arXiv:2201.06595.

Morgensztern, D., Du, L., Waqar, S. N., Patel, A., Samson, P., Devarakonda, S., … & Puri, V. (2016). Adjuvant chemotherapy for patients with T2N0M0 NSCLC. *Journal of Thoracic Oncology,* 11(10), 1729-1735.

Pearl, J., & Bareinboim, E. (2011). *Transportability of causal and statistical relations: A formal approach.* In Twenty-fifth AAAI conference on artificial intelligence.

Strauss, G. M., Herndon, J. E., Maddaus, M. A., Johnstone, D. W., Johnson, E. A., Harpole, D. H., … & Green, M. R. (2008). Adjuvant paclitaxel plus carboplatin compared with observation in stage IB non–small-cell lung cancer: CALGB 9633 with the Cancer and Leukemia Group B, Radiation Therapy Oncology Group, and North Central Cancer Treatment Group Study Groups. *Journal of clinical oncology,* 26(31), 5043.

Stuart, E. A., Cole, S. R., Bradshaw, C. P., & Leaf, P. J. (2011). The use of propensity scores to assess the generalizability of results from randomized trials. J*ournal of the Royal Statistical Society:* Series A (Statistics in Society), 174(2), 369-386.

Wu, L., & Yang, S. (2021). Transfer learning of individualized treatment rules from experimental to real-world data. *arXiv preprint* arXiv:2108.08415.

Wu, L., & Yang, S. (2022). *Integrative R-learner of heterogeneous treatment effects combining experimental and observational studies.* In Conference on Causal Learning and Reasoning (pp. 904-926). PMLR.

Yang, S. (2021). Semiparametric efficient estimation of structural nested mean models with irregularly spaced observations. *Biometrics,* doi.org/10.1111/biom.13471.

Yang, S., & Ding, P. (2020). Combining multiple observational data sources to estimate causal effects, *Journal of American Statistical Association,* 115, 1540–1554.

Yang,S., & Kim, J.K. (2020). Statistical data integration in survey sampling: a review, *Japanese Journal of Statistics and Data Science*, 3, 625–650.

Yang,S., Kim, J.K., & Hwang, Y. (2021). Integration of data from probability surveys and big found data for finite population inference using mass imputation. *Survey Methodology,* 47, 29–58.

Yang, S., Kim, J. K., & Song, R. (2020). Doubly robust inference when combining probability and nonprobability samples with high dimensional data. *Journal of the Royal Statistical Society:* Series B (Statistical Methodology), 82(2), 445-465.

Yang, S., Zeng, D., & Wang, X. (2020a). Elastic integrative analysis of randomized trial and real-world data for treatment heterogeneity estimation. *arXiv preprint* arXiv:2005.10579.

Yang, S., Zeng, D., & Wang, X. (2020b). Improved inference for heterogeneous treatment effects using real-world data subject to hidden confounding. *arXiv preprint* arXiv:2007.12922.

Yang, S., Wang, J., & Ding, P. (2019). Causal inference with confounders missing not at random, *Biometrika,* 106, 875–888.

Yang, S., & Zhang, Y. (2020). Multiply robust matching estimators of average and quantile treatment effects. *Scandinavian Journal of Statistics,* doi:10.1111/sjos.12585. ■

EVALUATING THE IMPACT OF DIFFERENT RANDOMIZATION RATIOS IN DESIGNING HYBRID CONTROL TRIALS

Chenqi Fu (Penn State), Herbert Pang (Genentech), Jiawen Zhu (Genentech)

Introduction

Randomized controlled trials (RCTs) are the gold standard for evaluating the efficacy and safety of experimental treatments. While the adequately conducted RCTs have irresistible strengths, the availability of data sources and well-developed methodologies to utilize the external information bring new possibilities to successfully and thoroughly investigate the experimental intervention (Pocock, 1976; Lewis et al., 2019). Appropriate borrowing existing control data can increase the efficiencies of clinical trials by reducing RCT control arm patient enrollment and study duration. By incorporating external controls, more recruited patients are saved from an internal control arm of a typical RCT and can have a higher chance to be assigned to the treatment arm. The augmented RCT with proper study design is also able to provide high-quality evidence similar as traditional RCTs for statistical and clinical inferences (Lin et al., 2018, 2019).

Methods have been proposed and investigated to use external evidence to augment RCTs. Viele et al. reviewed and summarized several prevailing approaches to borrowing external control subjects, including pooling, test-then-pool, and dynamic

borrowing (Viele et al., 2014). Among them, Bayesian methods enjoy an exceptional advantage of dynamic borrowing which provide flexible and objective approaches to automatically down-weigh external control subjects according to the degree of heterogeneity they introduce. Therefore, Bayesian dynamic borrowing methods can reduce the confounding introduced by incorporating external control and simultaneously optimize the effective sample size which provides essential information (Dron et al., 2019). Several Bayesian dynamic borrowing methods have been developed, including the power prior introduced by Ibrahim and Chen, the commensurate prior model developed by Hobbs et al., and the meta-analytic-predictive approach proposed by Neuenschwander et al. (B. Hobbs et al., 2011; Ibrahim & Chen, 2000; Neuenschwander et al., 2016). The implementation of dynamic borrowing methods has evolved recently as well. In this paper, we focus on the evaluation of the commensurate prior model and use the *psborrow* R package to conduct simulation studies (Lu et al., 2021).

When external controls are incorporated in an RCT, serving as a supplement to the concurrent control group, the randomization ratio of treatment to concurrent control should be adjusted accordingly. Since the external controls provide additional information, fewer subjects will be needed in the concurrent control arm so that the randomization ratio can be higher than 1:1 to achieve the target power.

In this article, we conducted a simulation study to assess the sensitivity of the commensurate prior method to various sample sizes and randomization ratios for a target power. We empirically evaluated the performance of dynamic borrowing under various scenarios with different degrees of heterogeneity based on different sample sizes and randomization ratios of RCTs.

Method

Let t_i denote the time-to-event variable and X_i denote the vector of observed covariates for patient i. β denotes the coefficients of the corresponding covariates. We denote z_i as the indicator of the treatment that z_i=1 if the patient i is enrolled in the experimental arm of RCT and z_i=0 if the patient is either included in the RCT control arm or external control arm. Let k_i denote the trial indicator that $k_i = 1$ if the patient i is an external control subject and otherwise, $k_i = 0$. The hazard function for patient i is assumed to be

$$h_i(t_i|\lambda, \beta, \gamma, \delta) = h_0(t_i|\lambda) \exp(X_i\beta + z_i\gamma + k_i\delta), \quad (1)$$

where γ is the log hazard ratio reflecting the treatment effect and δ models the between-trial heterogeneity among the concurrent and external control subjects. We assume that the baseline hazard $h_0(t_i)$ follows a Weibull distribution with the shape parameter having a prior distribution of exp(0.0001) and the scale parameter following a prior distribution of lognormal(0,10000). We specify a hyperprior distribution for δ as N(0, $1/\tau$), and the precision parameter τ is assumed to follow a half-Cauchy prior with location parameter 0 and scale 0.2 and non-informative priors for β and γ as N(0,10000) (Gelman A, et al 2006).

Simulation

We conduct simulation studies to investigate the sensitivity of the proposed method for borrowing external controls with respect to different randomization ratios, sample sizes and degree of bias from the external control data source, given a fixed number of external control subjects to guild possible future clinical study designs when utilizing a known external control dataset.

Our simulation setup is motivated by a lung cancer trial with 3 variables x_1, x_2, and x_3, where x_1 is a continuous variable and both x_2 and x_3 are binary variables. The survival outcomes for RCT subjects were simulated according to

$$h_i(t_i|\lambda,\beta,\gamma,\delta) = \lambda \exp(x_{i1}\beta_1 + x_{i2}\beta_2 + z_i\gamma)\,, when\ \ k_i = 0 \qquad (2)$$

while the time-to-event for external control subjects was generated from the hazard function

$$h_i(t_i|\lambda,\beta,\delta) = \lambda \exp\left(x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3\right), when\ k_i = 1 \qquad (3)$$

β_1, β_2, λ and γ are parameters estimated from the lung cancer trial. Note that x_{i3} are equal to 0 for all the RCT subjects, but it is not the case for the external control subjects. Thus when $\beta_3 = 0$, the external control and RCT control subjects are homogenous, while when $\beta_3 \neq 0$, there is a between-trial heterogeneity.

Prior to investigating the impact of different randomization ratios in the presence of bias introduced by external controls, we select several sample sizes and randomization ratios for concurrent RCT, which achieve a target power without between-trial heterogeneity and are comparable when introducing bias. We specifically assume that there are 200 subjects available from a historical control arm. We search the combinations of the sample size of RCT arms that each arm could have 100, 200, 300,

or 400 subjects. We aim to determine RCT scenarios that have similar target power assuming the external control is homogenous to the RCT control subjects, i.e. $\beta_3 = 0$ (Fig 1). Our simulation study was based on 2000 simulated datasets/trials for each pair of sample sizes.

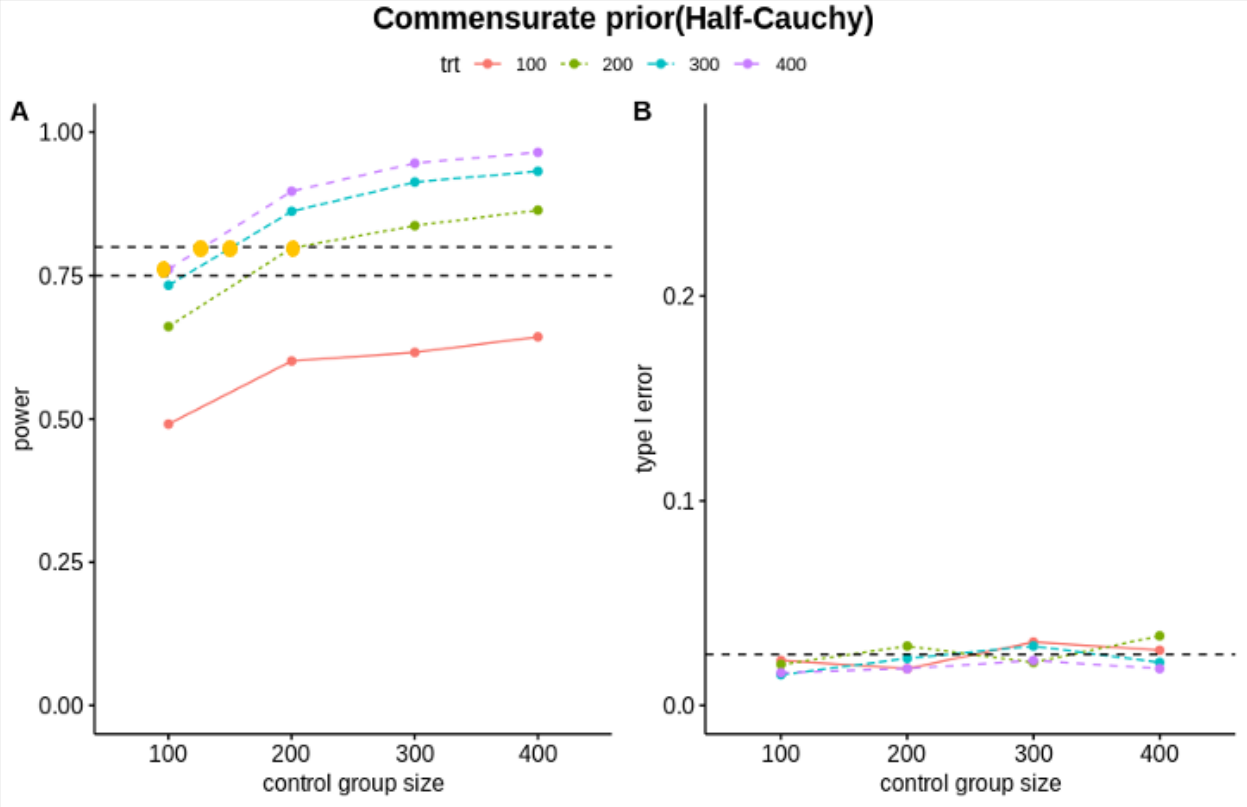


Fig.1 Sample size grid search. *(A) power and (B) type I error for different pairs of the sample size of RCT treatment and control arms using the dynamic borrowing method to incorporate 200 external control subjects. Different colors represent different sample sizes of the RCT treatment arm.*

Given 200 external control subjects, all pairs of the sample size of treatment and control arms in RCT yield overall good type I errors in the absence of between-trial heterogeneity. 4 scenarios of sample size and randomization ratio for RCT are selected

for 0.76 - 0.83 power after the sample size grid search, such as (200, 200), (300, 150), (400, 130) and (400,100) for RCT treatment and control arms, respectively (Table 1).

Table 1 Sample size grid search result.

RCT treatment	RCT control	RCT sample size	Ratio	Power
200	200	400	1:1	0.773
300	150	450	2:1	0.811
400	130	530	3:1	0.833
400	100	500	4:1	0.766

Note. 4 selected scenarios of sample size and randomization ratios.

To further evaluate the performance of the commensurate prior method with the commensurability parameter τ following a half-Cauchy prior accounting for the between-trial heterogeneity, we allow β_3 to vary from (-log(2), log(2)). When $\beta_3 \neq 0$, the hazards ratio of RCT control to external control will range from 0.5 to 2. The treatment effect estimated was adjusted for x_1 and x_2. We use 4 selected pairs of the sample size from the previous step.

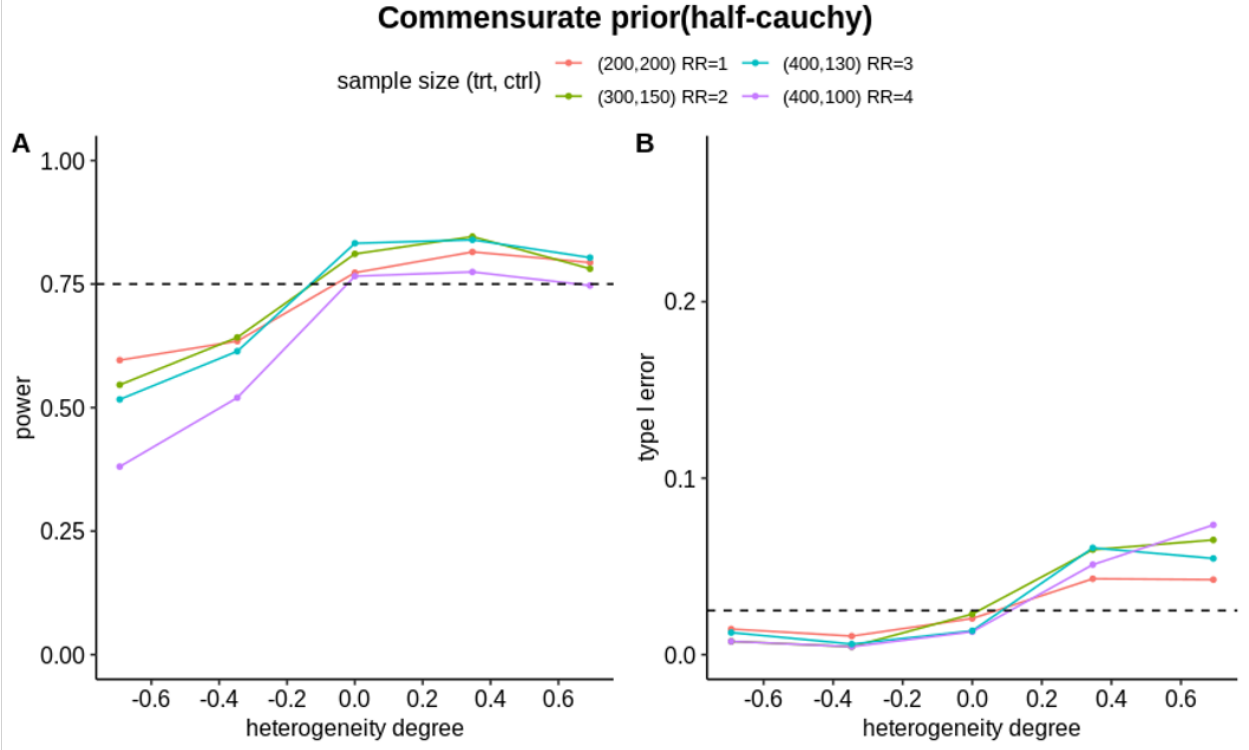


Fig.2 Comparison of randomization ratios. *(A) power and (B) type I error for different randomization ratios given 200 external control subjects when using the dynamic borrowing method. Different color represents different scenarios of randomization ratio.*

Figure 2 demonstrated the power and type I error for 4 randomization ratios (RR) with respect to various degrees of between-trial heterogeneity. In general, both power and type I error increase as the β_3 increases, in which the difference between RCT treatment and external control becomes large and exaggerates the true treatment effect. The scenario that RR=4 (in purple) has the lowest power among the four scenarios and the type I error inflated the most when there is a large between-trial heterogeneity (β_3=log(2)). The scenarios that RR=2 (in green) and RR=3 (in blue) have comparable power and similar type I error. The scenario that RR=1 has higher power than the other

scenarios when β_3 <0 and a slightly lower power than RR=2 and RR=3 when β_3>0. The type I error is better controlled when RR=1.

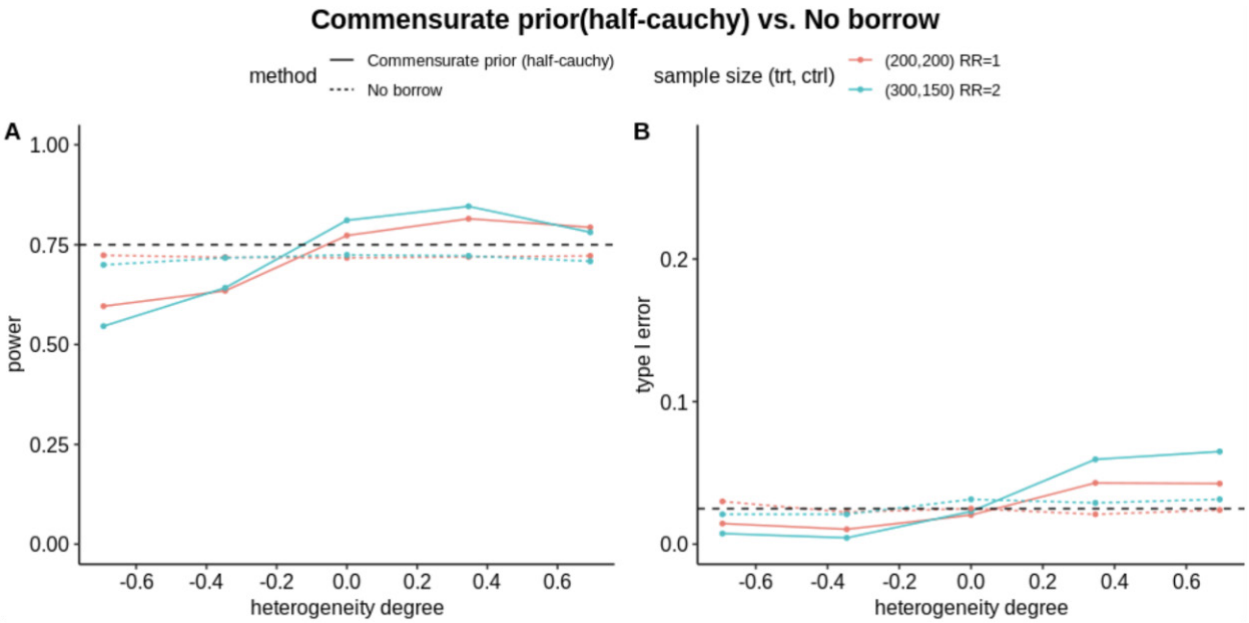


Fig.3 Comparison to not borrowing. *(A) power and (B) type I error for different randomization ratios given 200 external control subjects using dynamic borrowing method compared to not borrowing scenarios. Solid lines show dynamic borrowing results. Dashed lines show no borrowing case. Different color represents different scenarios of randomization ratio.*

We select scenarios with RR=1 and RR=2 for further investigation. The scenarios with RR=2 and RR=3 have similar performance. Considering that the total sample size of the scenarios with RR=2 is smaller than the scenario with RR=3, the scenario with RR=2 was selected. Figure 3 shows the comparison between dynamic borrowing external control subjects using the commensurate prior method and not borrowing any external control information. In general, borrowing external control gains power when $\beta_3 > 0$, in which the between-trial difference exaggerates the treatment effect. Though both

scenarios have type I error inflation, the scenario with RR=1 has just slightly higher type I errors than not borrowing.

Discussion

In this study, we evaluate the performance of the commensurate prior method in hybrid control trials using different sample sizes and randomization ratios with a given power. Incorporating a fixed number of external controls, our results suggest assigning more patients to an experimental treatment when there is no between-trial heterogeneity among the concurrent and external control subjects (the external controls subjects are unbiased). To further evaluate the sensitivity, we proceed with 4 selected pairs of the sample size for the treatment and control arms that provide a target power. In the presence of non-negligible between-trial heterogeneity, when the randomization ratio is 1:1, the commensurate prior method has a good overall performance at various heterogeneity levels and has the lowest total sample size required among all randomization ratios investigated. When there is no strong evidence for a substantial between-trial heterogeneity, randomization ratios of 2:1 and 3:1 increase the efficiency of trials without a high risk of biased estimation (Hobbs et al., 2013).

Through simulations, we found that different randomization ratios demonstrate degree of sensitivity to varied between-trial heterogeneity. In general, the hybrid trial has more consistent power and more stable type I error control when the ratio of the treatment to the control of the RCT is smaller. The scenarios with RR=2 and RR=3 have comparable

performance, which may result from different total sample sizes. It suggests that the ratio of the sample size of RCT internal control to external control is another factor that jointly impacts the operating characteristics of hybrid trials. The degree of heterogeneity plays a role in determining the effective sample size of external controls. The higher the effective sample size in the external control group, the fewer concurrent subjects needed in the RCTs (Hobbs et al., 2013).

In this study, we conducted retrospective analyses to estimate the treatment effect with a set of fixed randomization ratios of an RCT. Researchers have investigated the usage of external controls in adaptive trials to prospectively adjust the randomization ratio. Hobbs et al. proposed a design which assesses the between-trial heterogeneity at interim analyses and accordingly adjust allocation probability to achieve a balance of total information (concurrent and external) among treatment arms (Hobbs et al., 2013). Both retrospective and prospective analyses confirm that when incorporating external controls, adapting the randomization ratio enhances the efficiency of hybrid control trials. Evaluating the randomization ratio in the presence of multiple sources of external controls and historical treatment subjects is a natural extension of interest for future investigations. While further operational and practical considerations should be taken into account, our study provides valuable guidance to determine the randomization ratio of hybrid control trials in practice.

Acknowledgements

This work is partially supported by a grant from the FDA (NIH U01 FD007206).

Reference

Dron, L., Golchi, S., Hsu, G., & Thorlund, K. (2019). Minimizing control group allocation in randomized trials using dynamic borrowing of external control data–an application to second line therapy for non-small cell lung cancer. *Contemporary clinical trials communications, 16,* 100446.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian analysis,* 1(3), 515-534.

Hobbs, B. P., Carlin, B. P., Mandrekar, S. J., & Sargent, D. J. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics, 67*(3), 1047-1056.

Hobbs, B. P., Carlin, B. P., & Sargent, D. J. (2013). Adaptive adjustment of the randomization ratio using historical control data. *Clinical Trials, 10*(3), 430-440.

Ibrahim, J. G., & Chen, M. H. (2000). Power prior distributions for regression models. *Statistical Science,* 46-60.

Lewis, C. J., Sarkar, S., Zhu, J., & Carlin, B. P. (2019). Borrowing from historical control data in cancer drug development: a cautionary tale and practical guidelines. *Statistics in Biopharmaceutical Research,* 11(1), 67-78.

Lin, J., Gamalo-Siebers, M., & Tiwari, R. (2018). Propensity score matched augmented controls in randomized clinical trials: A case study. *Pharmaceutical Statistics, 17*(5), 629-647.

Lin, J., Gamalo-Siebers, M., & Tiwari, R. (2019). Propensity-score-based priors for Bayesian augmented control design. *Pharmaceutical Statistics, 18*(2), 223-238.

Lu, Y., Lin, A., Pang, H., & Zhu, J. (2021). Bayesian Dynamic Borrowing Tool for Complex Innovative Trial Designs. ASA *Biopharmaceutical Report Summer, Volume 28,*(3), 11–19.

Neuenschwander, B., Wandel, S., Roychoudhury, S., & Bailey, S. (2016). Robust exchangeability designs for early phase clinical trials with multiple strata. *Pharmaceutical Statistics, 15*(2), 123–134.

Pocock, S. J. (1976). The combination of randomized and historical controls in clinical trials. *Journal of Chronic Diseases, 29*(3), 175–188.

Viele, K., Berry, S., Neuenschwander, B., Amzal, B., Chen, F., Enas, N., Hobbs, B., Ibrahim, J. G., Kinnersley, N., Lindborg, S., Micallef, S., Roychoudhury, S., & Thompson, L. (2014). Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical Statistics, 13*(1), 41–54. ■

CMC STATISTICAL SUPPORT FOR COVID-19 VACCINE DEVELOPMENT: GOING THE EXTRA MILE!

Aili Cheng (Pfizer), Brad Evans (Pfizer), Jia Liu (Pfizer), Heliang Shi (Pfizer), Peili Wang (Pfizer), Ke Wang (Pfizer), Xiaoman Zhai (Pfizer)

The Covid-19 pandemic has caused dramatic global disruption. At the same time, it also became a driving force of miracles. One of the miracles is the very first COVID-19 vaccine, developed by Pfizer and BioNTech. It is not only the first approved COVID-19 vaccine, but also the first messenger RNA (mRNA) vaccine approved in the US. Pfizer CEO, Dr. Albert Bourla, called this achievement a "moonshot" (Bourla, 2022). It took only two hundred and forty-eight days from the time Pfizer and BioNtech announced the collaboration to the emergency use authorization (EUA) submission. About eight months later, the biological license application (BLA) was approved by the FDA (Figure 1). It took a highly collaborative and dedicated "army" to make this miracle happen. Many colleagues worked day and night on the development of this vaccine including our CMC (chemistry manufacturing and control) statistical team. The challenges were unprecedented, but the resilience and excellence the team has demonstrated were amazing. The stories shared below are just a quick snapshot of what the Pfizer CMC statistical team accomplished during this special journey; hopefully these stories will inspire others to operate at "light-speed" in bringing important new medicines to patients around the world to dare to achieve more in the future.

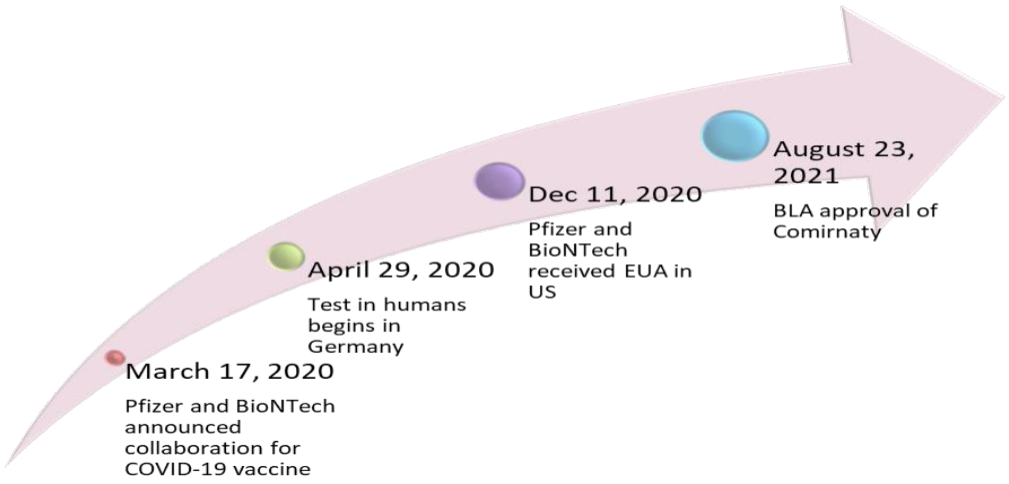


Figure 1. The key milestones of Covid vaccine development. The web links to the above milestones are listed as follows:

March 17, 2020 *https://www.globenewswire.com/news-release/2020/03/17/2001593/0/en/Pfizer-and-BioNTech-to-Co-develop-Potential-COVID-19-Vaccine.html*

April 29, 2020 *https://www.globenewswire.com/news-release/2020/04/29/2023929/0/en/BioNTech-and-Pfizer-announce-completion-of-dosing-for-first-cohort-of-Phase-1-2-trial-of-COVID-19-vaccine-candidates-in-Germany.html*

December 11, 2020 *https://www.fda.gov/news-events/press-announcements/fda-takes-key-action-fight-against-covid-19-issuing-emergency-use-authorization-first-covid-19*

August 23, 2021 *https://www.fda.gov/media/151710/download*

COVID-19 vaccine was categorized as our "light-speed" project. Such a high-speed project, and the ongoing pandemic, came with unparalleled challenges especially in manufacturing. As our CEO Dr. Albert Bourla said, ""In order to ensure that every country can have access to our COVID-19 vaccine two conditions had to be met: a price that anyone can afford and reliable manufacturing of enough vaccine for all…Meeting the second condition was much more challenging" (Bourla, 2021). For CMC statisticians, the challenge was mainly reflected in five areas:

1. mRNA was a new vaccine modality and there was limited prior manufacturing and development experience which posed significant challenges in process development, scale-up, trouble shooting, and specification setting.
2. To accelerate the development and manufacturing, two companies with multiple sites and organizations across many different time zones worked together 24 hours a day and 7 days a week. It was almost impossible to find a common slot for meetings. To better accommodate our scientists' lab schedule, the CMC statisticians had to constantly re-arrange personal schedules to attend the meetings.
3. Accelerating does not mean cutting corners. Instead, a lot of activities were running in parallel which required more work and posed significant challenge in decision making. Objective assessment using statistics became even more critical in this situation.
4. Working on site and exchanging ideas and thoughts in the hallway with scientists used to be part of CMC statisticians' routine life. However, since the start of Covid pandemic, we had to work from home and used digital devices (like Microsoft Teams chat or phone call) to get hold of our busy colleagues to get the job done. Quite often this meant weekend and late-night meetings.
5. EUA filing was new to most of us. Although our scientist partners were working very hard, getting last minute urgent requests became routine for the COVID-19 vaccine project. As a result, we constantly adjusted the delivery plans of other projects to accommodate the need of the COVID-19 vaccine project.

Despite all these challenges, our delivery time for the COVID-19 vaccine related tasks ranged from hours to just a few days. How did we do that with just a few statisticians and still maintain consistent support for other projects? As demonstrated in the examples below, collaboration, innovation, and commitment all played significant roles in our success.

Example 1: Innovative Design of Experiment (DoE) to improve the mRNA transcription

mRNA, the active ingredient of the COVID-19 vaccine, is synthesized from plasmid DNA via the In Vitro Transcription (IVT) step. A statistical design was needed to determine which factors to modify to increase the RNA integrity in this step. Three factors were considered. The typical design for this type of study is a central composite design (CCD) with 16-18 runs, which allows us to build a full response model including main effects, 2FI and quadratic terms for curvature if any. However, only up to 7 runs/block and no more than 2 blocks could be run in this case. In other words, 14 was the maximum number of experiments that our scientists could do, and it must be done in up to 2 blocks. Due to the special constraints, the typical CCD was not appropriate anymore. Instead, our statistician decided to use a hybrid design (Roquemore,1976). The hybrid design allows us to start with a two-input central composite circumscribed (CCC) design and embed a third factor to it. As a result, this design accommodated all the restrictions and still allowed for full response surface model fit.

Not only was it our first time applying the hybrid design to biological process development, it was also one of the first two studies that benefited from the Auto-reporting tool that was recently developed in RShiny to automate the reporting process (Figure 2). With literally five button clicks, a draft DOE report could be generated in minutes; this reporting step used to take days to weeks to accomplish. The RShiny tool significantly increased the report consistency and flexibility as well. It now becomes a daily tool for DoE reports.

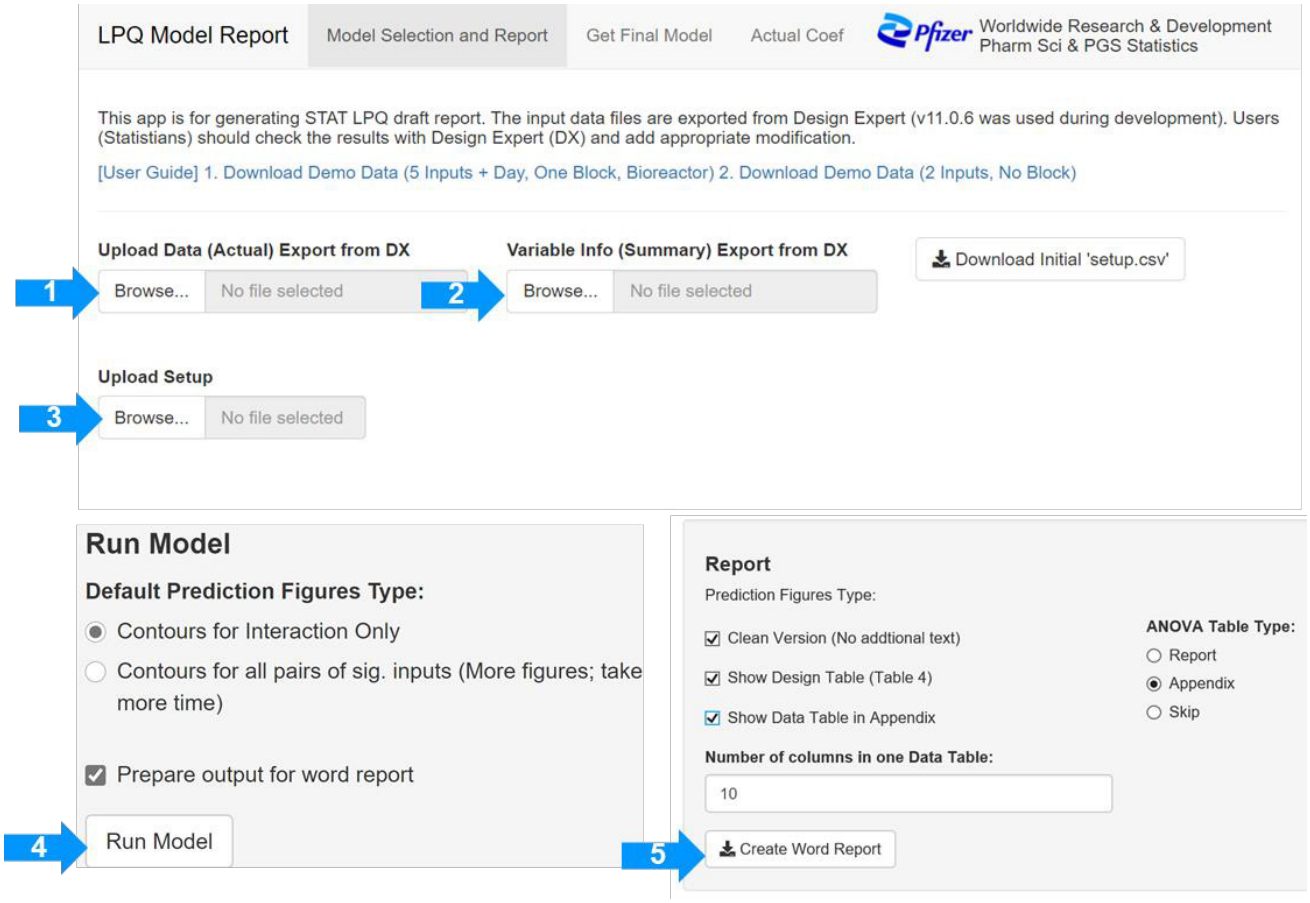


Figure 2. The interface of RShiny tool for DoE report automation

Example 2: DoE for the lipids

COVID-19 vaccine is mRNA encapsulated in lipid nanoparticles (LNP) formed by four different types of lipids. These lipids are essential for the mRNA protection and transfection. They also play an important role in decreasing the immunogenicity and improving biodistribution (Swingle, Hamilton, Mitchell, 2021). However, they are not easy to manufacture. Pfizer statisticians worked closely with scientists, designed two DoEs and built more than 40 models to understand the functional relationship between process parameters and the quality of lipids. The work was used in the delivery of a robust manufacturing process.

Example 3: In-vitro Expression (IVE) assay acceptance criteria

The quality of the vaccine is tested both during the processing and right before being released to the market. A series of tests need to be done which take multiple days

to complete. IVE is one of the critical assays which is used to confirm the presence and the in-vitro expression of the vaccine.

An acceptance criterion (i.e., the minimum expression level) needed to be determined before this assay could be used for batch release; a batch can only be released to the market if it meets this criterion, otherwise, it would be rejected. IVE is measured using a cell-based assay, which are generally more variable than biochemical and biophysical assays. The assay was validated at the time, but there was a very limited number of batches and analytical data prior to the EUA submission.

The typical strategy for situations like this is to wait until more batch data are available to better quantify the process variability and analytical variability. However, the COVID-19 vaccine could not wait. Instead, all the relevant analytical data were employed in decision making, along with the batch data. Also, various analyses were performed to understand in vivo and in vitro correlation, the impact of increasing the number of test replicates, and the feasibility to further refine assay acceptance criteria. The final decision was made based on the expected assay precision assuming the recommended replication strategy would work as intended and assuming all the planned method optimization actions would improve method precision. The resultant criterion was quite stringent and could potentially lead to high out of specification (OOS) rate if any of these expectations was not met. Fortunately, the method performance improved, as expected. We not only set a specification limit, but also reduced the method variability by optimizing the assay replication strategy.

Example 4: 2-8°C Storage Time Extension

When the COVID-19 vaccine was first released in December 2020, the recommendation was that it could only be left in the refrigerator for no more than 5 days. This was because there was not enough stability data to support a longer storage time. To add more convenience to the Pharmacy and make the vaccine more readily available, thorough analyses were performed as soon as new stability data became available. To expedite the process, the statisticians prepared the analysis method, code, display plots, and report template prior to receipt of the data. When the data finally arrived on a Sunday night, the statistical results together with the report were done right away. The analysis showed that the vaccine could be stored at 2-8°C for up to 31 days. This was then quickly approved by global regulatory authorities and almost instantly, this result got to the news: NPR reported "Pfizer Vaccine Can Stay Longer At Warmer Temperatures Before Being Discarded" (Greenhalgh, 2021). CNN said "FDA: Pfizer vaccine can now be stored at standard refrigeration temperatures for up to a month" (Sealy, 2021). Global Alliance for Vaccines and Immunizations (GAVI) announced that "The vaccine can be stored at higher temperatures in Gavi-supported countries too."(Geddes, 2021). This was probably the first time CMC statistical work so quickly made news headlines. It took only 30 days from the time we got the data to the FDA approval, which is also record-breaking.

Summary

High speed does not mean "skip steps/process". To achieve the aggressive timelines, a lot of work must be done in parallel, which often resulted in more work. Just working longer hours was not enough. Innovation, standardization, and automation (ISA) are also critical to success. Besides regular project support, ISA has been the focus of Pfizer CMC statistics team for many years. Besides the innovative DoE and RShiny tool mentioned in the examples, we have standardized and automated many other common practices. The previous time and effort clearly paid off in the COVID-19 vaccine project.

The entire development journey was like a relay race from one function to another. Every single second was used without getting a break. "Working around the clock" is probably the best phrase to describe everyone on the Covid project. Although most worked remotely, we have never worked so closely before. Also, with so many people watching and waiting, we were all fully committed to getting the work done no matter the circumstances. It is the high level of collaboration and commitment that has led to the success of the vaccine.

COVID-19 vaccine development is also a very rewarding journey. Our work progress has been reported in TV, newspaper, and websites. As Pfizer employees, we constantly received thanks and appreciation from relatives, neighbors, and friends. People even came to our working site to thank us in person (Sokolow, A., 2021). All this attention and appreciation made us feel honored and humbled at the same time. We could not be more grateful and proud to be part of this journey.

Acknowledgement

The authors would like to acknowledge David Cirelli, Rodney Combs, Ranga Godavarti, Leslie Hawley, Kim Vukovinsky, and Nicholas Warne for their helpful comments and advice (the names are in alphabetic order).

Reference

Bourla, A. (2022) Moonshot: Inside Pfizer's Nine-Month Race to Make the Impossible Possible, Harper Business

Bourla, A. (2021) An Open Letter from Pfizer Chairman and CEO to Colleagues, *https://www.pfizer.com/news/articles/why_pfizer_opposes_the_trips_intellectual_property_waiver_for_covid_19_vaccines#:~:text=In%20order%20to%20ensure%20that,of%20enough%20vaccine%20for%20all*.

Geddes, L. (2021) Change to cold-storage conditions could make Pfizer COVID-19 vaccine more widely available, *https://www.gavi.org/vaccineswork/change-cold-storage-conditions-could-make-pfizer-covid-19-vaccine-more-widely*

Greenhalgh, J. (2021) Pfizer Vaccine Can Stay Longer At Warmer Temperatures Before Being Discarded, *https://www.npr.org/sections/coronavirus-live-updates/2021/05/19/998420256/pfizer-vaccine-can-stay-longer-at-warmer-temperatures-before-being-discarded*

Roquemore, K. G. (1976) Hybrid Designs for Quadratic Response Surfaces, *Technometrics,* 18:4, 419-423, DOI: 10.1080/00401706.1976.10489473

Sealy, A., (2021) FDA: Pfizer vaccine can now be stored at standard refrigeration temperatures for up to a month, *https://www.cnn.com/world/live-news/coronavirus-pandemic-vaccine-updates-05-19-21/index.html*

Sokolow, A. (2021) Emerson Hospital thanks Pfizer scientists for their work on COVID-19 vaccine, The Sun, E-edition

Swingle, K.L., Hamilton, A.G., Mitchell, M.J. (2021) Lipid Nanoparticle-Mediated Delivery of mRNA Therapeutics and Vaccines, Trends in Molecular Medicine, Vol. 27, No. 6, *https://doi.org/10.1016/j.molmed.2021.03.003* ■

INTERVIEW WITH SANDEEP MENON (PFIZER) ON LEADERSHIP

Ling Wang, Associate Editor at ASA BioP Report

Sandeep Menon

Whether it's in academia, government, or industry, visionary leadership is a much sought-after quality in today's workplace. It is the vision that you can grow overtime, and that can differentiate oneself from others and provide you with a fulfilling career path. In June 2022, I sat down with Sandeep Menon, Chief Scientific Officer, Artificial Intelligence and Digital Sciences, and SVP, Head of Early Clinical Development at Pfizer, to have him share his leadership experience and journey with his fellow ASA Biopharmaceutical Section members. Sandeep's responsibilities include overseeing many areas in drug development, such as Clinical Science, Biostatistics, AI in Clinical Sciences, Clinical and Quantitative Systems Pharmacology, Precision Medicine and Digital Medicine. He is an elected fellow of the ASA and adjunct faculty at Boston University School of Public Health, Tufts University School of Medicine and the Indian Institute of Management. You can find his detailed bio at the end of the interview. But first things first, here is the transcript of our conversation.

Ling Wang: Hi Sandeep, it's such a great pleasure to see you virtually again, and for our chat on leadership for the ASA Biopharmaceutical report. To get us started, can you tell us a little bit about yourself, especially your journey from completing medical school and practicing medicine and then expanding your skills into Applied Mathematics and Biostatistics to growing into a senior leadership role at Pfizer?

Sandeep Menon: Thank you so much Ling. I am honored and excited to be here. I started my journey as a medical student at University of Karnataka in India (now University of Bangalore) where I received my medical degree and practiced medicine for two years. I then came to the US and obtained my Ph.D in Biostatistics from Boston University as I've always been interested in applied mathematics. As I continue to grow in my professional career in both the pharmaceutical industry and academia, I realized that I have a great passion to keep learning, and teaching is one of the best ways to learn. When I started my career here at Pfizer in late 2010, I was already an adjunct faculty member at Boston University School of Public Health. I continued to teach part time at various institutions over the years, including applied mathematics/statistics, computing courses, machine learning, and courses that blend precision medicine and quantitative sciences. In addition, I've been teaching in management schools, mostly on quantitative leadership. I believe my fascination with the evolving components of sciences, passion to teach and remain a "curious student" has helped me tremendously in my journey.

Ling Wang: Your responsibilities encompass several other areas beyond Biostatistics, such as clinical sciences, Precision Medicine, Clinical Pharmacology and Quantitative System Pharmacology, AI in clinical, digital medicine and translational imaging, and overseeing the Pfizer Innovation and Reasearch Lab (PfIRE). What learnings from your journey can you share for quantitative scientists to take to become impactful leaders?

Sandeep Menon: It is clear today that quantitative science is one of the fastest growing fields. Every industry wants to hire talent with applied mathematics and statistics backgrounds. We are very fortunate that our skills are so marketable. However, if we want to stand out as an innovator and make an impact, there are few areas that I would recommend we keep working on. I would start again with being a curious and continuous learner and making sure we are upskilling ourselves. I would say, "stay curious and even naïve at times." For example, in my early career when I studied medicine, there was not a lot of precision medicine or genomics in the curriculum, but now we cannot survive without knowing about it. We bridge the gap by continuous learning – taking courses or learning on the job. The second piece I have learned is that it is important to be "brave and authentic" and build an envoirnment where we can make mistakes. We learn from our mistakes even more than our successes. If we mean it, say it, and be authentic and humble on what we can and cannot do. The third one is developing communication skills which includes listening. I feel that I learn more from listening to others than listening to myself. Leaders who listen are more influential than leaders who want to dominate the discussion and want to be heard. Finally, it is about enjoying the journey, being patient and serving the organization and colleagues whom we work with. As I was moving up the career ladder, one of my mentors gave me an excellent advice which has stayed with me, "ask the question: how can I help my organization, my colleagues, and our patients whom we serve, instead of what can I get from the organization." When we have this mindset and attitude, success follows.

Ling Wang: Thank you so much for sharing that. During the years of your career, there may be challenges that you have experienced. Could you share a few examples of how did you overcome these challenges?

Sandeep Menon: Good question. In my journey, what I have learned in the past 10 years working at Pfizer, and about 15-16 years in the industry, especially as an applied mathematician with a medical background, is that the biggest challenge has been trying to make sure that we are able to talk in a similar language as other non-statistical, or non-mathematical colleagues. That has been a challenge, at the same time, an opportunity for being able to learn their domain, communicate and educate them in quantitative sciences. I've often felt this has been a challenge for many of the applied mathematicians and statisticians in general. Another challenge, or I would call it an opportunity, is the importance of keeping up with the science. In my current role at Pfizer, one of my responsibilities is to oversee Precision medicine and Digital medicine. We have a lab called the PfIRe lab (Pfizer innovation research lab), which leverages state of the art technology and enables dynamic and remote monitoring of human behaviors to develop meaningful and quantitative digital endpoints. The only way to succeed is to educate ourselves and educate the people around us about Precision medicine and Digital medicine. The third challenge is that, oftentimes when we bring in something new, there's always going to be some resistance for adoption. For example, when Adaptive designs and Bayesian designs in clinical trials came in the mid to late 2000s, like 2008-2009, there was a lot of resistance from colleagues to not have an interim analysis, or just stop a futile clinical research project. We learned that it takes time to get new methods off the ground, adopted and embraced. Innovation comes eventually and we need to be patient and persistent with it.

Ling Wang: Absolutely. I think your story on the Adaptive and Bayesian designs will be very helpful for quantitative scientists. You touched upon the topic of communication. How can we become better communicators, in terms of working with Physicians, Biologists, Chemists, Scientists, and Engineers, in addition to just improving the communication skills?

Sandeep Menon: My disclaimer here is that I am also still learning, I am far from being a great communicator. ASA Biopharm has been a great platform, and I personally have benefited greatly from the work ASA Biopharm has done for the quantitative disciplines to

grow a culture of communication. Even though we call ourselves American Statistical Association, technically it is much broader than that. It is a lot of data scientists, data engineers, machine learning experts and others coming under this big umbrella. A few pieces of advice I'll give on communication and goals is to be focused, which means that instead of chasing 15 consolation prizes, focus on 3-4 gold medals that have impact. For technical experts, you need to go very deep into what you know. Go 100 feet deeper in your field, know your stuff extremely well. At the same time, when you are collaborating, you should dive at least 5 feet deep into how does the Biology and the Science work, how does the Engineering Technology work, how does the Medical Information work. This will make you an invaluable communicator because you are able to meet your collaborators at common ground.

Ling Wang: What are the actions statisticians who are well-versed in our technical field can take to better develop their business acumen, and on the other side of the coin, do you have any suggestions for those who have experience and business sense, but having less time to keep up with all of the new methodology developments and technology?

Sandeep Menon: I'm far from an expert on this, but I must mention that the more you know, the better; information is an asset. It is great to specialize in certain areas, but to be successful you need to understand the bigger picture. Take the time to learn about your organization, about your company or your university to find out how it operates and get familiar with the basics. If you're in the industry, even getting familiar with the financial statements, customer life cycle, physician perspective and the patient journey is very valuable. If you know nothing about an area, go out for a coffee with a colleague and have them walk you through what they are trying to do. In any role, we should never stop asking questions. Every person that we interact with is someone we can learn from.

Ling Wang: What advice would you give to young statisticians/data scientists/applied math-

Sandeep Menon is the Chief Scientific Officer of AI and Digital Science and SVP, Head of Early Clinical Development at Pfizer Inc., and holds Adjunct faculty positions at Boston University School of Public Health, Tufts University School of Medicine and the Indian Institute of Management. At Pfizer, he is in the Worldwide Research, Development and Medical Leadership Team and leads a multi-functional global team which includes experts in Clinical Sciences, Biostatistics and Bioinformatics, Clinical Pharmacology, Quantitative Systems Pharmacology, Precision Medicine including labs, Digital Medicine which includes Pfizer Research and Innovation (PfIRE) lab, Translational Imaging and Early Scientific Planning and Operations. His responsibilities span into multiple therapeutic areas including Inflammation and Immunology, Oncology, Rare Disease, Anti-Infectives and Cardiovascular and Metabolism. He also leads PfIRe (Pfizer Innovation and Research) lab with a remit to leverage state of the art technology to enable dynamic and remote monitoring of human behaviors to develop meaningful novel quantitative digital endpoints. During his years at Pfizer Sandeep has held leadership positions of increasing responsibility, from Discovery through Pivotal Studies. Prior to joining Pfizer, he held late-phase leadership roles at Biogen Idec and Aptiv Solutions (now ICON). Before joining the industry, he practiced family medicine in Mumbai and was Resident Medical Officer.

Sandeep is an elected fellow of the American Statistical Association (ASA), awarded the Young Scientist Award by the International Indian Statistical Association, received the Statistical Excellence Award in Pharmaceutical Industry by Royal Statistical Society, UK and recently received the Distinguished Alumni Award from Boston University School of Public Health. He received his medical degree from Bangalore (Karnataka) University, India, and later completed his Masters in Epidemiology and Biostatistics and Ph.D. in Biostatistics at Boston University and research assistantship at Harvard Clinical Research Institute. He is on the advisory board for the M.S. program at Boston University. Sandeep served as an associate editor of the ASA journal Statistics in Biopharmaceutical Research and served as an invited committee member of the prestigious Samuel S. Wilks Memorial Award offered by ASA. He has published more than 50 scientific original publications and book chapters and co-authored /co-edited 7 books. He has received several awards for academic, teaching and research excellence.

ematicians who just started their career journey? What about those leaders in statistics and data science who are looking to further develop their career beyond statistics?

Sandeep Menon: First, I would also like to congratulate them for beginning such a great and fulfilling career. When you are starting your career, always be curious and keep learning. If you are just starting out in the industry, I would encourage you to join organizations such as ASA Biopharm to find not only mentors, but peers starting out in the industry. It is also vital in the first few years of your career to deepen your interests as they could lead to new skills. For more seasoned quantitative scientists, I'd encourage continuing your education because it's very important to keep learning. Even for me at this point of my career, I take some courses, I talk to a lot of junior colleagues who are just out of school and who tutor me on new methodologies and new science that I can learn every day. That is a commitment we make for ourselves and we owe it to our organizations.

For leaders aiming to grow their career beyond statistics, for example, I can give you an industry example. If you are in drug development, tell yourself, 'I'm a drug developer first, Statistician/Mathematician next.' If you come to every table as a statistician, and not a drug developer, you may be successful statistician, but you will not be able to influence your colleagues beyond your line. You need to understand drug development, understand the biologist, basics of chemistry, translational pharmacology, medical practice and competitive landscape. We need to stay rigorous about statistical methodology and authentic about statistics and mathematics, but at the same time, we need to start wearing the bigger picture hat and have a holistic approach.

Ling Wang: Lastly a fun question - I remember in one of the meetings you mentioned the movie "3 Idiots". Can you tell us a little bit why is this your favorite movie and what you have learned from it that is relevant to your career?

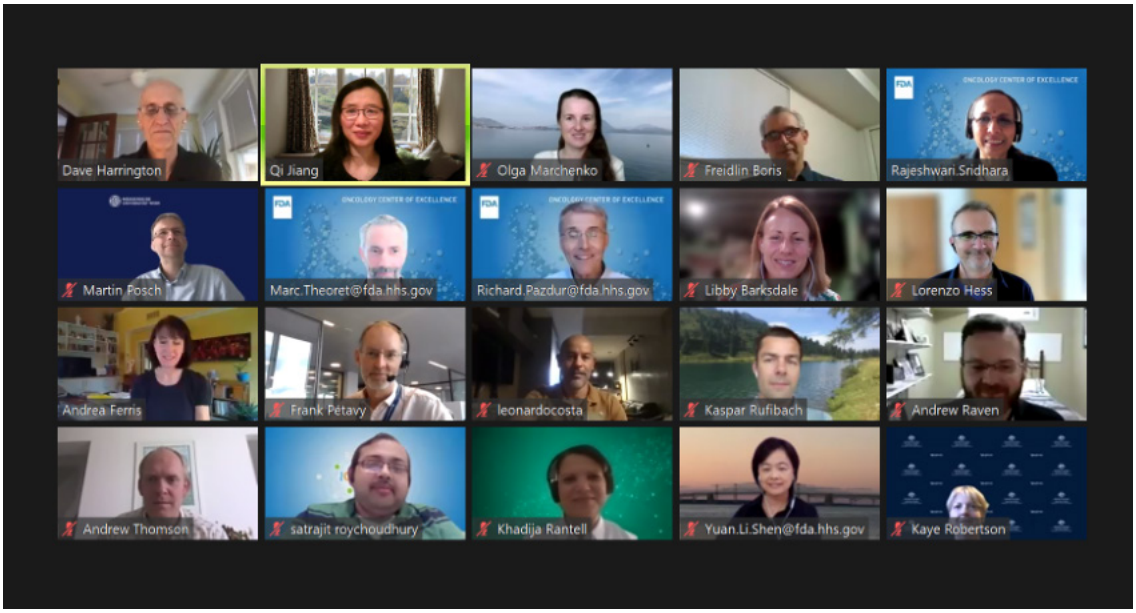
"For leaders aiming to grow their career beyond statistics, for example, I can give you an industry example. If you are in drug development, tell yourself, 'I'm a drug developer first, Statistician/Mathematician next.'"

Sandeep Menon: Great memory Ling! It has been several years. For me, that was my best Bollywood movie. Sorry I will be bragging a bit here. That is because I related myself with one of the lead roles. Growing up, for my entire academic career including medical school, I never used to worry about grades. It was always about what can I learn, and I studied with a lot of passion. Even today a few of my classmates tell me that they think of me when watching that movie. By the time I was in the second year in medical school, I remember helping first year students and even my classmates, on physiology, pathology, and some general medicine. When we were in the final year, some of our professors actually delegated me to assist teaching few courses regularly. In late 1990's / early 2000's there was little access to internet, I would just read myself or go and talk to other physicians or my seniors and learn. Eventually, it paid off with my grades as well as I aced every year which was not easy in India especially given the competition and the volume of students every year taking the board exams. The experience and the lesson for me from that movie is 'Follow your passion with excellence and success will follow.'

Ling Wang: This is really great, thank you so much Sandeep for your time, advice and comments. This was great fun. ■

SUMMARY OF ASA BIOP SECTION'S VIRTUAL DISCUSSION WITH REGULATORS ON TIME-TO-EVENT ENDPOINTS IN CANCER TRIALS IN THE PRESENCE OF NON-PROPORTIONAL HAZARDS

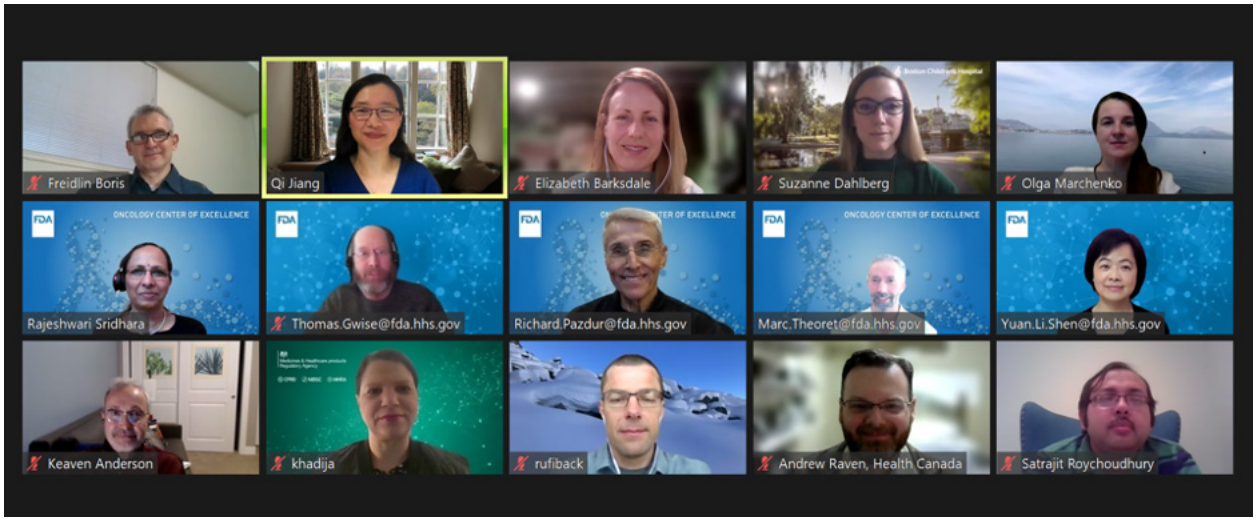
Rajeshwari Sridhara (FDA), Olga Marchenko (Bayer), Qi Jiang (Seagen), Elizabeth Barksdale (LUNGevity), Richard Pazdur (FDA), Marc Theoret (FDA)



On September 16, 2021, and March 17, 2022, the American Statistical Association (ASA) Biopharmaceutical Section (BIOP) and LUNGevity Foundation hosted open discussions with biostatisticians, clinicians, and regulators regarding impacts of non-proportional hazards as part of a series of discussions conducted for the US FDA Oncology Center of Excellence's initiative, Project SignifiCanT (Statistics in Cancer Trials). The goal of Project SignifiCanT is to advance cancer drug development through collaboration and engagement among stakeholders in the design and analysis of cancer clinical trials. Organized jointly by the ASA BIOP Statistical Methods in Oncology Scientific Working Group, LUNGevity Foundation, and the FDA Oncology Center of Excellence, the overarching theme for these two meetings was how best to evaluate and summarize results when non-proportional hazards are observed in randomized cancer clinical trials with time-to-event endpoints.

The speakers/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, MHRA, SMC, TGA, Brazil), academicians, patient advocates and expert statistical consultants. In addition, over 100 members attended the virtual meetings including representatives from other International Regulatory Agencies (e.g., from Japan, Israel, Singapore). The discussions were moderated by the BIOP Statistical Methods in Oncology Scientific Working Group co-chairs, Dr. Qi Jiang from Seagen and Dr. Olga Marchenko from Bayer, Dr Elizabeth Barksdale from LUNGevity Foundation, and Dr. Rajeshwari Sridhara, contractor from Oncology Center of Excellence, FDA.

In many randomized cancer clinical trials where a time-to-event outcome (e.g.: overall survival) is the primary outcome of interest, non-proportional hazards (NPH) are



observed. The limitations of the use of Cox-proportional Hazards Regression Model in estimating treatment effect in such scenarios are widely understood. Comparisons between different rank-based tests and combination tests have been reported in literature. While the most used log-rank (LR) test may suffer from loss of power when NPH is present, in general, the LR test is relatively robust and can detect significant differences between survival curves. However, summarizing and reporting the treatment effect remains a challenge as the hazard ratio is difficult to interpret when NPH exists.

The September 2021 forum focused on 'Summarizing treatment effect when comparing time-to-event outcomes in cancer trials in the presence of non-proportional hazards.' While different rank-based tests and combination tests have been reported in literature, there are merits and drawbacks to each of these alternative methods. The discussion in September of 2021 among multi-disciplinary experts explored optimal summary measures that can describe the average treatment effect under various NPH conditions. The March 2022 forum discussion among multi-disciplinary experts was a continuation of the discussion held in September of 2021 and focused on 'Statistical methods for evaluating treatment effects of time-to-event outcomes in cancer trials in the presence of non-proportional hazards.' The discussion in March of 2022 explored how best to design a cancer trial and pre-specify analysis methods when NPH is a possibility.

In September 2021, an introductory presentation noting the different types of NPH was followed by presentations by a cross-pharma working group representative and an academician. While estimated medians and hazard ratio using Cox-proportional model are not ideal, there are several options to summarize observed data albeit there are limitations with each measure such as weighted hazard ratio, hazard ratio over time, milestone survival, and restricted mean survival time. A few suggestions by the presenters included use of parametric models and use of simultaneous confidence bands for difference of survival functions.

The panel discussion that followed focused on what summary measures are suitable (and clinically interpretable) alternatives to hazard ratio and estimated medians when NPH are observed in randomized cancer clinical trials. The panelists opined the following. The concept of NPH came to forefront especially with anti-PD1 drugs. Understanding why NPH are observed is important. The clinical context, such as, available subsequent therapies, a subgroup effect, mechanism of action of the drug, trial design, and others, needs to be examined before interpreting the results. Among the observed NPH, delay in treatment effect with observed late separation of the survival curves is generally not as challenging to interpret as crossing survival curves. From a patient's perspective, hazard ratio and log-rank test or other statistical tests are often difficult to understand. Generally, patients do not read the product labels and expect the treating physicians to explain the risks and benefits of the treatment. The important thing is to provide the most useful information for the physicians and patients to make their decisions. A single summary measure may not be adequate when treatment effect changes over time and multiple pre-specified summary measures may be needed. Kaplan-Meier curves with confidence bands can summarize the totality of information on treatment effect.

The March 2022, an introduction to the scope of the discussion was followed by presentations by a representative from a FDA biostatistician and a representative of the cross-Pharma working group. The regulatory representative presented an evaluation (Shen Y et.al. 2022))

of MaxCombo test (Lin R.S. et.al. 2020) originally proposed by the cross-pharma working group, using a few examples from applications submitted to FDA. The evaluation suggests that undesirable properties of the test include rejecting the null hypothesis in favor of both the experimental treatment and control treatment, as well as difficulty in interpreting the results when the survival curves cross in the evaluation of treatment effect from randomized studies. The industry presentation highlighted that the primary analysis method must be prespecified as per the ICH E9 guideline requirement and that, of the available methods, the modified MaxCombo test is robust and agnostic to the types of NPH which can be prespecified. The industry representative proposed a group sequential strategy based on log-rank and MaxCombo test when planning interim analyses.

The panel discussion that followed these presentations addressed the challenges of pre-specification, how to choose an appropriate statistical method, and how the results are interpreted when NPH is observed in cancer clinical trials that are intended for regulatory decision making. The panelists indicated that a majority of cancer trials are designed assuming proportional hazards and use log-rank test as the primary analysis method to test the hypothesis of treatment effect. When NPH is observed in such trials, there is often post-hoc examination of the reasons for non-proportionality on a case-by-case situation. Only if a statistically significant treatment effect is established using the pre-specified primary analysis could one consider alternate summary measures to estimated median and hazard ratio. For certain drug classes, the possibility of NPH may be known before conducting Phase III trials and such information should be utilized in their design. Adequate follow-up and number of events are critical in establishing treatment effect when NPH is observed. At the design stage, weighted log-rank test and MaxCombo test are potential alternative options to log-rank test; however, both tests have limitations. This in-depth discussion highlighted the uncertainties at the design stage and the difficulty in pre-specifying a robust primary method of analysis to evaluate the treatment effect.

This forum, similar to previous ones, provided an opportunity to have open scientific discussions among diverse stakeholders focused on emerging statistical issues in cancer drug development. We plan to continue with similar multi-disciplinary open forum discussions in the future on a variety of important topics that include statistical aspects in cancer drug development with various stakeholders.

Acknowledgement: Authors thank Ms. Joan Todd (FDA) and Ms. Diana Aldecoa (LUNGevity Foundation) for their logistical and technical support, and Dr. Sunhee Ro (Sierra Oncology) and Dr. Rong Liu (BMS) for taking the meeting minutes.

References:

Shen Y, Wang X, Mushti S, Mulkey F, Zhou J, Gao X, Zhang L, Gwise T, Tang S, Theoret M, Pazdur R, Sridhara R. (2022). Nonproportional Hazards – An Evaluation of the MaxCombo Test in Cancer Clinical Trials, Published online in Statistics in Biopharmaceutical Research. DOI: 10.1080/19466315.2021.2008485.

Lin RS, Lin J, Roychoudhury S, Anderson, KM, Hu T, Huang B, Leon LF, Liao JJZ, Liu R, Luo X, Mukhopadhyay P, Qin R, Tatsuoka K, Wang X, Wang Y, Zhu K, Chen TT, Iacona R & Cross-Pharma Non-proportional Hazards Working Group. (2020). Alternative Analysis Methods for Time to Event Endpoints Under Nonproportional Hazards: A Comparative Analysis. Statistics in Biopharmaceutical Research 12(2): 187-198.

***Speakers/Panelists:** Dr. Keaven Anderson (Merck), Dr. Yolanda Barbachano (MHRA, UK), Dr. Michael Coory (TGA, Australia), Dr. Suzanne Dahlberg (Boston Children's Hospital), Ms. Andrea Ferris (LUNGivity Foundation), Dr. Leonardo Filho (ANVISA, Brazil), Dr. Boris Freidlin (NCI), Dr. Thomas Gwise (FDA), Prof. David Harrington (Harvard University), Dr. Lorenzo Hess (SMC, Switzerland), Dr. Richard Pazdur (FDA), Prof. Martin Posch (Center for Medical Statistics, Informatics, and Intelligent Systems at the Medical University of Vienna, Austria), Dr. Khadija Rantell (MHRA, UK), Andrew Raven (Health Canada), Dr. Kaye Robertson (TGA, Australia), Dr. Satrajit Roychoudhury (Pfizer), Dr. Kaspar Rufibach (Roche), Dr. Yuan Li Shen (FDA), Dr. Marc Theoret (FDA), Dr. Andrew Thomson (EMA). ■

CENTRAL STATISTICAL MONITORING – WHY WE NEED TO KNOW MORE

Tim Rolfe (GSK), Susan Talbot (Amgen), Rakhi Kilaru (PPD), Sharon Love (UCL)

Central monitoring is the most efficient way to ensure patient safety, trial integrity and data quality in multicentre clinical trials 1-4. and its use is recommended by both FDA & EMEA and in the ICH-E6(R2) Good Clinical Practice guidelines 5-7. However, early implementation of Central Monitoring models focussed primarily on Key Risk Indicators (KRIs) as a simple, implementable solution for identifying site to site variation 8-9. However, a more holistic approach to assessing data quality using complex statistical analytics, rather than simple univariate assessment of KRIs, can aid detection of systemic issues, data irregularities and potential fraud with the greatest potential for jeopardizing the validity of study results.

Although it might seem reasonable to assume that the incidence of misconduct or fraud in science in general and in clinical trials, in particular, is low; the true incidence is difficult to estimate 3. However, in 2021 Richard Smith, former editor of the BMJ concluded that the problem of fraud in medical research "is huge" 11.

Kirkwood et al 10 brought together the thinking on central statistical monitoring methodology (CSM) and published R-programs for others to use. They classified individual participant-level monitoring and site-level monitoring that was required and gave methods to look for recording and entry errors, procedural errors and fraud. They envisioned a time in the future when these checks would be automated and routinely carried out. Almost a decade later, trial sponsors have interpreted which checks to run and how and when to run them and many have extended the methodologies. However, limited information is available on the methods used and the results of any checks are not routinely reported.

PSI, along with EFSPI and ASA-BIOP have set up a group to create a forum for collaboration and discussion of CSM strategies and methodologies, including quality tolerance limits (QTLs). Quality tolerance limits7 are used to proactively control systematic risks to factors critical to quality. QTLs combined with statistical monitoring techniques can reduce spending on inefficient on-site monitoring practices potentially resulting in diverting resources to increase sample size or conduct more trials

The goal is to review current available methods described in the central statistical monitoring literature and recommend best practices for the broader statistical community on how robust central statistical monitoring can be achieved. Once current practices are summarised the special interest group plans to look to risk/issue detection methods from other industries where practices are potentially more advanced e.g., financial and gambling and the field of data science to evaluate new methodologies and enhancements in statistical monitoring.

On behalf of the CSM/QTL Special Interest Group (SIG), a joint collaboration including PSI, ASA BIOP & EFSPI (*https://psiweb.org/sigs-special-interest-groups/csm-qtl-sig*)

For further information, or to join the CSM/QTL SIG, please contact the co-chairs:

Susan Talbot (Amgen), *sshepher@amgen.com*
Tim Rolfe (GSK), *timothy.e.rolfe@gsk.com*

References:

1. Baigent C., Harrell F.E., Buyse M., et al. Ensuring trial validity by data quality assurance and diversification of monitoring methods. Clin Trials 2008;5:49–55.

2. Venet D, Doffagne E, Burzykowski T, et al. A statistical approach to central monitoring of data quality in clinical trials. Clin Trials 2012;9:705–7013.

3. George S.L., Buyse M. Data fraud in clinical trials. Clin Investig (Lond) 2015;5(2):161–173.

4. Agrafiotis D.K., Lobanov V.S., Farnum M.A., et al. Risk-Based Monitoring of Clinical Trials: An Integrative Approach. Clin Therapeutics 2018;40:1204-1212.

5. US Food and Drug Administration. Guidance for industry. Oversight of clinical investigations – a risk-based approach to monitoring. Available at: https://www.fda.gov/media/116754/download. Published August 2013. Accessed July 11, 2022.

6. European Medicines Agency. Reflection paper on risk-based quality management in clinical trials. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/11/WC500155491.pdf. Published November 28, 2013. Accessed July 11, 2022.

7. International Committee for Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH harmonized guideline. Integrated addendum to ICH E6(R1): Guideline for good clinical practice. E6(R2). Available at: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-6-r2-guideline-good-clinical-practice-step-5_en.pdf. Published December 1, 2016. Accessed July 11, 2021.

8. TransCelerate. Position paper: risk-based monitoring methodology. Available at: http://www.transceleratebiopharmainc.com/wp-content/uploads/2016/01/TransCelerate-RBM-Position-Paper-FINAL-30MAY2013.pdf. Published May 30, 2013. Accessed July 11, 2022.

9. Wilson B., Provencher T., Gough J., et al. Defining a Central Monitoring Capability: Sharing the Experience of TransCelerate BioPharma's Approach. Therapeutic Innovation & Regulatory Science 2014;48:529-535

10. Kirkwood A.A., Cox T., Hackshaw A. Application of methods for central monitoring in clinical trials. Clin Trials 2013;10:783-806

11. Smith R. The BMJ Opinion: Time to assume that health research is fraudulent until proven otherwise? Available at: https://blogs.bmj.com/bmj/2021/07/05/time-to-assume-that-health-research-is-fraudulent-until-proved-otherwise/ Published July 5. 2021. Accessed July 11, 2022 ■

HAPPY BIRTHDAY BIOPHARM!

In 2021 we celebrated the 40th anniversary of our Biopharmaceutical Section virtually. This year, **IN PERSON,** let's connect and celebrate the 40+1 anniversary of our Section!

At JSM, we'll celebrate during our Biopharmaceutical Section Business Meeting and Mixer. Mark your calendars for Tuesday, August 9 starting at 5:30pm in the Marriot Marquis Liberty L reception.

At RISW, we'll celebrate during the 40+1 BIOP Section Celebratory Reception on Wednesday, September 21 starting at 5:45pm.

Reminisce with current and past colleagues, former classmates and new acquaintances!

Prepare for the mixer by catching up on last year's Biopharmaceutical Report. Each issue *https://community.amstat.org/biop/biopharmreport* featured an article summarizing each decade of our Section's history.

Special thanks to the 40th Anniversary Committee!

Meg Gamalo, Jennifer Gauvin, Veronica Bubb, Lisa Lupinacci, Meijing Wu, Richard Zink

UPCOMING CONFERENCES

2022 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 three days (September 20, 2022 –September 22, 2022), with invited sessions co-chaired by statisticians from industry, academia, and the FDA and short courses on related topics offered on the first day of the workshop. To find out more visit: *https://ww2.amstat.org/meetings/biop/2022/*

Early Conference Registration Closes:
August 17, 2022
Hotel Reservations Deadline
August 29, 2022

Women in Statistics and Data Science Conference

The 2022 Women in Statistics and Data Science Conference in St. Louis, Missouri from October 6, 2022 – October 8, 2022, aims to bring together hundreds of statistical practitioners and data scientists. WSDS 2022 will highlight the achievements and career interests of women in statistics and data science.

To register visit here: *https://ww2.amstat.org/meetings/wsds/2022/conferenceinfo.cfm*

Early Registration Ends: August 25, 2022
Housing Deadline: September 3, 2022
Regular Registration: October 8, 2022

ASQ Fall Technical Conference - Mining for Quality with Statistics and Data Science

This meeting will be held between October 12, 2022 and October 14, 2022. It is organized by the ASQ and ASA Section on Physical and Engineering Sciences and Section on Quality & Productivity. The goal of this conference is to engage researchers and practitioners in a dialogue that leads to more effective use of statistics to improve quality and foster innovation. Attendees should have a desire to participate, network, and discover new techniques to solve real-world problems!

To register visit here:
https://falltechnicalconference.org/
Early registration ends:
September 12, 2022
Last day to book group rates for hotel:
September 12, 2022. ■