Volume 30, No. 2 • Summer 2023

BIOPHARMACEUTICAL REPORT

Chair: Brian Millen **Editors:** Ling Wang, Meijing Wu, Donghui Zhang

CONTENTS

Notes from the editors

Wish everyone is enjoying the summer, taking some time off and staying cool! There are so many great events this year, especially the Joint Statistical Meetings recently held in Toronto, and the upcoming FDA Industry Workshop that will be at the end of September. For this summer issue of the biopharmaceutical report, our theme is "Relentless Efforts in Statistical Innovation". We highlight topics such as external controls, dose optimization, adaptive designs, and much more.

In this issue, we start with the topic of leadership development by sharing with all of you the in-person interview we had with **Anne Heatherington**, Takeda's Senior Vice President and Head of the Data Science Institute. Anne had shared many insights she had on drug development, how statisticians and data scientist can be more effective, and AI/ML's role and future directions. Many of you will enjoy her great advice and wisdom! Next is our featured articles that highlight this issue's theme on statistical innovation. The first one, written by **Melanie Poulin-Costello** (Roche), discussed external controls in drug development in which she focused on challenges and opportunities in utilizing external controls and commented on the future directions of the field. Next, on the topic of adaptive designs, **Keaven Anderson** (Merck), **Sabrina Wan** (Merck) and **Hongtao Zhang** (Merck) shared with us great insights on practical considerations on implementing adaptive clinical trial designs. Their topics range from planning, execution to interim analysis and futility decisions, and more. The third featured article is by **Yuan Ji** (The University of Chicago) and **Dehua Bai** (The University of Chicago). They shared a coherent framework for dose finding and optimization for oncology clinical trials. On a similar topic, our fourth featured article, written by **Bo Huang** (Pfizer) and **Ying Yuan** (The University of Texas MD Anderson Cancer Center), discussed design strategies for dose optimization in oncology drug development. Continuing our discussion on the topic of clinical trial diversity from the last issue, **Julie Shah** (Boehringer Ingelheim), **Xiang Tang** (Boehringer Ingelheim) and **Nancy Bauer** (Boehringer Ingelheim) opined on diversity in clinical trials and shared a diversity dashboard where one can monitor trial diversity as it progresses. When it comes to quality, **Ajaz S. Hussain** (Independent Advisor and Consultant) shared with us advice for statisticians to be more effective in underpinning assurance of pharmaceutical quality for the patients. Our last featured article is by **Xin Huang** (Abbvie), **Haiyan Zheng** (University of Cambridge) and **Steven Novick** (Eli Lilly), in which they commented on the role of statistical innovation in enhancing the success of translational research.

Later in the issue, we have updates on progress from ASA-DahShu IDSWG Multidisciplinary Master Protocol Working Group and Section on scientific working group proposal committee. It's important to hear about the great work that these groups are doing. Then you can see summary of four virtual discussions organized by ASA BIOP section's Statistical Methods in Oncology Scientific Working Group, the FDA Oncology Center of Excellence, and LUNGevity Foundation. The topics of these four discussions are: 1) Assessing bias in cancer studies with real world data elements, 2) Statistical Considerations in the Early Interim Overall Survival Analysis in Indolent Cancers for Evaluation of Risk, 3) Non-inferiority Cancer Clinical Trial Design Considerations when Data from a Single Foreign Country is Available, 4) Considerations of Innovative Cancer Clinical Trial Designs for Post-Market Dose-Optimization Studies. As September is here, the co-chairs of the 2023 BIOP Regulatory and Industry Statistics Workshop (RISW), **Erik Bloomquist** (FDA) and **Fanni Natanegara** (Eli Lilly), are extending a warm invitation to all of you to join them in Washington DC for a few days of scientific exchanges and gathering. They have also highlighted the conference's topics and the progress they have made so far. We close this issue with a list of upcoming conferences that may be of interest to the BIOP community at large. Many thanks to the contributing authors and ASA colleagues who have made this summer issue of BIOP report 2023 possible!

INTERVIEW WITH ANNE HEATHERINGTON (TAKEDA) ON LEADERSHIP

Ling Wang and Meijing Wu, Editors, *ASA Biopharmaceutical Report*

In today's ever-changing science and technology with abundant data and information, how to embrace new mindsets and cultivate a data-driven culture will define the next generation data science leaders in the pharmaceutical and biotech industry. Those leaders who relentlessly pursue the value that data brings and empower teams to unlock new opportunities are set to drive innovation and unlock new opportunities for the future. In May 2023, the editors of the Biopharmaceutical Report were fortunate enough to sit down and have an in-depth conversation on leadership with **Anne Heatherington**, SVP R&D Chief Data & Technology Officer and Head, Data Sciences Institute (DSI) at Takeda. Anne leads the overall strategy and execution of all quantitative sciences including clinical pharmacology, biostatistics, global outcomes research, epidemiology, technology and digital sciences. You can find her detailed bio in the blue sidebar on this page. During the conversation, she shared great insights with us on her journey in the field of data science and its application in drug development. She also emphasized the importance of data and how statisticians, data scientists, and quantitative researchers contribute to the pharma/biotech industry. Furthermore, Anne shared her thoughts on the role of AI/ML and how to lead in this ever-changing field nowadays. She gave invaluable advice on what success means for a data science organization, and the qualities which the best data science leaders possess. It's an enlightening conversation and we are delighted to share this with all of you here.

Ling Wang: It is great to finally meet you. You're such an inspiration to all of us. I was wondering if you could start by telling us a little bit about yourself, and how did you find your passion in data science through your career journey?

Anne Heatherington: Thank you for asking me to do this interview. It's a real privilege to be interviewed

Anne Heatherington, Ph.D., is the R&D Chief Data & Technology Officer and Head, Data Sciences Institute (DSI) at Takeda where she leads overall strategy and execution of all quantitative sciences including clinical pharmacology, biostatistics, programming, global outcomes research, epidemiology, technology, data sciences and digital sciences.

Throughout her 20+ year career, Anne has led organizations and programs in large pharma, mid-size biotechs and start-up organizations. Prior to Takeda, Anne served as Head of Clinical Development at Summit Therapeutics. Prior to Summit, Anne served as Vice President and Head of Quantitative Clinical Sciences at Pfizer where she oversaw Clinical Pharmacology and Biostatistics supporting research and development activities across several Research Units.

Anne received her bachelor's degree in pharmacy from Queen's University Belfast, Northern Ireland, and her Ph.D. in pharmacokinetics from the University of Manchester, England. She completed post-doctoral training in the Centre for Bioengineering at University of Washington, Seattle and spent her early career at Amgen Inc, Thousand Oaks, California and Pfizer Ltd (Sandwich UK and Cambridge MA).

by the ASA. I'm a clinical pharmacologist by training. Actually, my first degree is pharmacy. My second degree is pharmacokinetics. I was lucky enough to do my Ph.D at Manchester University with Professor Malcolm Rowland who is one of the godfathers of pharmacokinetics. I wouldn't call myself a data scientist by any means, at least not in the traditional definition. But I really look at how we can maximally use data. My whole career has focused on key questions, including "How do we use our data to inform drug development optimally?" and "How do we generate high-quality data that allows us to make decisions?"

Ling: That's very interesting and inspiring at the same time, thank you. A follow-on question is from your strategic and business point of view, how do you think statisticians, data scientists and quantitative researchers can make the best contributions to the pharm and biotech industry?

Anne: My own passion is driven by how do we generate the best quality data to enable decision making, and then how do we analyze and use our data maximally. You need a full range of skills to do that well, in drug discovery, drug development, registration, evidence generation and beyond. Data is our goldmine really, and it's all about how we use it. Our statisticians use it primarily to help with trial design, thinking about our endpoints and the robust analyses we do in our trials, whereas our clinical pharmacologists think about data in terms of dose selection and more longitudinal type modeling, particularly when we're thinking about pharmacometrics. Our epidemiologists think about how we can use real-world data to inform trial design or to provide additional evidence. Our data scientists really get into the meat of large data sets and how we analyze and apply these data - from digital tools to genomics.

It's really all about making sure you have the right expertise working with the right data types with the right analyses to answer the right questions. At Takeda within DSI, we've come to define "data science" as all of those things.

Ling: To follow on that question, can you share some experience because you have a synergistic data science organization which is also rare. What are the examples you can give to make the best out of this whole group?

Anne: I can give you one example I love, which is around disease data strategies. When we go into a new disease there are many things we need to know. We need to understand the endpoints, the standard of care, what the competitors are doing, how patients interact with clinical trials. We need to understand the longitudinal nature of the disease and the types of biomarkers and whether they are predictive or prognostic. So, for each new disease we go into, we now implement a disease data strategy. Initially, we define the key questions that we need to answer about this disease and bring the clinicians, the statisticians, clinical pharmacologists, our digital experts, our pharmaco-epidemiologists, everybody into that discussion to define the key questions. We then discuss what data types we would need to answer those questions as well as the potential data sets. Those data sets could be those that we have internally, they could be those from consortia, real-world data, or any other sources. By taking this approach, we are able to ensure that the data sets satisfy all the needs for the programs and for the analysis by multiple talented people to help initiate the quantitative learning for that disease area. To me, that's a real benefit of the way we're organized. We're also starting to standardize how we bring data in, how we store it and then make sure we have the right governance to ensure appropriate data access.

Ling: Next, a very trendy question, with the emergence of AI tools, such as ChatGPT and others, where do you think the pharma/biotech industry can leverage this new technology? What's your vision on thinking about how the field will evolve in the near term, and maybe in ten years?

Anne: About three or four years ago, we started a collaboration with MIT focused on the application of AI/ML to health, and it's been an incredible collaboration.

The way we approached this collaboration was to initially define the big business problem within Takeda; we then asked if we had the right data sets to address the problem and if AI/ML were the best techniques for solving that problem. We initially jointly selected 12 such business problems, and then we collaborated with the amazing researchers at MIT to develop the algorithms. Like most collaborations, it takes time, but we've had real successes, particularly in the area of manufacturing and pharmaceutical sciences because of the volume of data generated, as well as the data standardization and intentionality. This area is really ripe for AI/ML techniques in terms of processing and automation methods. We've also looked at diagnosis of several diseases using real-world datasets with varying degrees of success. Part of the reason for the MIT collaboration was to develop our internal expertise, as well as our approach to the framework, considering rigor, repeatability, ethics, bias etc, in which we conduct such work. We have had more than 150 internal colleagues involved in the MIT program and we've developed a really nice framework for how we approach AI to make sure we're asking all the right questions.

To answer your question about where generative AI can add value, the first place is in very mundane tasks such as writing notes from a meeting or in document review, you can ask the generative AI to summarize it for you. Of course, all of the outputs require very careful human review, but overall, they're likely low risk. We should then be looking at efficiencies around automation - which could be as simple as standardizing our tables listings and figures after database lock or generation of clinical study reports. But before we would implement GenAI into our highly controlled regulatory environment, we would need to have a clear framework in place for acceptance and review. Imagine the time we'd save with just TLFs and CSRs; and then imagine investigators' brochures and summary documents! But again, this would all need to be considered within a framework in partnership with our technologists and business experts. Otherwise, we're going to get in trouble.

Ling: Now let me flip the question. What's your philosophy of saying "no" to something because AI has so much attention? A lot of people are trying to use it everywhere. I've talked about the term "responsible AI" because we wanted to make the right decision to use it in the right place. What's your philosophy on that?

Anne: My philosophy is, first of all, there has to be a clear business case. Secondly, as I said, we have to operate within a framework - any application that relates to our portfolio in R&D is assessed within the context of this framework for responsible AI. But of course, ultimately, it's all about the data. If you don't have the right data, AI isn't going to do the job for you.

Ling: Next question is, what do you think success looks like for a data science organization?

Anne: What I care about most is getting good drugs to patients faster. I tell my organization if you have an idea and you tell me it helps us get good drugs to patients faster, the answer is probably yes. Success for me is where we have used our skills to help make more informed decisions, to make the process more streamlined, to improve quality, and to create drugs that really matter to patients. If we've done all of those things with data at the core, then that's success.

Ling: Nowadays talent is the most important piece in the Data Science organization. What are the best leadership qualities that you're looking for when you identify future data science leaders?

Anne: It's a really interesting question. Of course, capability is key, right? That doesn't necessarily mean you've gone to the best schools. Many of us didn't go to the best school. It doesn't necessarily mean you've got the best skills or that you grew up in an English-speaking country. It means you have core education and training, combined with one of the best capabilities - you are a hard worker and you wish to learn.

Layered on top of that is the ability to interact and to explain what you're doing in a way that other people can

make informed decisions with the information you're bringing. Too many times, us data junkies, love our data and our methods so much and we love talking to each other, but we're hopeless at explaining it to someone else. Let me emphasize, this isn't an English-as-a-second-language problem. This is "I am so into my analysis that you should understand my analysis." We have to turn it around and be able to explain what we are doing in a way that anyone can make an informed choice. The ability to communicate clearly is very important.

Thirdly, the willingness to engage and interact and not be a purist. Oftentimes there's no right or wrong way. It's okay for some trials to have 60% power, and for some other trials, maybe you want to have 90%. Maybe P-values don't matter and it's more about effect size or maybe you use Bayesian methods or confidence intervals. Being open to those negotiations, listening, and understanding where extreme rigor is essential, and then being able to determine where you can have some more flexibility.

Then as you become more senior, the ability to listen, and the willingness to engage and continually extend the edges of your knowledge outside of your field becomes much more and more important.

Ling: In the area of data science and drug development, how do we navigate that from a career perspective when we work on the career ladder, even outside of data science? What's your perspective on that?

Anne: In DSI, I love that colleagues have the option to diverge. However, I think my career advice to everyone is to develop your absolute foundational knowledge in your own discipline, become an expert and be known as the person that delivers – on-time and high quality. There's so much opportunity in this space if you've got your foundation, are willing to learn and you're in an organization that supports career growth and movement.

When I have new starters here in my organization, I tell them "I hope you can spend your entire career here because there is a place for you in different therapeutic areas, across different quantitative groups, and throughout the various stages of development". We are so lucky to have so much opportunity for development and growth!

Ling: Last question we have is, whether you can share a fun fact about yourself with us?

Anne: One thing that I love to do is to go to concerts and music festivals. This summer I have tickets for Ed Sheeran, Pink, Arctic Monkeys, Alicia Keys and I think I might fit in The Killers (one of my favourites) when I am back in Belfast. I absolutely love doing that with my husband, kids or girlfriends!

Ling: Thank you so much for this fantastic interview, Anne. Have fun at those concerts! ■

EXTERNAL CONTROLS IN DRUG DEVELOPMENT

Melanie Poulin-Costello, *Hoffmann-La Roche*

Two disparate sources of data, randomized controlled trials (RCT) and real world data (RWD), are each useful in drug development. The RCT is our optimal tool for assessing safety and efficacy of an experimental treatment. RWD clarifies the effectiveness of treatments outside of the restrictions of an RCT. Statisticians are exploring the intersection of RCT and RWD to meet the dynamic needs of drug development, regulations, and patients. One such tool statisticians are currently testing, and in fact using, is the external control arm (Rahman et al., 2021).

The RCT as we know it today emerged from a long history of experimenters. In 1747 Dr James Lind's "Treatise on Scurvy" compared treatments in a non-randomized, 6 arm trial with N=12 patients (Collier, 2009). The clinical trial continued to evolve and not until 1948 was a controlled, randomized trial, designed by British statistician Sir Austin Bradford Hill, first published (Crofton & Mitchison, 1948). The current advances in RCT use data external to the trial to supplement a single arm or small sample size or, in some bold cases, as a fully external control arm. It is no surprise that a qualified, innovative pool of statisticians is evolving the RCT. The external control raises many opportunities in drug development but also has had many challenges (Hall et al., 2021; Lambert et al. 2022).

The FDA, in their draft guidance on externally controlled trials, provides the following definition:

"In an externally controlled trial, outcomes in participants receiving the test treatment according to a protocol are compared to outcomes in a group of people external to the trial who had not received the same treatment. The external control arm can be a group of people, treated or untreated, from an earlier time (historical control), or it can be a group of people, treated or untreated, during the same time period (concurrent control) but in another setting." (U.S. Food and Drug Administration, 2023)

Almost 50 years ago, Pocock published pioneering work on methods to combine RCT and historical data on the control treatment (Pocock, 1976). Today in 2023, as rich data sources have become more accessible, statisticians are exploring the use of external data. Both historical trial data (HTD) from RCT and RWD sources are being used as external controls. Patient anonymization techniques enable data sharing compliant with patient privacy laws (General Data Protection Regulation (GDPR) in Europe (EU), for example). In collaboration with organizations such as TransCelerate (*https://www.transceleratebiopharmainc.com*), and Vivli (*https://vivli.org*), pharmaceutical companies are providing platforms for sharing HTD from RCT. RWD sources from electronic health records (EHR) are available commercially (e.g., Flatron (*https://flatiron.com*)) as well as through various registries and public datasets. As our access to data increases, our understanding of data's usefulness in an externally controlled clinical trial setting expands.

So let's explore what statisticians have been learning.

Challenges with External Controls

Many industry proponents of externally controlled trials may claim cost and time savings to justify their use; however, the pharmaceutical industry's profit margins may not always justify such a need. More generally, avoiding randomizing patients to a placebo or substandard care arm may justify pulling external data in for a control arm rather than a traditional RCT. This is most evident in rare disease and pediatric trials which struggle to randomize sufficient patient numbers. A single arm study can often suffice for regulatory approval in those settings, especially when an external control can be brought in to address causal inference questions (Sola-Morales et al.,2023; Arondekar et al., 2021).

External controls, however, come with many caveats. Statisticians have identified several challenges, as well as potential solutions, for constructing external controls whether using historical clinical trial data, or a RWD source.

RWD and historical clinical trial data serving as an external control introduce multiple sources of bias into the causal inference question that statisticians are attempting to answer with RCT (Schmidli et al., 2019; Thorlund et al., 2020; Burger et al., 2021; *Historical Trial Data Sharing - TransCelerate*, 2021; U.S. Food and Drug Administration, 2023). In addition to biases, limitations on HTD and on RWD sources hinder their use as external controls. Some of these biases and limitations are listed in Table 1.

Several statistical methods and study design considerations have been proposed to address biases and limitations,

as shown in Table 1. However, none of the statistical methods is sufficient to replace the benefits of randomization: balancing known and unknown confounders between treatment arms to perform causal inference (Wang, 2021).

The FDA states that the "likelihood of credibly demonstrating the effectiveness of a drug of interest with an external control is low, and sponsors should choose a more suitable design" (U.S. Food and Drug Administration, 2023). This however still leaves room to innovate solutions for drug development and internal decision making in early phase studies. Statisticians can also address treatment access challenges with Health Technology Assessment (HTA) by employing external data sources. In the rare disease space, regulatory authorities are willing to engage in early discussions to assess the usefulness of external controls (Sola-Morales et al., 2023; Ji et al., 2022).

Table 1: Some issues when using RWD or historical RCT data in an externally controlled trial.

Issue	Impact on Causal Inference
Immortal time bias	Misclassification of person-time at risk when comparing RWD with RCT.
Bias in healthcare	Healthcare system's bias are present in RWD (e.g. female vs male heart attack diagnosis).
Temporal bias	Historical data may incorporate biases as healthcare and treatments improve over time.
Missing data	RWD missing is complex: may be an artifact of patient care, inconsistent data capture, patient migration, healthcare access, or is truly missing.
Endpoint assessment	Physician assessments in clinical practice differ from RCT per protocol assessments. Some endpoints (e.g. death) are not typically part of EHR and may require data linkages
Timing of assessments	Physician assessments in clinical practice are not as frequent or regular as RCT, per protocol assessments (Adamson et al. 2022)
Data capture	What data is captured and how it is captured may differ from clinic to clinic; some data may be captured by primary care physician and not be available via a specialist (e.g. oncology). Result is inconsistency and bias.
Unblinded	Patients, physician nurses are all aware of the treatment given.

Opportunities with External Controls

Many publications since Pocock's article in 1976 provide robust solutions to the limitations and biases of externally controlled trials. Key publications (though not a systematic review) are provided in Table 2.

Table 2: Some publications on implementing external data into a clinical trial

Issue		Impact on Causal Inference
Pocock	(1976)	RCT historical control arms
Lodi	(2019)	3-step approach for the comparison of effect estimates from existing randomized trials and observational studies
Schmidli	(2019)	Methods for matching external control to single arm trial
Thorlund	(2020)	Specific considerations to address bias, matching and validation
Burcu	(2020)	Sources of data and considerations, including biases
Burger	(2021)	Very good summary of sources of bias and mitigation strategies
Hall	(2021)	Historical RCT as external control
Liu	(2021)	Propensity score methods and Bayesian meta-analytic-predictive (MAP) prior
Rahmund	(2021)	An overview on the use of external control data in design and analysis
Journal of biopharmaceutical statistics Volume 32 Issue 4	(2022)	Special Issue on Real World Evidence
Mhatre	(2022)	Endpoint assessment in oncology: real world PFS
Rippin	(2022)	Estimands for external controls
Ton	(2022)	Endpoint assessment in oncology OS, PFS and ORR
Incerti	(2022)	Endpoint assessment, time to event outcomes, historical RCT
Chen	(2023)	Use of RWD to inform trial design
Li	(2023)	Propensity Score Methods
Muller	(2023)	Bayesian approaches

Table 3 Recent examples of external data supporting submissions

Date (Reference) *type of external data*	**Selection Criteria**	**Results**
01JAN2015 - 20AUG2021 (Sola-Morales et al., 2023, Curtis et al., 2023) *RWD, HTD*	All **submissions** to FDA, EMA, HTA (IQWIG, NICE, G-BA, HAS)	•EMA 165 total submissions 26 (16%) with RWD or HTD •FDA 408 total submission 38 (9%) with RWD or HTD •HTA 772 total submissions 70 (9%) with RWD or HTD - IQWIG 12 - NICE 16 - G-BA 25 - HAS 17
2016–2021 (Wang et al., 2023) *RWD, HTD*	Oncology **approvals** by EMA	•103 total approvals 18 (17%) RWD or HTD submissions using 24 external control arms 15/18 were accepted by EMA
01JAN2019 to 30JUN2021 (Purpura et al., 2022) *RWD*	All NDA BLA **approvals** by FDA	•136 total approvals 116 (85%) any RWD 8 - primary 57 - supportive
2015 to 2020 (Arondekar et al., 2022) *RWD*	Oncology NDA BLA **approvals** FDA	•133 NDA BLA approvals 11 (8%) RWD •249 sNDA/ sBLA approvals 2 (<1%) RWD

IQWIG - Institute for Quality and Efficiency in Health Care; NICE - National Institute for Health and Care Excellence, G-BA Federal Joint Committee Germany, HAS - Haute Autorité de Santé France, NDA - New Drug Application; BLA - Biologics License Application; sNDA - supplemental NDA; sBLA -supplemental BLA

Data external to a clinical trial has successfully been included in regulatory and HTA submissions. See Table 3 for summaries on how external data has been used with FDA and European Medicines Agency (EMA) submissions and approvals as well as HTA submissions.

Notably, external controls have played a part in access decisions with HTA (Sola-Morales et al., 2023). Payers typically prioritize data that supports a patient's actual experience from RWD sources, which lends itself to external control in evidence packages submitted to HTA (Schad & Thronicke, 2022). Statisticians should proactively consider reimbursement questions and the need for RWD early in drug development, as RWD could be a critical part of the decision to reimburse (Curtis et al., 2023).

The use of external controls for internal decision making is difficult to ascertain from the literature, but very impactful for drug development. External controls can be used to determine if and when to move a molecule from early phase development to a later phase, whether phase 1b to 2 or phase 2 to 3. Since early phase studies are often lacking a concurrent control arm and have small sample sizes, RWD and HTD provide cost effective and ethically viable solutions for decision-making. RWD provides additional value to RCT planning in that it can inform on many aspects of trial planning such as inclusion and exclusion criteria, target population, unmet medical need, control arm assumptions, stratification factors.

The Future of External Controls

The future of external controls in clinical trials hinges on regulatory context, and also on statisticians continuing to innovate for early phase decision making and contributing to overall drug development with HTA in mind. Synthetic data (Azizi et al., 2021; Myles et al., 2021), virtual human twins (Sun et al., 2023; Viceconti et al., 2023), and *in silico* trials (Viceconti et al., 2021) are all now part of a drug developer's vocabulary and these methods may provide further paths forward as external controls.

Statisticians are working at an exciting time to pursue the use of externally controlled trials with a future outlook for drug development. Experimentation and commitment to scientific rigor can both be achieved as we continue to evolve the RCT. ■

References

Adamson, B. J., Ma, X., Griffith, S. D., Sweeney, E. M., Sarkar, S., & Bourla, A. B. (2022). Differential frequency in imaging-based outcome measurement: Bias in real-world oncology comparative-effectiveness studies. *Pharmacoepidemiology and Drug Safety*, 31(1), 46-54.Arondekar, B., Duh, M. S., Bhak, R. H., Der-Sarkissian, M., Huynh, L., Wang, K., Wojciehowski, J., Wu, M. V., Wornson, B., Niyazov, A., & Demetri, G. D. (2021). Real-World Evidence in Support of Oncology Product Registration: A Systematic Review of New Drug Application and Biologics License Application Approvals from 2015–2020. *Clinical Cancer Research,* 28(1), 27–35. https://doi.org/10.1158/1078-0432.ccr-21-2639

Azizi, Z., Zheng, C., Mosquera, L., Pilote, L., & Emam, K. E. (2021). Can synthetic data be a proxy for real clinical trial data? A validation study. *BMJ Open*, 11(4), e043497. https://doi.org/10.1136/bmjopen-2020-043497

Burcu, M., Dreyer, N. A., Franklin, J. M., Blum, M. J., Critchlow, C. W., Perfetto, E. M., & Zhou, W. (2020). Real□world evidence to support regulatory decision□making for medicines: Considerations for external control arms. *Pharmacoepidemiology and Drug Safety*, 29(10), 1228–1235. https://doi.org/10.1002/pds.4975

Burger, H. U., Gerlinger, C., Harbron, C., Koch, A., Posch, M., Rochon, J., & Schiel, A. (2021). The use of external controls: To what extent can it currently be recommended? *Pharmaceutical Statistics*, 20(6), 1002–1016. https://doi.org/10.1002/pst.2120

Chen, J., Ho, M., Lee, K. Y., Song, Y., Fang, Y., Goldstein, B. I., He, W., Irony, T., Jiang, Q., Van Der Laan, M. J., Lee, H., Lin, X., Meng, Z., Mishra-Kalyani, P. S., Rockhold, F. W., Wang, H., & White, R. (2021). The Current Landscape in Biostatistics of Real-World Data and Evidence: Clinical Study Design and Analysis. *Statistics in Biopharmaceutical Research*, 15(1), 29–42. https://doi.org/10.1080/19466315.2021.1883474

Collier, R. (2009). Legumes, lemons and streptomycin: A short history of the clinical trial. *Canadian Medical Association Journal*, 180(1), 23–24. https://doi.org/10.1503/cmaj.081879

Crofton, J., & Mitchison, D. A. (1948). Streptomycin resistance in pulmonary tuberculosis. *British medical journal,* 2(4588), 1009.

Curtis, L. H., Sola□Morales, O., Heidt, J., Saunders-Hastings, P., Walsh, L., Casso, D., Oliveria, S., Mercado, T., Zusterzeel, R., Sobel, R. E., Jalbert, J. J., Mastey, V., Harnett, J., & Quek, R. G. (2023). Regulatory and hta considerations for development of real-world data derived external controls. *Clinical Pharmacology & Therapeutics*. https://doi.org/10.1002/cpt.2913

Custom Data Solutions | Flatiron Health. (n.d.). https://flatiron.com/real-world-evidence/custom-data-solutions

Hall, K. T., Vase, L., Tobias, D. K., Dashti, H., Vollert, J., Kaptchuk, T. J., & Cook, N. R. (2020). Historical Controls in Randomized Clinical Trials: Opportunities and Challenges. *Clinical Pharmacology & Therapeutics,* 109(2), 343–351. https://doi.org/10.1002/cpt.1970

Historical Trial Data Sharing - TransCelerate. (2021, April 23). TransCelerate. https://www.transceleratebiopharmainc.com/initiatives/historical-trial-data-sharing/

Incerti, D., Bretscher, M. T., Lin, R., & Harbron, C. (2022). A Meta-analytic framework to adjust for bias in external control studies. *Pharmaceutical Statistics*, 22(1), 162–180. https://doi.org/10.1002/pst.2266

Ji, Z., Lin, J., & Lin, J. (2022). Optimal sample size determination for single-arm trials in pediatric and rare populations with Bayesian borrowing. *Journal of Biopharmaceutical Statistics*, 32(4), 529–546. https://doi.org/10.1080/10543406.2022.2058529

Lambert, J., Lengliné, E., Porcher, R., Thiébaut, R., Zohar, S., & Chevret, S. (2022). Enriching single-arm clinical trials with external controls: possibilities and pitfalls. *Blood Advances*. https://doi.org/10.1182/bloodadvances.2022009167

Li, H., & Yue, L. Q. (2023). Propensity score-based methods for causal inference and external data leveraging in regulatory settings: From basic ideas to implementation. *Pharmaceutical Statistics*. https://doi.org/10.1002/pst.2294

Liu, M., Bunn, V., Hupf, B., Lin, J., & Lin, J. (2021). Propensity-score-based meta-analytic predictive prior for incorporating real-world and historical data. *Statistics in Medicine*, 40(22), 4794–4808. https://doi.org/10.1002/sim.9095

Lodi, S., Phillips, A. N., Lundgren, J. D., Logan, R., Sharma, S., Cole, S. R., Babiker, A., Law, M., Chu, H., Byrne, D., Horban, A., Sterne, J. a. C., Porter, K., Sabin,

C. A., Costagliola, D., Abgrall, S., Gill, J. S., Touloumi, G., Pacheco, A. G., . . . Hernán, M. A. (2019). Effect Estimates in Randomized Trials and Observational Studies: Comparing Apples With Apples. *American Journal of Epidemiology*, 188(8), 1569–1577. https://doi.org/10.1093/aje/kwz100

Myles, P., Ordish, J., & Branson, R. (2021). Synthetic data and the innovation, assessment, and regulation of AI medical devices. *RF Q.*, 1, 48-53.

Pocock, S. J. (1976). The combination of randomized and historical controls in clinical trials. *Journal of Chronic Diseases,* 29(3), 175–188. https://doi.org/10.1016/0021-9681(76)90044-8

Purpura, C. A., Garry, E. M., Honig, N., Case, A., & Rassen, J. A. (2021). The Role of Real-World Evidence in FDA-Approved New Drug and Biologics License Applications. *Clinical Pharmacology & Therapeutics*, 111(1), 135–144. https://doi.org/10.1002/cpt.2474

Rahman, R., Ventz, S., McDunn, J., Louv, B., Reyes-Rivera, I., Polley, M. Y. C., Merchant, F. J. A., Abrey, L. E., Allen, J. E., Aguilar, L. K., Aguilar-Cordova, E., Arons, D., Tanner, K. G., Bagley, S. J., Khasraw, M., Cloughesy, T. F., Wen, P. Y., Alexander, B. M., & Trippa, L. (2021). Leveraging external data in the design and analysis of clinical trials in neuro-oncology. *Lancet Oncology*, 22(10), e456–e465. https://doi.org/10.1016/s1470-2045(21)00488-5

Rippin, G., Ballarini, N. M., Sanz, H., Largent, J., Quinten, C., & Pignatti, F. (2022). A Review of Causal Inference for External Comparator Arm Studies. *Drug Safety*, 45(8), 815–837. https://doi.org/10.1007/s40264-022-01206-y

Schad, F., & Thronicke, A. (2022). Real-World Evidence—Current Developments and Perspectives. *International Journal of Environmental Research and Public Health*, 19(16), 10159. https://doi.org/10.3390/ijerph191610159

Schmidli, H., Häring, D. A., Thomas, M., Cassidy, A., Weber, S. A., & Bretz, F. (2019). Beyond Randomized Clinical Trials: Use of External Controls. *Clinical Pharmacology & Therapeutics*, 107(4), 806–816. https://doi.org/10.1002/cpt.1723

Sola-Morales, O., Curtis, L. H., Heidt, J., Walsh, L., Casso, D., Oliveria, S., Saunders-Hastings, P., Song, Y., Mercado, T., Zusterzeel, R., Mastey, V., Harnett, J., & Quek, R. G. W. (2023). Effectively Leveraging RWD for External Controls: A Systematic Literature Review of Regulatory & HTA Decisions. *Clinical Pharmacology & Therapeutics*. https://doi.org/10.1002/cpt.2914

Thorlund, K., Dron, L., Park, J. J. H., & Mills, E. J. (2020). Synthetic and External Controls in Clinical Trials – A Primer for Researchers. *Clinical Epidemiology, Volume 12*, 457–467. https://doi.org/10.2147/clep.s242097

Ton, T. G., Pal, N., Trinh, H., Mahrus, S., Bretscher, M. T., Machado, R. a. M., Sadetsky, N., Chaudhary, N., Lu, M. T., & Riely, G. J. (2022). Replication of Overall Survival, Progression-Free Survival, and Overall Response in Chemotherapy Arms of Non-Small Cell Lung Cancer Trials Using Real-World Data. *Clinical Cancer Research*, *28*(13), 2844–2853. https://doi.org/10.1158/1078-0432.ccr-22-0471

U.S. Food and Drug Administration. (2023) Considerations for the design and conduct of externally controlled trials for drug and biological products. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-design-and-conduct-externally-controlled-trials-drug-and-biological-products

Viceconti, M., Pappalardo, F., Rodriguez, B., Horner, M., Bischoff, J., & Tshinanu, F. M. (2021). In silico trials: Verification, validation and uncertainty quantification of predictive models used in the regulatory evaluation of biomedical products. Methods, 185, 120-127.

Viceconti, M., De Vos, M., Mellone, S., & Geris, L. (2023). From the digital twins in healthcare to the Virtual Human Twin: a moon-shot project for digital health research. arXiv preprint arXiv:2304.06678.

Vivli. (2023, June 10). Home – *GREI - Vivli*. Vivli - Center for Global Research Data. https://vivli.org/

Wang, J. (2021). To use or not to use propensity score matching? *Pharmaceutical Statistics*, *20*(1), 15–24. https://doi.org/10.1002/pst.2051

Wang, X., Dormont, F., Lorenzato, C., Latouche, A., Hernandez, R., & Rouzier, R. (2023). Current Perspectives for External Control Arms in Oncology Clinical Trials: Analysis of EMA approvals 2016-2021. *Journal of Cancer Policy*, 100403.

Practical Considerations For Adaptive Design

Keaven M. Anderson
Merck & Co., Inc., Rahway, NJ, USA
keaven_anderson@merck.com

Sabrina Wan
Merck & Co., Inc., Rahway, NJ, USA
shuyan.wan@merck.com

Hongtao Zhang
Merck & Co., Inc., Rahway, NJ, USA
hongtao.zhang1@merck.com

ABSTRACT

Rather than scientific approaches, we focus on practical considerations concerning design and implementation of adaptive clinical trials.

1. INTRODUCTION

Adaptive design may be considered when a trial needs to answer multiple questions. For example, a trial may select the most appropriate population or dose(s) at an interim analysis to save sample size compared to a design where all patients or doses continue throughout the trial. Asking about scientific issues is not sufficient to select a design, however. There are many practical considerations that need attention. This note will focus on these practical considerations as there is a very extensive literature on methods to consider.

2. ABBREVIATED LITERATURE REVIEW

There is a great deal of literature on adaptive design. One starting point of this field could be [1] as noted in the review article by [2] which summarizes adaptive design concepts over the following 25 years. Adaptive design has had considerable interest in the pharmaceutical industry. For example, the cross-industry Adaptive Design Working Group was initiated in 2005, resulting in several review papers (e.g., [7]). This and substantial academic interest have likely been motivation for regulatory guidance (e.g., [3], [8]). The types of practical considerations considered here have also been summarized by others ([4], [11]).

The primary focus here will be on randomized trials considering a single treatment strategy. Some adaptations that might be considered are blinded or unblinded sample size adaptation, population selection, or dose selection. Readers interested in basket, umbrella and platform trials will not find them here; consider [10] or [9], although there are many references available. There is also substantial literature on early-stage designs such as dose finding that is not covered here.

3. BASIC QUESTIONS TO CONSIDER WHEN SELECTING A DESIGN

When starting with a clinical design concept, it is wise to begin with which questions you need to answer. Questions evaluate not only methodology, but the practicality of implementing a design.

3.1 Does adaptation slow planning?

Lack of familiarity with designs can slow planning. Having standard design approaches, software tools and protocol templates available can expedite planning and approval of adaptive studies. Custom programming, large simulation needs that require complex summaries, and getting clinical trial teams, management and even regulators on board can add months to planning for a trial. Having internal and/or external expertise familiar with methods and implementation is a plus. [5] have suggested reporting standards for adaptive designs.

3.2 Does adaptation slow execution?

The standard operating model for a trial team and management will be to expedite quick trial completion. Operations teams may typically open as many study sites and enroll as many patients as fast as possible. Adaptive designs on the other hand, require follow-up data to obtain information to inform adaptation to the trial team. This may require longer follow-up on patients enrolled more slowly to be able to effectively adapt. Being able to collect and clean follow-up data efficiently on an ongoing basis is critical, so operations must be efficient. There is likely to be more work involved in adaptive design execution than in non-adaptive trials.

3.3 If you are trying to minimize costs, do you need to consider more than the expected number of patients to be enrolled?

If running a Phase 2/3 trial with the idea of saving costs with an early futility decision, it can be important to not open too many sites before a decision to go to Phase 3. Site activation can be expensive and accelerate both site costs and potential patient costs if many participants are enrolled before making an adaptation. A critical consideration is whether to continue enrollment of patients while data is cleaned for the Phase 2 analysis used to make adaptation. Pausing enrollment can save money but with the possibility that investigators may discontinue their participation. Continuing enrollment can over-enroll patients after the Phase 2 decision-making data that cannot be used in the formal Phase 2/3 inference. The data from these patients may not be consistent with the Phase 2 data used to make an adaptive decision, but this is only known after the adaptation. Cleaning of the data for Phase 2 decision needs to be done quickly with high quality to ensure appropriate and timely Phase 2 decisions. These are important caveats to the objective of expediting trial completion to accelerate program development. This also suggests simulation and other planning software might benefit by considering more than just sample size in determining trial costs.

3.4 Is interim decision-making done appropriately?

Adaptive decision-making at the time of interim can be complex. Since interim data is limited, decisions can be ambiguous. Integrity of eventual conclusions can depend on keeping operations teams blinded and avoiding bias introduced by adaptation or execution decisions being based on considerations other than the limitations imposed by the design. Independent data monitoring committees (DMCs) may not be accustomed to making strategic interim decisions.

3.5 Would operationally seamless or inferentially seamless adaptation better suit your needs?

An inferentially seamless trial is a trial where there is some adaptation made, but data before and after the adaptation is combined for inference and estimation. The thought is often that this can replace separate Phase 2 and Phase 3 trials with a combined Phase 2/3 design that answers some key Phase 2 questions before further focusing the Phase 3 design, e.g., select and confirm the best dose in a single trial. An inferentially seamless adaptive design has the potential to reduce sample size compared to the total of separate trials as well as to accelerate time to completion. As noted above, planning time can slow an inferentially seamless trial due to trial complexity.

An operationally seamless trial would combine patient enrollment for Phase 2 and Phase 3 in a single trial, but analysis of the two phases is done separately. This allows a complete examination of Phase 2 data to fine tune multiple design features for Phase 3. Combining into a single trial still has the advantages of only writing a single protocol and only operating a single trial. We note that an inferentially seamless trial can be converted to an operationally seamless trial at the time of the analysis for adaptation if the data at that time suggests more changes than can be considered in the inferentially seamless design.

3.6 Is the data used to adapt likely to be representative of the later part of the trial that is being adapted?

We consider unblinded sample size re-estimation based on an interim treatment effect. This is an example of adaptive design where homogeneity is required for adaptations to work as expected. That is, unblinded sample size estimation assumes the same treatment effect for observations before and after adaptation. During the COVID-19 epidemic we saw disease variants evolve every few months, often with a different prognosis for each variant. A treatment may not be equally effective for different variants as the target of a treatment drug may vary by virus variant. New countries and sites opening as a trial proceeds are examples that may introduce heterogeneity in treatment effect over time. Different sites simply enrolling patients with different prognoses may be important. Much was learned concerning practical matters on how to manage patients; see, for example, [6]. In the scenario above where we suggested not opening sites too fast prior to adaptation, we would have the potential downside that Phase 3 sites could produce different results than Phase 2, invalidating the assumptions of the adaptive methods applied.

3.7 Can simpler methods be more effective?

For the above example, our personal preference would be a group sequential design with early stopping to do adaptation by stopping for futility if a new treatment is ineffective or early stopping for efficacy if a new treatment is highly effective. At the time of trial design, the actual treatment effect is speculative. Thus, either ineffectiveness or high effectiveness are not out of the realm of possibilities. In addition, group sequential design is well understood with both free and commercial software available to enable quick planning to get a trial design written, reviewed and implemented.

3.8 Can you realistically expect to adapt doses or treatments studied in the middle of the trials?

If there are minor difference in effectiveness and safety for doses at an interim analysis, is it appropriate to choose a single dose going forward? That is, continuing multiple experimental groups to the end of the trial may be needed to confirm one or more safe and effective treatment. The thought experiment about this can be quite important to enable prompt and effective decision making at the time of an interim analysis.

A specific variant of this design may be considered for a combination of two treatments. It may not be well-established whether one or both treatments are required for effectiveness. A Phase 2/3 factorial design can be an effective way to limit investment in monotherapy arms that expected to be ineffective relative to the combination. That is, a trial sponsor could discuss with regulators if having little or no trend for improved outcomes with monotherapy (either single component of a combination) in the Phase 2 part of a trial is sufficient to omit them from the Phase 3 portion of the trial due to lack of efficacy. In this case the combined treatment would be the only experimental treatment continued in the Phase 3 portion of the trial. The question for regulators would be whether this could be sufficient to establish that each component of the combination therapy is contributing benefit.

3.9 Is futility analysis a good option?

Incorporating futility analysis in clinical trials can be an important way to limit investment in ineffective treatments, saving more to invest in more effective alternatives. Incorporating futility seems like a particularly simple adaptation, but considerations like delayed treatment effects, conditional power, increased Type II error (power reduction) and timing of analyses are worth careful thought. Waiting until something like 40% of planned statistical information, 85% power, and futility using an aggressive spending function can be a possibility. This allows a futility bound requiring a positive interim treatment trend without losing too much power. Requiring a positive trend to continue past the interim can reassure the large final investment is worthwhile. However, the tradeoffs for such a design make a difficult challenge and will vary on a trial-by-trial basis.

4. SUMMARY

Adaptive design can be a valuable tool in drug development to efficiently answer a number of questions required to bring an effective and safe new treatment to patients. However, with the complexity that adaptation brings substantial thought experiments are required to ensure effective designs are selected. Tradeoffs in complexity, duration of planning, risks of incorrect assumptions all need to be considered in the context of potential gains produced by effective adaptation.

5. REFERENCES

[1] Bauer, P. 1989. Multistage testing with adaptive designs. *Biometrie und Informatik in Medizin und Biologie*. 20, 4 (1989), 130–148.

[2] Bauer, P. et al. 2016. Twenty-five years of confirmatory adaptive designs: Opportunities and pitfalls. *Statistics in Medicine*. 35, 3 (2016), 325–347.

[3] Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research 2019. Adaptive designs for clinical trials of drugs and biologics. *United States Food and Drug Administration*. (2019).

[4] Coffey, C.S. et al. 2012. Overview, hurdles, and future work in adaptive designs: Perspectives from a national institutes of health-funded workshop. *Clinical Trials*. 9, 6 (2012), 671–680.

[5] Dimairo, M. et al. 2020. The adaptive designs CONSORT extension (ACE) statement: A checklist with explanation and elaboration guideline for reporting randomised trials that use an adaptive design. *bmj*. 369, (2020).

[6] Fadel, R. et al. 2020. Early short-course corticosteroids in hospitalized patients with COVID-19. *Clinical Infectious Diseases*. 71, 16 (2020), 2114–2120.

[7] Gallo, P. et al. 2006. Adaptive designs in clinical drug development—an executive summary of the PhRMA working group. *Journal of biopharmaceutical statistics*. 16, 3 (2006), 275–283.

[8] Medicinal Products for Human Use, C. for et al. 2007. Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design. *London: EMEA*. (2007).

[9] Park, J.J. et al. 2020. An overview of precision oncology basket and umbrella trials for clinicians. *CA: a cancer journal for clinicians*. 70, 2 (2020), 125–137.

[10] Park, J.J. et al. 2019. Systematic review of basket trials, umbrella trials, and platform trials: A landscape analysis of master protocols. *Trials*. 20, 1 (2019), 1–10.

[11] Quinlan, J. et al. 2010. Barriers and opportunities for implementation of adaptive designs in pharmaceutical product development. *Clinical trials*. 7, 2 (2010), 167–173.

Oncology Dose Optimization in Early-Phase Trials: A New Dawn

Yuan Ji and Dehua Bi

Department of Public Health Sciences The University of Chicago

Introduction

Cytotoxic oncology drugs (e.g., chemo-therapies) exert their efficacy through mechanisms that directly induce cell death, cancerous or not. Typically, a higher dose leads to more cell death, which leads to high efficacy and high toxicity. Therefore, for cytotoxic drugs the optimal dose for patients is the maximum tolerated dose (MTD), since it produces the highest efficacy among all the doses that can be tolerated. As a result, the vast majority of existing oncology phase I trials record the dose-liming toxicity (DLT) as the primary outcome [Ji et al., 2018] and aim to identify the MTD, the highest dose with no more than a DLT probability typically around 20% to 30%. Once the MTD is identified, the dose is often used in phase II and phase III trials for further testing of the drug's efficacy and effectiveness.

The above MTD-centric strategy (MCS) for early-phase oncology drug development has been the gold standards for decades since almost all the traditional oncology drugs are cytotoxic. An unintended consequence of this is that investigators and drug developers start to overlook the premise of MCS – efficacy increasing with dose – and instead view the MTD as the default optimal dose regardless the mechanism of action for an oncology drug. When the premise is not true, MTD is not necessarily the optimal dose.

Due to the explosive advancement in biological and genomics research since the human genome [Consortium, 2001, Venter et al., 2001] was sequenced in the early 2000's, oncology drugs have switched from directly attacking cells based on cytotoxic means to precisely targeting biological processes at the molecular level such as genetic and immune pathways. For example, PD-1 inhibitors work by blocking the PD-1 receptor on T cells, allowing these immune cells to recognize and kill cancer cells [Robert et al., 2014]. Since PD-1 inhibitors fight tumor by blocking as many PD-1 receptors as required to stimulate the immune response, exceeding the necessary amount does not necessarily enhance their therapeutic effects. Thus, the optimal therapeutic dose may not be the MTD, but instead a lower dose that achieves maximum receptor blocking with minimal side effects.

The US FDA's Project Optimus [FDA, 2023b] aims to adapt the approach of clinical trials to the new realities of cancer treatment. Under this project, the FDA encourages the development and application of novel trial designs and statistical methods that attempt to identify the optimal dose of oncology drugs instead the MTD. Several publications [Shah et al., 2022a,b, Zirkelbach et al., 2022] and an an FDA draft guidance [FDA, 2023a] have called

for changes to early-phase clinical trial designs. See Figure 1 for a summary of the draft guidance. In this article, I will review and present related statistical designs and methods for optimal dose finding in oncology, the challenges the new paradigm imposes, the need for Bayesian models and designs, and suggestions for future work.

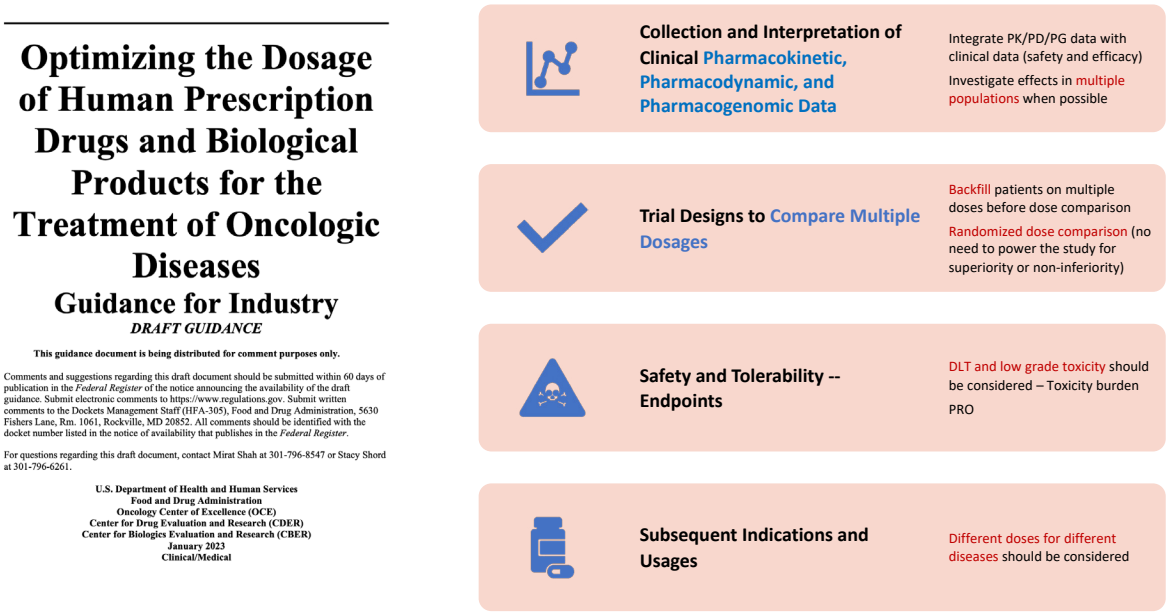
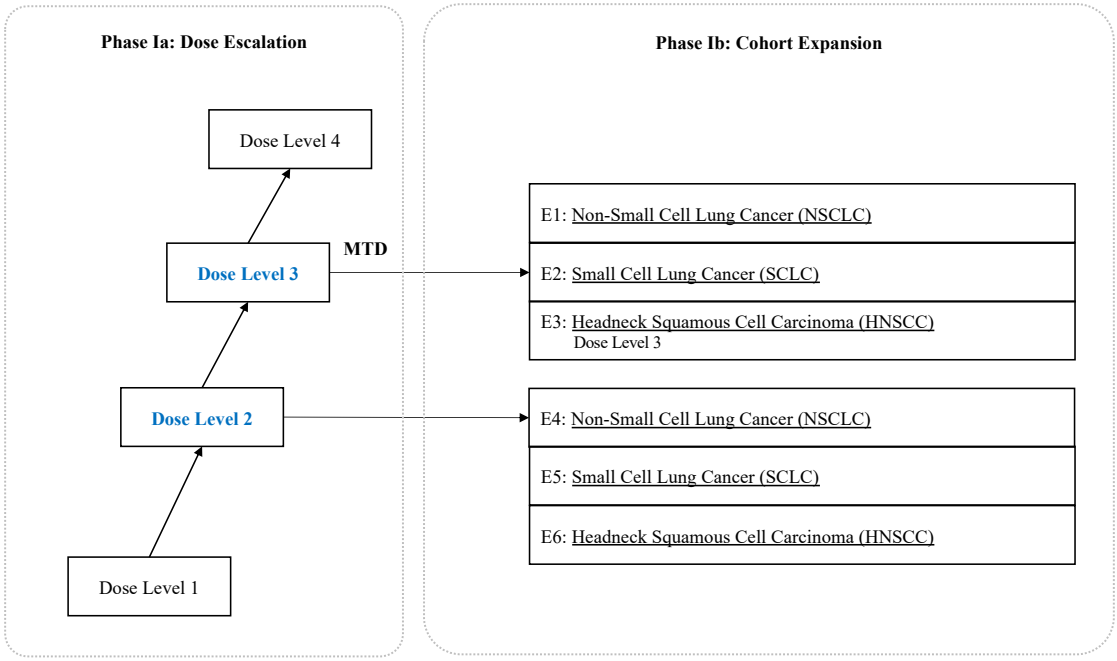


Figure 1: Summary of the FDA draft guidance on dose optimization.

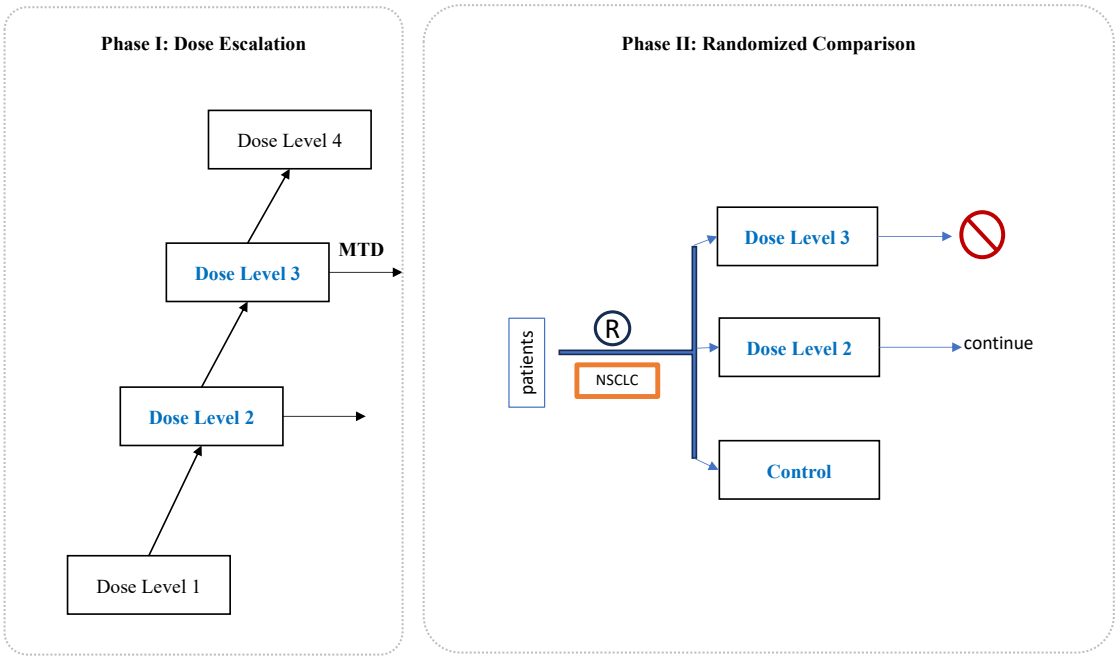
Past, Present and Future

Past For statisticians, it is not a new concept to find the optimal dose instead of the MTD as the primary objective for a dose-finding trial. There is a vast statistical literature of designs for finding an optimal biological dose (OBD) in early-phase oncology trials (Thall and Cook [2004], Li et al. [2017], Lin and Ji [2020, 2021], Zhou et al. [2019], among many others). The main idea in these EffTox designs is to model the joint efficacy and toxicity outcomes rather than just the DLT. By modeling the joint outcomes, the EffTox designs compare doses based on a tradeoff between efficacy and toxicity, thus achieving dose optimization. For example, the tradeoff may be $Pr(p_d < p_T, q_d > q_E|data)$ or a utility function $U(p_d, q_d) > 0$ that increases with q_d and decreases with p_d, where, p_d and q_d are the toxicity and efficacy probabilities of dose d, and p_T and q_E are some thresholds for toxicity and efficacy.

While the EffTox designs address dose optimization by modeling the joint efficacy and toxicity endpoints, they encounter a practical issue that hinders them from being easily and widely applied in real-world trials. In oncology, the binary efficacy outcome is usually based on tumor shrinkage, e.g., defined by RECIST [Eisenhauer et al., 2009], which takes 10-12 weeks post-treatment to measure. In contrast, the DLT is typically measured within the first treatment



(a) Phase Ia/Ib seamless



(b) Phase I-II seamless

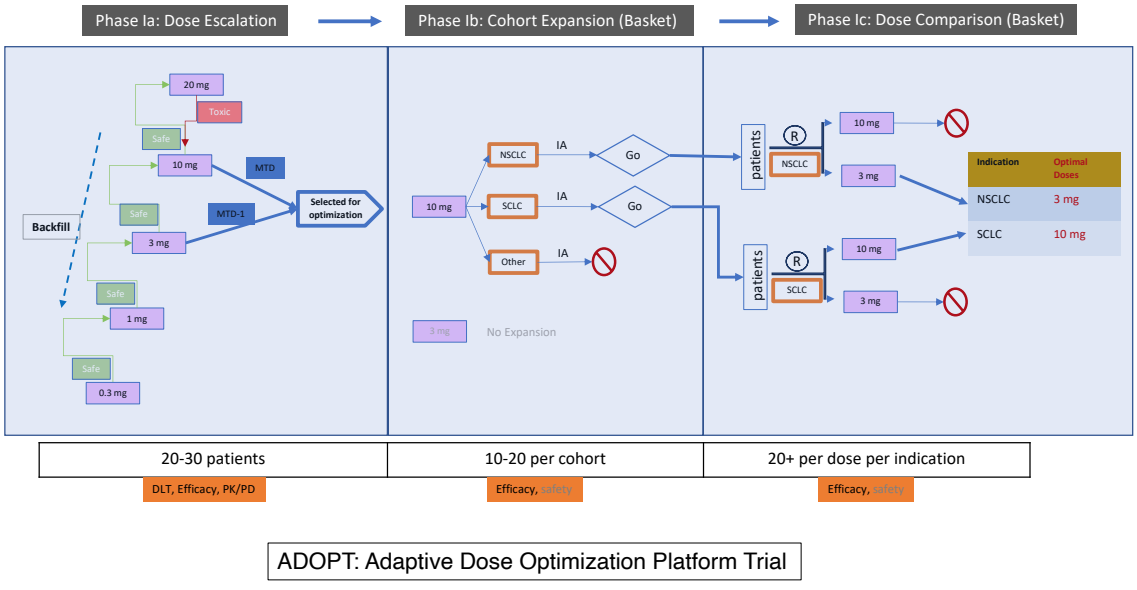
Figure 2: Seamless designs: (a) Phase Ia/Ib seamless, and (b) Phase I/II seamless. A phase Ia/Ib seamless design consists of dose escalation (phase Ia) and cohort expansion (phase 1b). A Phase I-II seamless design consists of dose escalation (phase I) and randomized dose comparison (phase II). In both seamless designs, the first stage is the same, a dose escalation trial, but labeled differently as phase Ia or phase I.

cycle, around 3-4 weeks post-treatment. Therefore, to apply the EffTox designs, one needs to wait much longer after treating each patient cohort than toxicity-based designs before making a dosing decision for the subsequent cohort. This delay is prohibitively long for practical implementation.

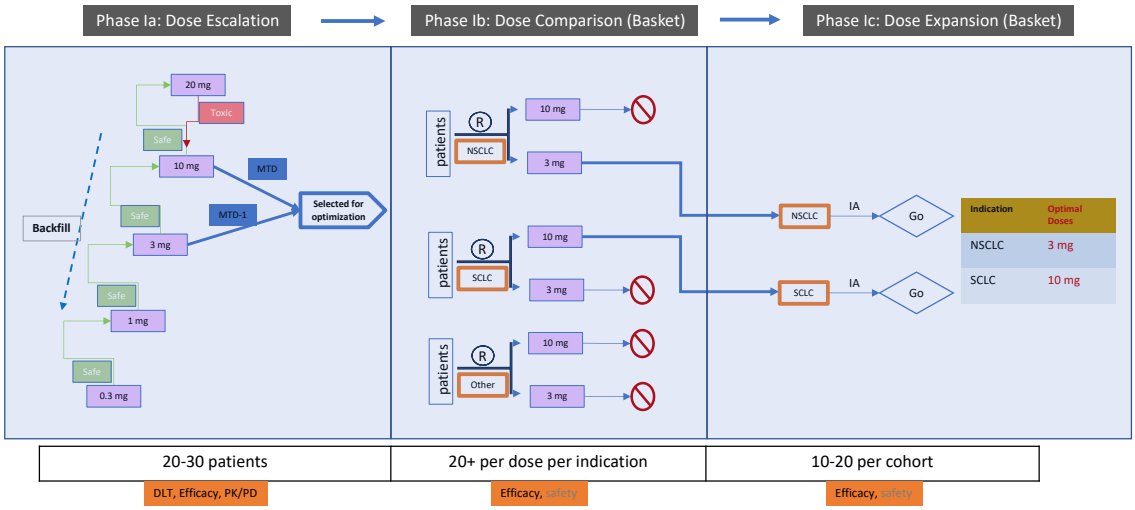
Alternatively, seamless phase Ia/Ib designs and phase I-II designs[Pan et al., 2014, Hoering et al., 2011] have been developed to overcome the issue caused by the long waiting time for efficacy outcome in the EffTox designs. For seamless phase 1a/1b designs (Figure 2a), the first stage takes the form of a standard phase I clinical trial. In the second stage, patients are enrolled as expansion cohorts [Lyu et al., 2023] at one or more indications to test the efficacy of one or more selected doses from the first stage. This forms a two-dimensional basket. Interim analyses and stopping rules may be applied to stop the expansion early due to futility, and toxicity monitoring is preferred so that a cohort with unacceptable toxicity may be terminated early. In contrast, for seamless phase I-II designs, a slightly different but important modification to the phase Ia/Ib design is the randomized phase II comparing two doses in the presence of a control arm. See Figure 2b.

Many other designs have also been developed in the research literature. It is impossible to cover all of them. In general, these advanced designs all recognize the issue of MCS and instead aim to select a dose that provides optimal benefits to patients. These past works provide a solid foundation for Project Optimus to quickly hit the ground running.

Present Figure 3 illustrates a novel scheme for early-phase dose optimization. We call this scheme ADOPT, standing for Adaptive Dose Optimization Platform Trial. ADOPT is structured as a phase I trial consisting of three seamless sub-phases, Ia, Ib, and Ic. Two versions of ADOPT are presented in Figure 3, denoted as ADOPT-V1 and ADOPT-V2. In both versions, phase Ia represents an improved dose escalation highlighted by novel features like patient backfill and the use of PK/PD data. At the end of phase Ia, doses 10mg (the MTD) and 3mg (the dose below MTD) are selected and sent to phases Ib and Ic for testing of efficacy. ADOPT-V1 (Figure 3a applies the MATS design [Jiang et al., 2023] for phases Ib and Ic. Specifically, phase Ib expands the higher dose 10mg in three indications, performs an interim analysis (IA) to determine which indication will be selected for a randomized comparison of 10mg and 3mg in phase Ic. ADOPT-V2 reverses the order of dose expansion and randomized comparison. The two versions of ADOPT may be suitable for different drug development programs and mechanisms of action. For example, if it is strongly believed that the higher dose is more efficacious than the lower dose, V1 might be a better design since it only tests the lower dose (in phase Ic) when the higher dose demonstrates promising efficacy. Otherwise, V2 (Figure 3b) might be preferred, which allows randomized comparison between the two doses immediately after dose escalation in phase Ia.



(a) ADOPT-V1



(b) ADOPT-V2

Figure 3: A stylized illustration of the Adaptive Dose Optimization Platform Trial (ADOPT). It consists of three seamless phases, Ia, Ib, and Ic. Phase Ia is for dose escalation. Phases Ib and Ic are basket trials for expansion and randomized dose comparison. Novel features like backfill and integration of PK/PD data can be considered in phase Ia. The order of phases Ib and Ic may change depending on specific settings in practice, shown as the two versions V1 in (a) and V2 in (b). In the end, different indications may have different optimal doses. For example, 3mg for NSCLC and 10mg for SCLC are selected as the optimal doses.

While ADOPT provides a framework for dose optimization and addresses some key points listed in the FDA draft guidance (Figure 1), it imposes practical and logistic challenges. For example, the sample size for ADOPT will be much larger than a conventional oncology phase I trial (see suggested sample sizes in Figure 3). Also, ADOPT is complex and requires dedication from the study team to ensure high-quality trial conduct. However, the benefits are clear. Through ADOPT, investigators and sponsors will accumulate substantial information about the tested drug and doses, gaining high confidence in the optimal doses selected for different indications. This increases the probability of success in late-phase trials. Various applications of ADOPT have already been attempted, as seen in some recent works in the literature such as patients backfilling in dose escalation [Dehbi et al., 2021, Liu et al., 2023] and seamless phase Ib/Ic for dose expansion/optimization[Jiang et al., 2023].

Project Optimus also emphasizes the importance of utilizing pharmacological data for dose optimization. Traditionally, PK/PD data are analyzed separately from clinical outcomes. However, designs and methods [Ursino et al., 2017, Su et al., 2022] that combine pharmacological observations and clinical outcomes have been explored and developed to improve the efficiency of dose selection.

Future The future is bright for finding an optimal dose for a new oncology drug. An early-phase oncology trial will resemble a master protocol like ADOPT which packs dose escalation, expansion, and randomized comparison in a single study. However, many details and questions are yet to be addressed before this becomes a reality, for example, the optimal sample size allocation for each stage in ADOPT. If the overall sample size for the entire trial is fixed, it is useful to find an optimal ratio of patient allocation to the three stages in ADOPT so that the overall trial efficiency is maximized. Also, the order of phase Ib and Ic may be switched, allowing randomized comparison of two doses across indications followed by expansion of a winning dose for each indication. These questions will lead to new research, which subsequently will improve the quality of drug development and patient care.

A challenge that will need to be addressed by the oncology pharmaceutical industry is the need to increase investment in resources and time for phase I trials. The new ADOPT trials will inevitably require more money and more time than the traditional phase I MTD dose-finding trials. This new challenge will disproportionally affect large pharma and small biotech, as the latter is typically constrained by reaching short-term milestones like completing a traditional dose-finding trial. It is imperative to adapt the investment mindset and milestones in the new era of dose optimization. This is because while early-phase oncology trials will now demand more upfront investment, they are expected to enhance the overall probability of success of oncology drug development by finding the best dose.

Be Bayesian

Project Optimus needs appropriate statistical models to empower trial designs and data analyses. Bayesian models and designs are the ideal choice to address new challenges and meet objectives from the project. Bayesian dose-finding designs have been routinely applied in practice (Ji et al. [2010], Liu and Yuan [2015], O'Quigley et al. [1990], Neuenschwander et al. [2008], among many others). The newly developed backfill designs also rely on Bayesian models [Dehbi et al., 2021, Liu et al., 2023] to allow more doses to be explored during a phase I trial. Furthermore, Bayesian designs and models can naturally integrate PK/PD and pharmacogenetics (PG) information and model efficacy of multiple doses in multiple populations [Ursino et al., 2017, Su et al., 2022]. And last but definitely not the least, randomized dose comparison, a novel and crucial part of Project Optimus, is best suited for Bayesian designs. The FDA draft guidance specifically states that "The trial does not need to be powered to demonstrate statistical superiority of a dosage or statistical non-inferiority among the doses." It actually is infeasible to do that. For example, based on a z-test, at the significance level of 0.1, with a superiority margin of 0.05, it would require 812 subjects per dose to achieve 80% power when the response rate for one dose is 0.3 and 0.4 for the other. A sample size of this magnitude is clearly too large for a phase I trial. Typically, sponsors and investigators are willing to consider a few dozen subjects per dose in the randomized comparison, which apparently would lead to miserable power and large type I error rates under standard frequentist calculation. Moreover, the goal of randomized comparison of selected doses is to test non-inferiority of the low dose in terms of efficacy, which would require a larger sample size under frequentist theory.

Bayesian designs and models justify a small sample size based on simple posterior calculations and simulations. For example, Figure 4 presents a heatmap of $\xi(n, \delta) = Pr(q_2 - q_1 > \delta \mid n, y_1, y_2)$, which calculates the posterior probability of the efficacy probability q_2 of the higher dose is larger than the lower dose q_1 by a margine of $\delta > 0$, assuming that n patients are randomized to each of the two doses and y_1 and y_2 patients respond with efficacy, respectively. One could then specify a threshold s, so that when $\xi(n, \delta) > s$ the higher dose is selected, and if not, the lower dose is selected. For example, let $n = 20$ be the sample size for each dose. And if $\delta = 0.05$ and if there are $y_2 = 10$ responders in the low dose, then with $s = 0.6, 0.7, 0.8$, or 0.9, in order to declare the higher dose is not more efficacious than the lower dose by a margin of $\delta = 0.05$, the higher dose must not have more than 10, 11, 12, or 13 responders. Also, in the MATS design [Jiang et al., 2023] the authors report the operating characteristics of the Bayesian design for different sample sizes, ranging from 20 to 40 per dose. This type of planning is likely to help investigators to decide an appropriate sample size for dose-optimization trials.

Future research on sample size planning based on Bayesian inference is needed to allow sponsors, investigators, and regulators to assess the risk and benefit tradeoffs for running the trial with different sample sizes. The main

objective would be to come up with a new framework and provide assurance of quality and reliability of the trial for a given sample size. Adaptive designs allowing sample size re-estimation may also be desirable since strict type I error control is no longer necessary. Instead, sponsors should be more focused on the probability of making a wrong decision in terms of dose and indication selection, which can be assessed through Bayesian multiplicity control [Scott and Berger, 2010].

Conclusion and Discussion

Project Optimus is heralding a new dawn for deciding an appropriate dose for novel oncology drugs. The movement is caused by the rapid advancements in scientific research revealing new biological and molecular mechanisms of carcinogenesis and progression. As a result, cancer drugs are changing their ways of eradicating tumor cells. The traditional MTD-centric approach is outdated, and oncology drug development is shifting towards finding an optimal dose that might be lower than the MTD and maximizes patient benefits. Given this shift, statistical designs for dose-finding trials need to adapt and novel methodologies must be developed.

Together with opportunities come challenges. A main question is how to efficiently perform a phase I oncology dose-optimization trial. This requires a new mindset and innovation. In this short article, I have reviewed and proposed some ideas that could help advance the statistical research for dose optimization. In particular, the proposed ADOPT framework may serve as a starting point for future trial designs. It is an exciting new era for statisticians working in cancer drug development. Many new problems and challenges will arise and it is upon the statisticians to help provide sound and efficient solutions.

References

International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*, 409:860–921, 2001. doi: 10.1038/35057062.

Hakim-Moulay Dehbi, John O'Quigley, and Alexia Iasonos. Controlled backfill in oncology dose-finding trials. *Contemporary Clinical Trials*, 111:106605, 2021.

E.A. Eisenhauer, P. Therasse, J. Bogaerts, L.H. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney, L. Rubinstein, L. Shankar, L. Dodd, R. Kaplan, D. Lacombe, and J. Verweij. New response evaluation criteria in solid tumours: revised recist guideline (version 1.1). *European Journal of Cancer*, 45(2):228–247, 2009.

US FDA. Optimizing the dosage of human prescription drugs and biological products for the treatment of oncologic

diseases guidance for industry. Technical report, U.S. Food and Drug Administration, 2023a. URL https://www.fda.gov/media/164555/download.

US FDA. Project optimus, 2023b. URL https://www.fda.gov/about-fda/oncology-center-excellence/project-optimus.

Antje Hoering, Mike LeBlanc, and John Crowley. Seamless phase i-ii trial design for assessing toxicity and efficacy for targeted agents. *Clinical Cancer Research*, 17(4):640–646, 2011.

Yan Ji, Jin Y Jin, David M Hyman, Geoffrey Kim, and Ajit Suri. Challenges and opportunities in dose finding in oncology and immuno-oncology. *Clinical and Translational Science*, 11(4):345–351, 2018.

Yuan Ji, Ping Liu, Yisheng Li, and B Nebiyou Bekele. A modified toxicity probability interval method for dose-finding trials. *Clinical trials*, 7(6):653–663, 2010.

Zhenghao Jiang, Gu Mi, Ji Lin, Christelle Lorenzato, and Yuan Ji. A multi-arm two-stage (mats) design for proof-of-concept and dose optimization in early-phase oncology trials. *Contemporary Clinical Trial*, page In press, 2023.

D. H. Li, J. B. Whitmore, W. Guo, and Y. Ji. Toxicity and efficacy probability interval design for phase i adoptive cell therapy dose-finding clinical trials. *Clinical Cancer Research*, 23(1):13–20, 2017.

Xiaolei Lin and Yuan Ji. The joint i3+ 3 (ji3+ 3) design for phase i/ii adoptive cell therapy clinical trials. *Journal of Biopharmaceutical Statistics*, 30(6):993–1005, 2020.

Xiaolei Lin and Yuan Ji. Probability intervals of toxicity and efficacy design for dose-finding clinical trials in oncology. *Statistical Methods in Medical Research*, 30(3):843–856, 2021.

Jiaxin Liu, Shijie Yuan, B Nebiyou Bekele, and Yuan Ji. The backfill i3+ 3 design for dose-finding trials in oncology. *arXiv preprint arXiv:2303.15798*, 2023.

Suyu Liu and Ying Yuan. Bayesian optimal interval designs for phase i clinical trials. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, pages 507–523, 2015.

Jiaying Lyu, Tianjian Zhou, Shijie Yuan, Wentian Guo, and Yuan Ji. Muce: Bayesian hierarchical modelling for the design and analysis of phase 1b multiple expansion cohort trials. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 72(3):649–669, 2023.

Beat Neuenschwander, Michael Branson, and Thomas Gsponer. Critical aspects of the bayesian approach to phase i cancer trials. *Statistics in medicine*, 27(13):2420–2439, 2008.

John O'Quigley, Margaret Pepe, and Lloyd Fisher. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics*, pages 33–48, 1990.

Haitao Pan, Fang Xie, Ping Liu, Jielai Xia, and Yuan Ji. A phase i/ii seamless dose escalation/expansion with adaptive randomization scheme (sears). *Clinical Trials*, 11(1):49–59, 2014.

Caroline Robert, Antoni Ribas, Jedd D. Wolchok, F. Stephen Hodi, Omid Hamid, Richard Kefford, Jeffrey S. Weber, Anthony M. Joshua, Wen-Jen Hwu, Tara C. Gangadhar, Amita Patnaik, Roxana Dronca, Hassane Zarour, Richard W. Joseph, Peter Boasberg, Bartosz Chmielowski, Christine Mateus, Michael A. Postow, Kevin Gergich, Jeroen Elassaiss-Schaap, Xiaoyun Nicole Li, Robert Iannone, Scot W. Ebbinghaus, S. Peter Kang, and Adil Daud. Pembrolizumab versus ipilimumab in advanced melanoma. *The New England Journal of Medicine*, 372(26):2521–2532, 2014. doi: 10.1056/NEJMoa1503093.

James G Scott and James O Berger. Bayes and empirical-bayes multiplicity adjustment in the variable-selection problem. *The Annals of Statistics*, pages 2587–2619, 2010.

M. Shah, A. Rahman, M. R. Theoret, and R. Pazdur. Optimizing dosing in oncology drug development q&a. Friends of Cancer Research, April 7 2022a.

Mirat Shah, Atiqur Rahman, Marc R. Theoret, and Richard Pazdur. How to get the dose right. *The ASCO Post*, May 10 2022b.

Xiao Su, Yisheng Li, Peter Müller, Chia-Wei Hsu, Haitao Pan, and Kim-Anh Do. A semi-mechanistic dose-finding design in oncology using pharmacokinetic/pharmacodynamic modeling. *Pharmaceutical statistics*, 21(6):1149–1166, 2022.

P. F. Thall and J. D. Cook. Dose-finding based on efficacy–toxicity trade-offs. *Biometrics*, 60(3):684–693, 2004.

Moreno Ursino, Sarah Zohar, Frederike Lentz, Corinne Alberti, Tim Friede, Nigel Stallard, and Emmanuelle Comets. Dose-finding methods for phase I clinical trials using pharmacokinetics in small populations. *Biometrical Journal*, 59(4):804–825, 2017.

J. Craig Venter et al. The sequence of the human genome. *Science*, 291(5507):1304–1351, 2001. doi: 10.1126/science.1058040.

Y. Zhou, J. J. Lee, and Y. Yuan. A utility-based bayesian optimal interval (u-boin) phase i/ii design to identify the optimal biological dose for targeted and immune therapies. *Statistics in Medicine*, 38(28):S5299–S5316, 2019.

J. Fourie Zirkelbach, M. Shah, J. Vallejo, J. Cheng, A. Ayyoub, J. Liu, R. Hudson, R. Sridhara, G. Ison, L. Amiri-Kordestani, S. Tang, T. Gwise, A. Rahman, R. Pazdur, and M. R. Theoret. Improving dose-optimization processes used in oncology drug development to minimize toxicity and maximize benefit to patients. *J Clin Oncol*, page JCO2200371, September 12 2022.

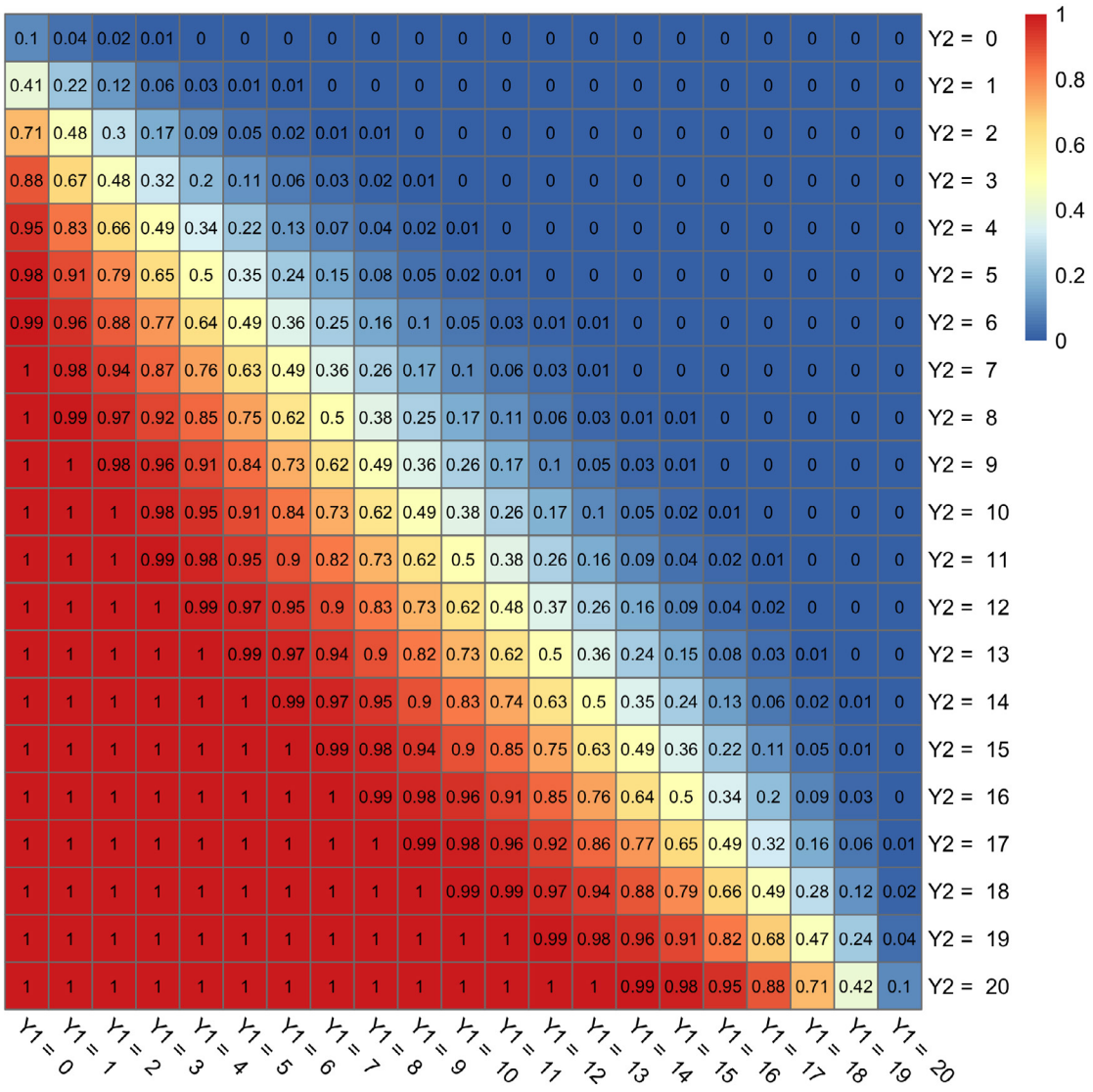


Figure 4: Posterior probabilities of the higher dose being more efficacious than the lower dose in a two-dose randomized comparison based on 20 patients per dose. Here $Y1$ and $Y2$ represent the number of responders in the low-dose and high-dose groups, respectively. For each pair of $(Y1, Y2)$, the posterior probability is calculated under a simple beta-binomial model.

DESIGN STRATEGIES FOR DOSE OPTIMIZATION IN ONCOLOGY DRUG DEVELOPMENT

Bo Huang[1] and Ying Yuan[2]

[1]Pfizer Inc., New York, NY, United States of America; Email address: bo.huang@pfizer.com [2]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.

1. Introduction

Cancer treatment is a complex process that requires careful consideration of various factors, including the stage of cancer, patient characteristics, and treatment options. One critical aspect of cancer treatment is determining the appropriate dosage of drugs to administer. The goal of dose optimization is to find the right balance between maximizing the effectiveness of the treatment and minimizing potential adverse side effects.

In 2021, the Oncology Center of Excellence launched Project Optimus to reform the dose optimization and dose selection paradigm in oncology drug development [1, 2, 3]. This multi-disciplinary initiative brings together medical oncologists, clinical pharmacologists, pharmacologists-toxicologists, and statisticians with the primary objective of addressing the challenges associated with dose optimization in oncology clinical trials. Some of the key considerations include utilizing pharmacodynamic (PD) biomarkers to inform dose optimization, taking into account factors beyond dose limiting toxicities (DLTs), such as tolerability issues like dose modification/interruption/ discontinuation and low-grade but persistent toxicities. Additionally, exploration of multiple doses after Phase I in a drug development program is being undertaken as part of this effort [2].

Drug development is a long, extremely expensive process. Dose optimization strategies that minimize the impact on both timeline and budget are of great importance and interest to drug developers. In this article, we will review and discuss various strategies for optimizing dose in oncology to facilitate the reform of the dose optimization and dose selection paradigm.

2. Dose Optimization in an Efficacy-integrated Dose Escalation Study

Over the past two decades, numerous efficacy-integrated dose escalation designs have been proposed to determine the optimal biological dose (OBD) [4-9]. In the statistical design literature, they are often known as phase 1/2 designs to highlight their feature of considering both toxicity (phase 1 endpoint such as DLT) and efficacy (phase 2 endpoint such as response or biomarker/pharmacokinetic [PK] endpoints) to prospectively identify the safe and effective dose and make the decision of dose escalation and de-escalation [10]. The primary advantage of these designs is that they continuously update the estimate of the most desirable dose, in light of the most recent data, to determine the dose assignment for the next cohort of patients, thereby tending to treat more patients at the optimal dose. The timeline aligns well with the current drug development paradigm, causing minimal disruption. Such designs are most suitable for sBLA/sNDA when new indications are investigated as a monotherapy or as novel combinations after initial market authorization where sufficient safety and efficacy evidence has been gained about the drug.

In clinical practice, however, the use of these designs is limited due to various factors, including longer evaluation time of efficacy posing challenges for real-time decision making, heterogeneity and "all-comer" nature of phase 1 dose escalation trials, low response rate impacting data availability for model fitting, complexity of statistical methods posing challenges for model fitting and dose selection decision, and potential confounding due to the lack of randomization.

Methods have been developed to address these challenges. For example, the recently proposed TITE-BOIN12 design [11] is an extension of the BOIN12 design [12] by incorporating a time-to-event method for bivariate outcomes. This design addresses some of the aforementioned limitations with a simpler design that utilizes partial data to reduce the time between cohorts. Another strategy is to prospectively identify biomarkers for dose escalation in the protocol and analyze the data retrospectively with backfill cohorts, as elaborated in Section 4.

This type of design serves as an initial step to establish a recommended phase 2 dose set (RP2S) for dose optimization, increase efficiency by continuously updating the estimate of the optimal dose and assigning more patients to potentially optimal dose(s), and offer preliminary data to support further dose optimization in subsequent stages, e.g., via randomization.

3. Randomized Dose Optimization in a Seamless Phase 1/2 Study

Randomized dose-ranging studies are routinely used, and often required by the FDA in the development of non-oncology drugs. In dose-ranging studies, patients are randomized to several dose groups for proof-of-concept and assessment of the dose-response relationship. However, such a design is rare in oncology due to a variety of reasons such as sample size restriction, pressing timeline, and higher tolerance for adverse drug reactions for maximizing efficacy.

According to the FDA's draft guidance [3], it is recommended to conduct a randomized multiple-dose phase 2 trial after the completion of phase 1 dose escalation and identification of the MTD. The guidance does not require that the randomized dose-ranging study be statistically powered in a rigorous way to demonstrate that one dose is superior to the other. Practically, one might treat each cohort as a single-arm trial with toxicity and futility monitoring rules (e.g., Simon two-stage design [13] or Bayesian optimal phase II (BOP2) design [14] to jointly monitor toxicity and efficacy) to make interim decisions (i.e., pick-the-winner or drop-the-loser), and generate adequate safety, efficacy, PK/PD data to support the exposure-response analysis and better inform the dose selection.

Some novel statistical designs have been proposed to improve the accuracy and efficiency of dose optimization using a randomized dose-ranging design. Guo and Yuan [15] proposed the DROID design, a dose-ranging approach for optimizing targeted oncology drug doses, bridging dose-ranging studies with oncology dose-finding designs. DROID has two stages: first, patients are adaptively assigned to different doses to establish the therapeutic dose range and the recommended phase 2 dose set (RP2S); second, patients are randomized to each dose in the RP2S to assess dose-response and identify the optimal dose. Yang et al. [16] proposed the MERIT method, which is a two-stage design that aims to address two critical questions coming from the FDA draft guidance on dose optimization in oncology: (a) how to determine the sample size, and (b) how to design a randomized multi-dose optimization trial in oncology. The design is formalized to determine an optimal admissible set that satisfies certain statistical properties, including Type I error and power. The randomized dose-ranging design has the advantage of minimizing bias by achieving balance in dose cohorts, but often requires a larger sample size to achieve reasonable power to identify the optimal dose from candidate doses, ranging from 20-50 patients per arm [16].

4. Sequential Dose Optimization after Early Sign of Efficacy in a Seamless Phase 1b/2 Study

It is well known that cancer drug development is increasing costly, and most investigational drugs fail before entering the confirmatory phase 3 stage. Given the high uncertainty and attrition rate, committing to a randomized study for dose optimization prior to establishing early sign of efficacy (eSOE) is often challenging in a budget-constrained corporate environment. A pragmatic sequential approach for dose optimization after eSOE at a 'no-regret' single dose (e.g. MTD or highest administered dose) expansion cohort enables that adequate resources can be allocated to more pipeline products and can be a useful strategy in oncology drug development.

This strategy can be a non-randomized dose-ranging design such as adding a backfill cohort based on emerging data and PK modelling following eSOE. However, the lack of randomization can lead to bias (e.g., time bias and population bias) and increase the risk of uninterpretable data for dose selection. This strategy can also be a staggered randomization design at 2 or 3 doses following eSOE for more rigorous decision making, despite at the expense of further prolonging the development timeline. This strategy is most appropriate when either confidence in the investigational product or strategic value is of high uncertainty and there is a need to establish eSOE prior to investment in dose optimization.

5. Simultaneous Dose Optimization in a Seamless Phase 2/3 Study

Seamless phase 2/3 designs in clinical trials offer several advantages over traditional sequential designs. By combining the two phases, researchers can streamline the drug development process without the need for separate protocols, saving time and money with the same clinical trial infrastructure and accrual sites while also potentially reducing the number of patients required for the trial (i.e., inferential seamless phase 2/3 design). Given the timeline concern with the incorporation of dose optimization, it is appealing to consider simultaneous dose optimization in a seamless phase 2/3 study. However, the enrolment pauses between Phase 2 and Phase 3 to collect the data required for dose optimization will cause operational challenges and may discourage site participation.

Operational Seamless Design

An operational seamless Phase 2/3 design is conceptually simple, consolidating the randomized dose-ranging and pivotal components into one protocol. The primary goal of this design is to minimize the gap between Phase 2 and Phase 3, particularly if the sponsor and the health regulators can agree on the dose optimization strategy and Phase 3 design during a Type-B meeting. Doing so reduces the enrolment pause before initiating the Phase 3 part. Additionally, conducting a well-designed, adequately sized, randomized dose-ranging Phase 2 trial can provide robust dose optimization decisions. By utilizing one protocol, operational efficiency can be achieved, allowing for synchronized database setup and site initiation, ultimately leading to faster enrolment in Phase 3.

Inferential Seamless Design

Compared to an operationally seamless design, an inferential seamless design combines the data from the Phase 2 component with the data from the Phase 3 component in the primary efficacy analysis with proper multiplicity adjustment, thereby is more statistically efficient. Jiang and Yuan [17] discuss different types of phase 2/3 designs and show that the sample size saving by using inferential seamless design can be 20-30%. However, unlike an operational seamless design, the inferential seamless design requires the same control arm in both the Phase 2 and Phase 3 parts, and the patient population should be identical. The Phase 2 component serves as Stage 1 of the pivotal Phase 3 trial for dose selection and early futility analysis, with typically one dose moving to Phase 3.

Because of the dose selection, standard methods based on pooled phase 2/3 data lead to inflated Type I error rate. A closed testing procedure can be used to control the familywise Type I error rate, which rejects a hypothesis H_i at level α only when all possible intersection hypotheses involving H_i are rejected by using valid local level α tests. For the interaction test of null hypotheses, one can use the Simes test [18] for improved efficiency since it is reasonable to assume that the p-values for testing H_i of different doses are non-negatively correlated [19]. To combine the data from Phase 2 and Phase 3, a combination test can be used. Two commonly used combination tests are Fisher combination test [20] and inverse normal combination test [21]. Friede and Stallard [23] provided a review of methods to control familywise Type I error for phase 2/3 trials.

Although an inferential seamless design boasts statistical efficiency, from an operational standpoint, the Phase 2 should be treated the same as the Phase 3. This requires the sponsor to appoint an independent steering committee to work alongside an external data monitoring committee (DMC) to manage the dose selection, with a secure firewall established between the committee and the study team. Furthermore, dose selection decisions may require more diverse data, including efficacy data, adverse events, pharmacokinetics, patient-reported outcomes, and so on, and may necessitate additional decision rules. Therefore, the decision of whether the gained efficiency justifies the increased complexity and operational restrictions should be made case-by-case.

6. Dose Optimization in a Randomized Phase 3 Study

Another option is multi-arm randomized phase 3 design, testing two or more dose levels versus a control arm in the pivotal trial. One can make the most statistically rigorous decisions with this option because of the larger sample size in a randomized setting. Enrolment will also be faster as Phase 3 studies are usually global trials with hundreds of investigational sites. Furthermore, the sponsor will be able to assess the long-term survival outcome when comparing multiple dose levels. However, statistical penalty on Type I error rate adjustment can make the study prohibitively large with higher cost.

A variate of this design is to have a run-in cohort (e.g., 30-40 patients per arm) to quickly assess safety, PK and early efficacy data to make a decision on dose selection and proceed with one dose afterward. It is similar but different from a seamless Phase 2/3 design as dose optimization is achieved in a Phase 3 trial. To avoid multiplicity adjustment due to bias arising from the dose selection, the final analysis should be performed using all data from the investigational drug (including the dropped dose arm). Such a design requires a smaller sample size for the pivotal trial and is faster and more efficient. However, there will be operational challenges for quick decision making and the follow-up time for patients may not be sufficient to make a more informed decision.

7. Timeline Impact of Various Design Strategies

Several factors affect the timeline of drug development in clinical trials for dose optimization. These factors include Chemistry, Manufacturing, and Control (CMC) as the availability of commercial formulation and its analytic/clinical validation can impact the start date of a pivotal trial. Regulatory interactions and reviews of the sponsor's dose optimization plan and data from earlier studies can also cause uncertainty in the development timeline.

Figure 1 provides a visual overview of the general timelines for each design strategy and the projected timing of the dose optimization decision.

In general, the impact of dose optimization on the development timeline can be reduced by implementing it earlier in the clinical development plan and by discussing the dose optimization strategy with regulators. Figure 1 shows the timeline for each design strategy, demonstrating that the earlier dose optimization is initiated, the less impact it has on the development timeline.

4. Discussions

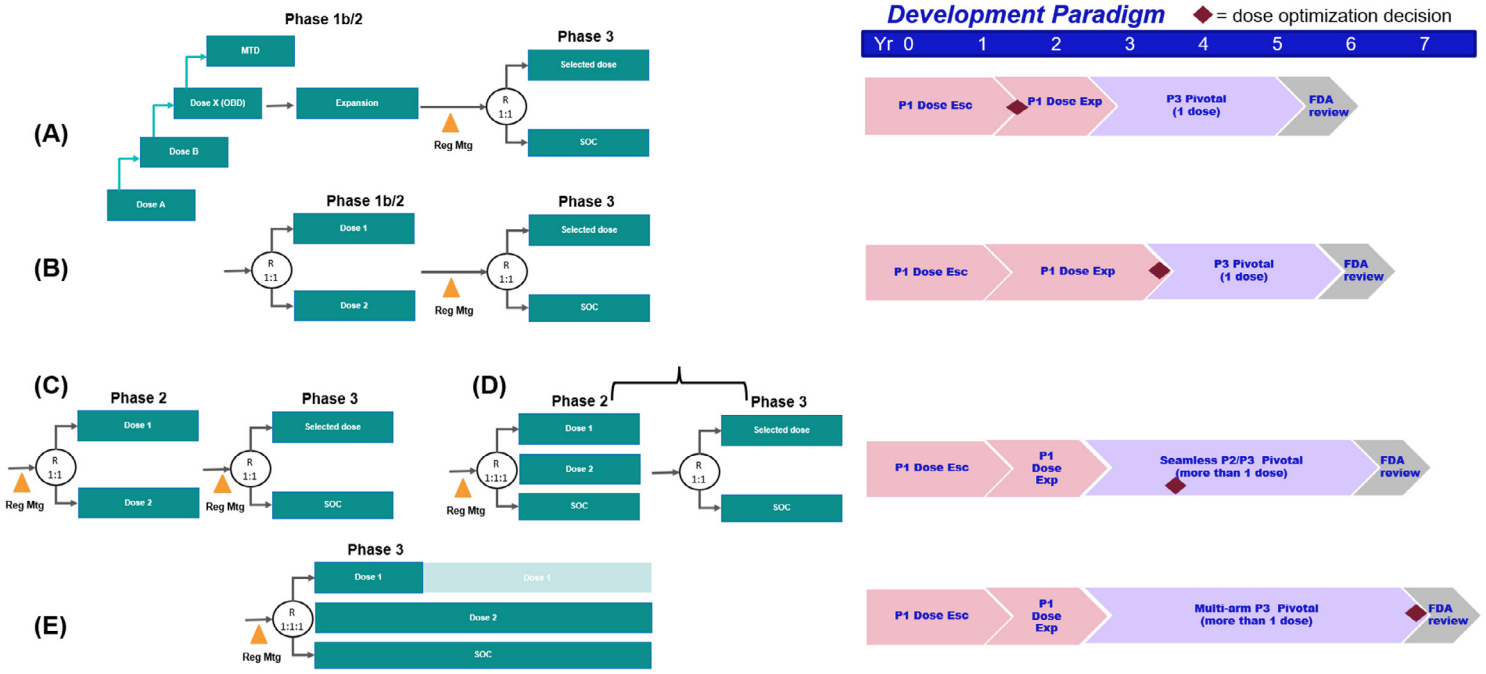
Dose optimization for novel oncology agents poses new challenges. A key challenge in dose optimization is weighing the additional sample size and potentially longer time to develop an optimized dosage versus the goal of minimizing the time to deliver an efficacious but possibly suboptimal dosage to patients [1].

The most effective and efficient strategy should include dose optimization in a seamless manner in the current drug development paradigm. Initial safety dose escalation should take into account the mechanism of action of the investigational drug, possibly allowing a flexible definition of the MTD (e.g., targeting 20% or 15% DLT rate) or revising the DLT definition by including clinically relevant grade 2 adverse events [23].

It is also important to emphasize that although statistical dose optimization designs are based on certain efficacy and safety endpoints to make the problem trackable and the design executable, the final decision of dose selection should take into consideration the totality of evidence, including safety, tolerability, short-term and potentially long-term clinical efficacy outcomes, pharmacokinetics, biomarkers (e.g. receptor occupancy, kinase inhibition). Quantitative exposure-response modelling for safety and efficacy together with structured qualitative benefit-risk assessment are equally important.

Ideally, dose optimization should be incorporated in the early stage of clinical development and discussion with the health regulators should occur as soon as possible, as early as the pre-IND setting to reduce potential delays in approval. Early investigation would facilitate more efficient trial designs and reduce the failure rate caused by undue toxicity. ■

Figure 1. Various design strategies and estimated timeline for dose optimization.



References

[1] U.S. Food And Drug Administration. (2022) Project Optimus, https://www.fda.gov/about-fda/oncology-center-excellence/project-optimus

[2] Zirkelbach JF, Shah M, Vallejo J, Cheng J, Ayyoub A, Liu J, Hudson R, Sridhara R, Ison G, Amiri-Kordestani L, Tang S. Improving Dose-Optimization Processes Used in Oncology Drug Development to Minimize Toxicity and Maximize Benefit to Patients. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2022:JCO2200371.

[3] U.S. Food And Drug Administration. (2023) Optimizing the Dosage of Human Prescription Drugs and Biological Products for the Treatment of Oncologic Diseases. Rockville, MD: Food and Drug Administration.

[4] Thall, P. F. and Cook, J. D. (2004). Dose-finding based on efficacy-toxicity trade-offs. *Biometrics,* 60:684–693.

[5] Bekele, B. N. and Shen, Y. (2005). A Bayesian approach to jointly modeling toxicity and biomarker expression in a phase I/II dose- nding trial. *Biometrics*, 61:343{354.

[6] Yin, G., Li, Y. and Ji, Y. (2006). Bayesian dose finding in phase I/II clinical trials using toxicity and efficacy odds ratios. *Biometrics*, 62(3), pp.777-787.

[7] Liu, S. and Johnson, V. E. (2016). A robust Bayesian dose-finding design for phase I/II clinical trials. *Biostatistics*, 17:249–263.

[8] Liu, S., Guo, B., and Yuan, Y. (2018). A Bayesian phase I/II trial design for immunotherapy. *Journal of the American Statistical Association*, 113:1016–1027.

[9] Takeda, K., Taguri, M., and Morita, S. (2018). Boin-et: Bayesian optimal interval design for dose-finding based on both e cacy and toxicity outcomes. *Pharmaceutical statistics*. 17(4):383-395.

[10] Yuan, Y., Nguyen, H. Q., and Thall, P. F. (2017). Bayesian designs for phase I-II clinical trials. Chapman and Hall/CRC, Boca Raton.

[11] Zhou, Y., Lin, R., Lee, J. J., Li, D., Wang, L., Li, R., & Yuan, Y. (2022). TITE-BOIN12: A Bayesian phase I/II trial design to find the optimal biological dose with late-onset toxicity and efficacy. *Statistics in Medicine*, 41(11), 1918-1931.

[12] Lin R, Zhou Y, Yan F, Li D, Yuan Y. BOIN12: Bayesian optimal interval phase I/II trial design for utility-based dose finding in immunotherapy and targeted therapies. *JCO precision oncology*. 2020 Nov;4:1393-402.

[13] Simon R. Optimal two-stage designs for Phase II clinical trials. *Controlled Clinical Trials* 1989; 10:1–10.

[14] Zhou H, Lee JJ, Yuan Y. BOP2: Bayesian optimal design for phase II clinical trials with simple and complex endpoints. *Stat Med*. 2017 Sep 20;36(21):3302-3314.

[15] Guo B, Yuan Y. DROID: Dose-ranging Approach to Optimizing Dose in Oncology Drug Development. arXiv preprint arXiv:2209.01655. 2022 Sep 4.

[16] Yang P, Li D, Lin R, Huang B, Yuan Y. Design and Sample Size Determination for Multiple-dose Randomized Phase II Trials for Dose Optimization. arXiv preprint arXiv:2302.09612. 2023 Feb 19.

[17] Jiang L, Yuan Y. Seamless Phase II/III Design: A Useful Strategy to Reduce the Sample Size for Dose Optimization. J Natl Cancer Inst. 2023 May 27:djad103. doi: 10.1093/jnci/djad103. Epub ahead of print. PMID: 37243720.

[18] Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3), 751-754.

[19] Sarkar, S. K., & Chang, C. K. (1997). The Simes method for multiple hypothesis testing with positively dependent test statistics. *Journal of the American Statistical Association*, 92(440), 1601-1608.

[20] Fisher RA. Statistical methods for research workers. Springer New York; 1992.

[21] Folks JL, Chhikara RS. The inverse Gaussian distribution and its statistical application—a review. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1978 Jul;40(3):263-75.

[22] Friede T, Stallard N: A comparison of methods for adaptive treatment selection. *Biom J* 50(5): 767-781, 2008.

[23] Huang B. Some statistical considerations in the clinical development of cancer immunotherapies. *Pharmaceutical Statistics*. 2018 Feb;17(1):49-60.

DIVERSITY IN CLINICAL TRIALS

Julie Shah, Xiang Tang, and Nancy Bauer, *Biostatistics and Data Science, Boehringer Ingelheim Pharmaceuticals, Inc.*

Clinical trials are a critical component of the drug development process, providing evidence of safety and efficacy for new treatments. However, historically, clinical trials have not always been representative of the diverse populations they aim to serve. This lack of diversity can have significant implications for the generalizability of trial results and the ability to identify potential differences in treatment response across different patient populations. In recent years there has been a growing recognition of the importance of diversity in clinical trials, and efforts are being made to improve the representation of underrepresented groups not only by industry but also by academia and regulatory agencies. Two important documents have been published by regulatory agencies on this topic. The Food and Drug Administration (FDA) published the draft guidance in April of 2022 entitled '*Diversity Plans to Improve Enrollment of Participants from Underrepresented Racial and Ethnic Populations in Clinical Trials*', which aims to increase the representation of different patient populations in clinical trials [1]. The guidance emphasizes the importance of including diverse patient populations in clinical trials to ensure that the results are generalizable and applicable to all patients. In January 2018, the European Medicines Agency (EMA) published '*Reflection paper on physical frailty: instruments for baseline characterisation of older populations in clinical trials*', a paper focused on the gender and age allocation of subjects [2].

Why Diversity Matters in Clinical Trials

Using a diverse population in a clinical trial, at least one that is aligned with the prevalence of the study indication, is important for several reasons. First, it ensures that the results of clinical trials are applicable to the broader population. If a trial only includes a narrow subset of the population, the results may not be generalizable to other groups. This can lead to disparities in healthcare, with certain groups being excluded from the benefits or potential negative effects of new treatments.

Second, diversity in clinical trials can help identify potential differences in treatment response across different patient populations. For example, certain medications may be more effective in certain racial or ethnic groups, or in men versus women, or young versus old. Without adequate representation of these groups in clinical trials, these differences may go unnoticed.

Third, diversity in clinical trials is important from an ethical standpoint. All patients should have the opportunity to participate in clinical trials and benefit from new treatments, regardless of their race, ethnicity, sex, or other factors.

Overall, increasing diversity in clinical trials requires a collaborative effort from a variety of stakeholders, including researchers, healthcare providers, patient advocacy groups, community organizations and patients and their families.

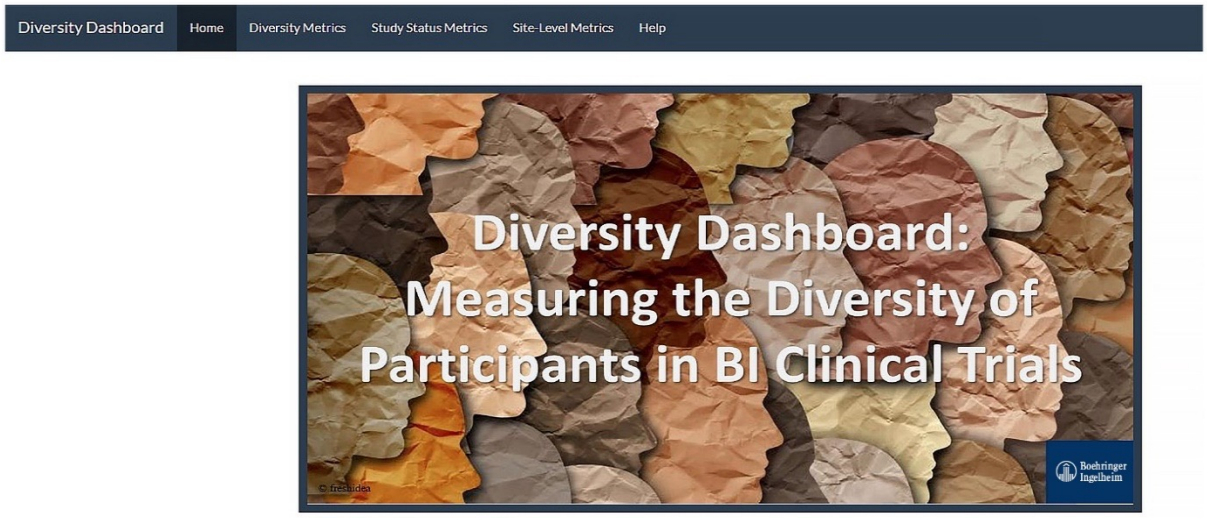
Strategies to Improve Diversity in Clinical Trials

Strategies can be employed to improve diversity in clinical trials such as: increasing education and outreach, using culturally sensitive recruitment strategies, providing language services, ensuring accessible trial sites, engaging with community organizations and leaders, and providing incentives for participation.

Enrolling and maintaining a diverse population in a clinical trial can be hindered by the lack of daily tracking of subject diversity metrics during trial conduct. This can lead to difficulties in adjusting the screened population dynamics based on "live" data and mitigating early discontinuation causes of the study, particularly for underrepresented populations. The Diversity Dashboard application was created to aid in the enrolling and maintaining a diverse patient population in the clinical trial.

Diversity Dashboard

The Diversity Dashboard was created for the Boehringer Ingelheim (BI) local Clinical Operations staff to monitor diversity metrics proactively and independently in clinical trials to aid in reaching diversity trial targets. This includes overseeing diversity metrics by various categories such as therapeutic areas (TAs), indications, drug substance, trials, investigator sites, and primary investigators. Additionally, the dashboard is designed to track patient recruitment and retention diversity by trial and site. Historical data can be mined to identify sites that have successfully recruited diverse populations for potential reuse for a given TA, indication or drug substance.



Methodology

The Diversity Dashboard application gathers various information at different time points and processes the data into the application. Subject data is collected from the sites every weekday from information gathered the day before. Other details such as investigator and site information are refreshed on weekly basis. Prevalence rates for a given indication are provided by the BI Real-World Data & Analytics group to the Clinical Operations staff and used to help define the diversity study targets for the specific trial. Based on user selections made in the Diversity Dashboard, prevalence or study targets data may not be appropriate to display and thus the current USA census is applied to the output. The Diversity Dashboard focuses on key diversity factors including sex, race, ethnicity and age groups. Recent factors added to the application are childbearing potential and gender identity.

Functionality

The Diversity Dashboard application contains three key features:

1. Diversity Metrics: Summarizes the diversity factors and compares it to benchmark data (e.g., USA census data, prevalence data, or study target data)
2. Study Status Metrics: Summarizes the disposition of subjects by diversity factors
3. Site-Level Metrics: Summarizes diversity factor data by USA sites and investigators

Dashboard Filters

The Diversity Dashboard application defined standard user selections across the application.

- The first selection criteria 1 includes specific levels of data to summarize: Therapeutic Areas, Indications, Substance Names, Project Codes, Study Numbers, USA Primary Centre Names or USA Principal Investigators.
    - Its associated sub-criteria selection 2 drills down to the details.
- The third selection 3 is based on the patient population of interest: Screened/Enrolled, Not Randomized/Not Entered or Randomized/Entered.
- The next criteria 4 is based on Study Status: Ongoing Studies and or Ended Studies.
- The Group by selection 5 summaries the data by: USA vs the Rest of the World or Year of Enrollment.
- The last two selections are Subject Characteristic criteria 6 and its sub-criteria 7: Sex, Race, Ethnicity, Age Group, Childbearing Potential and Gender Identity.

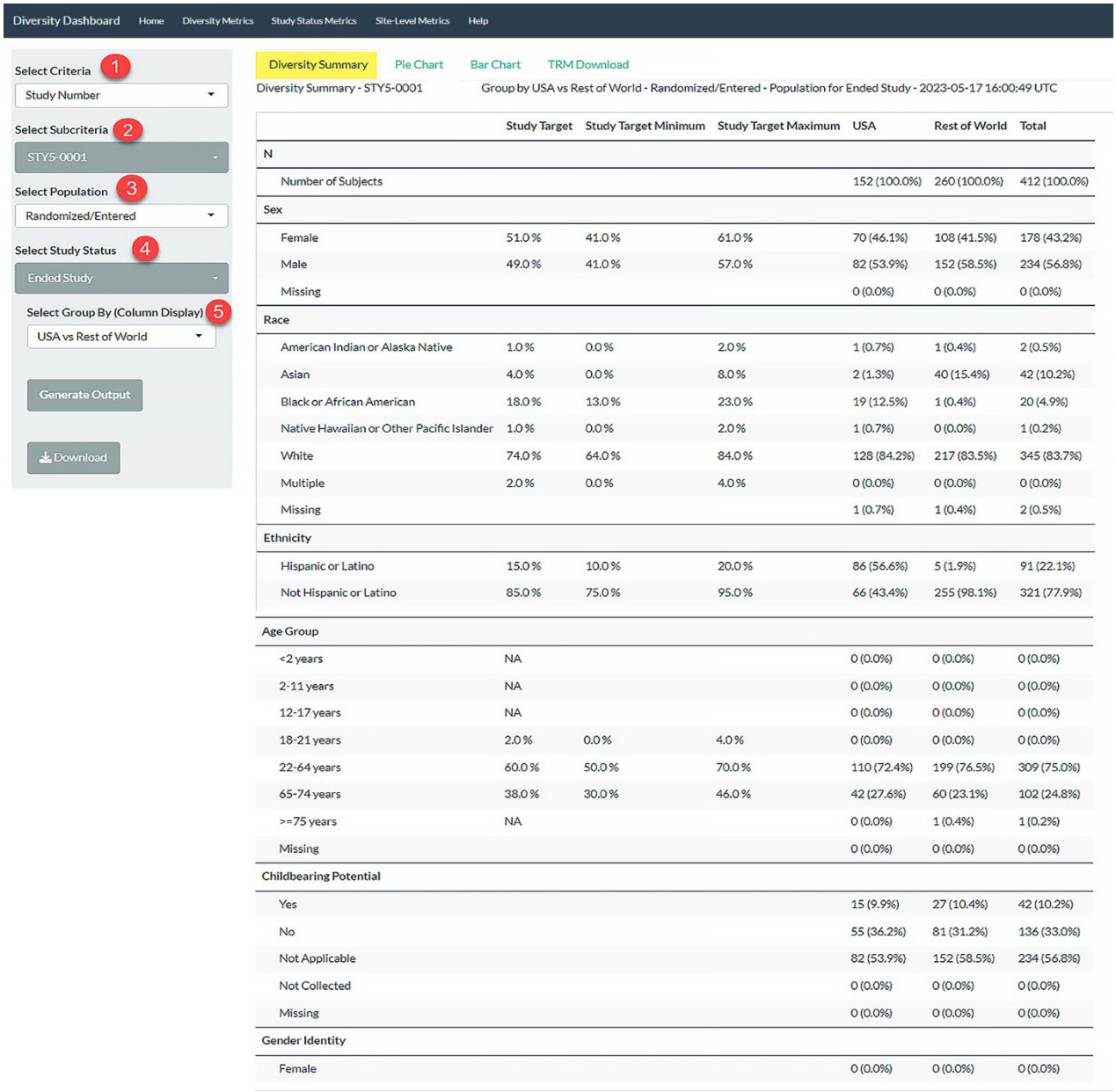
Three examples showing the use of the Diversity Dashboard are described below.

The **Diversity Metrics** tab contains three key features: Diversity Summary, Pie Chart and Bar Chart.

In the first example (Table 1), a single study was selected for an ended study, along with the 'Randomized/Entered' population and to group the output USA data verses the Rest of the World. Take note of the first diversity factor 'Sex' as listed on the output. As this study has already ended, the study under-randomized the Study Target (51%) for females and over-randomized the Study Target for males (49%). In both cases, the Study Target Minimum for females and the Study Target Maximum for males were met. Similarly, the Study Targets for the Age Groups, 18-21 and 65-74 were also not met.

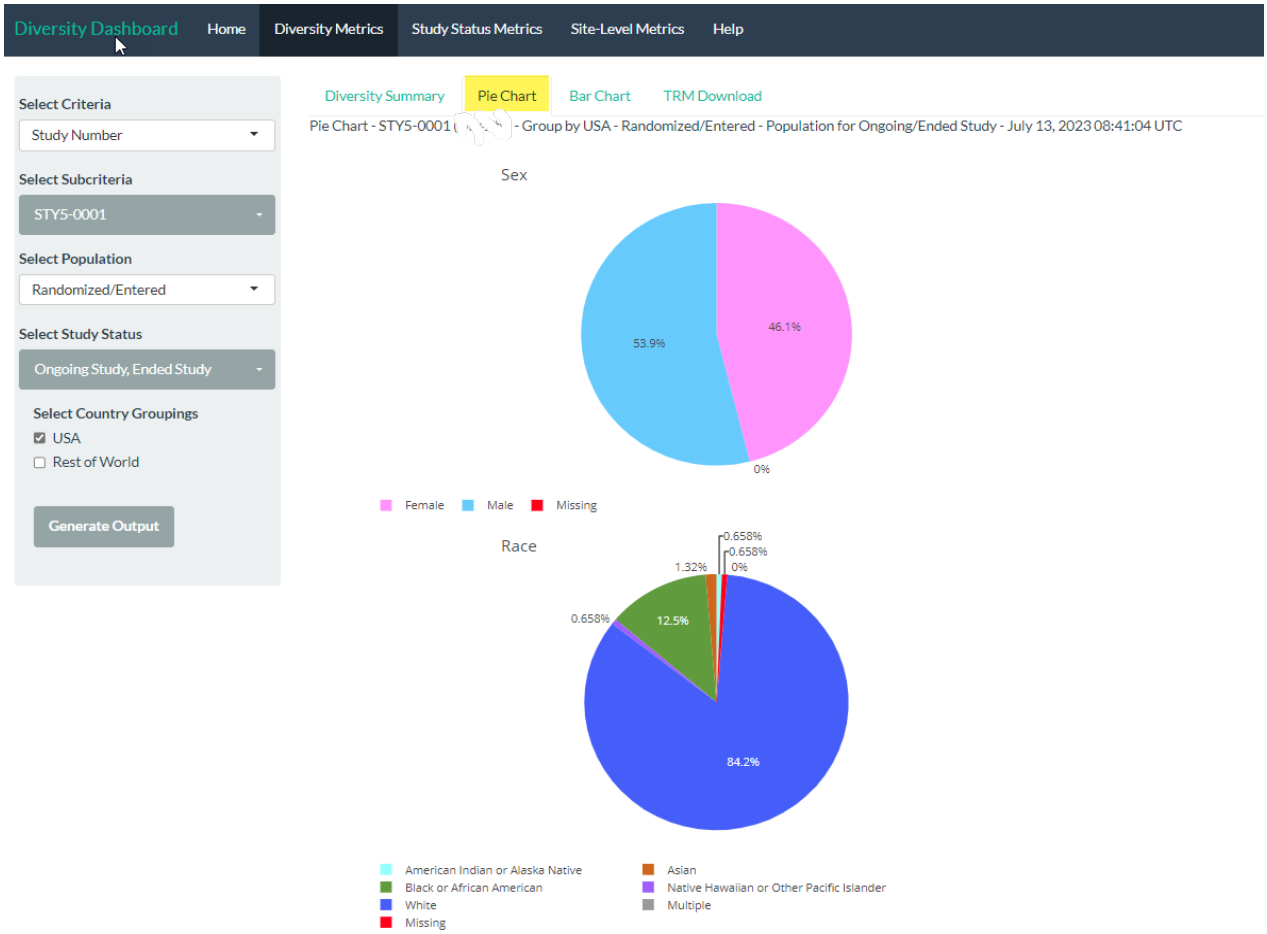
For a visual representation of the first example, see the Pie Chart (Figure 1) with the same criteria selected.

Table 1. Diversity Summary of the USA Randomized Population for Study STY5-0001



	Study Target	Study Target Minimum	Study Target Maximum	USA	Rest of World	Total
N						
Number of Subjects				152 (100.0%)	260 (100.0%)	412 (100.0%)
Sex						
Female	51.0 %	41.0 %	61.0 %	70 (46.1%)	108 (41.5%)	178 (43.2%)
Male	49.0 %	41.0 %	57.0 %	82 (53.9%)	152 (58.5%)	234 (56.8%)
Missing				0 (0.0%)	0 (0.0%)	0 (0.0%)
Race						
American Indian or Alaska Native	1.0 %	0.0 %	2.0 %	1 (0.7%)	1 (0.4%)	2 (0.5%)
Asian	4.0 %	0.0 %	8.0 %	2 (1.3%)	40 (15.4%)	42 (10.2%)
Black or African American	18.0 %	13.0 %	23.0 %	19 (12.5%)	1 (0.4%)	20 (4.9%)
Native Hawaiian or Other Pacific Islander	1.0 %	0.0 %	2.0 %	1 (0.7%)	0 (0.0%)	1 (0.2%)
White	74.0 %	64.0 %	84.0 %	128 (84.2%)	217 (83.5%)	345 (83.7%)
Multiple	2.0 %	0.0 %	4.0 %	0 (0.0%)	0 (0.0%)	0 (0.0%)
Missing				1 (0.7%)	1 (0.4%)	2 (0.5%)
Ethnicity						
Hispanic or Latino	15.0 %	10.0 %	20.0 %	86 (56.6%)	5 (1.9%)	91 (22.1%)
Not Hispanic or Latino	85.0 %	75.0 %	95.0 %	66 (43.4%)	255 (98.1%)	321 (77.9%)
Age Group						
<2 years	NA			0 (0.0%)	0 (0.0%)	0 (0.0%)
2-11 years	NA			0 (0.0%)	0 (0.0%)	0 (0.0%)
12-17 years	NA			0 (0.0%)	0 (0.0%)	0 (0.0%)
18-21 years	2.0 %	0.0 %	4.0 %	0 (0.0%)	0 (0.0%)	0 (0.0%)
22-64 years	60.0 %	50.0 %	70.0 %	110 (72.4%)	199 (76.5%)	309 (75.0%)
65-74 years	38.0 %	30.0 %	46.0 %	42 (27.6%)	60 (23.1%)	102 (24.8%)
>=75 years	NA			0 (0.0%)	1 (0.4%)	1 (0.2%)
Missing				0 (0.0%)	0 (0.0%)	0 (0.0%)
Childbearing Potential						
Yes				15 (9.9%)	27 (10.4%)	42 (10.2%)
No				55 (36.2%)	81 (31.2%)	136 (33.0%)
Not Applicable				82 (53.9%)	152 (58.5%)	234 (56.8%)
Not Collected				0 (0.0%)	0 (0.0%)	0 (0.0%)
Missing				0 (0.0%)	0 (0.0%)	0 (0.0%)
Gender Identity						
Female				0 (0.0%)	0 (0.0%)	0 (0.0%)

Figure1. Pie Chart of the USA Randomized Population for Study STY5-0001

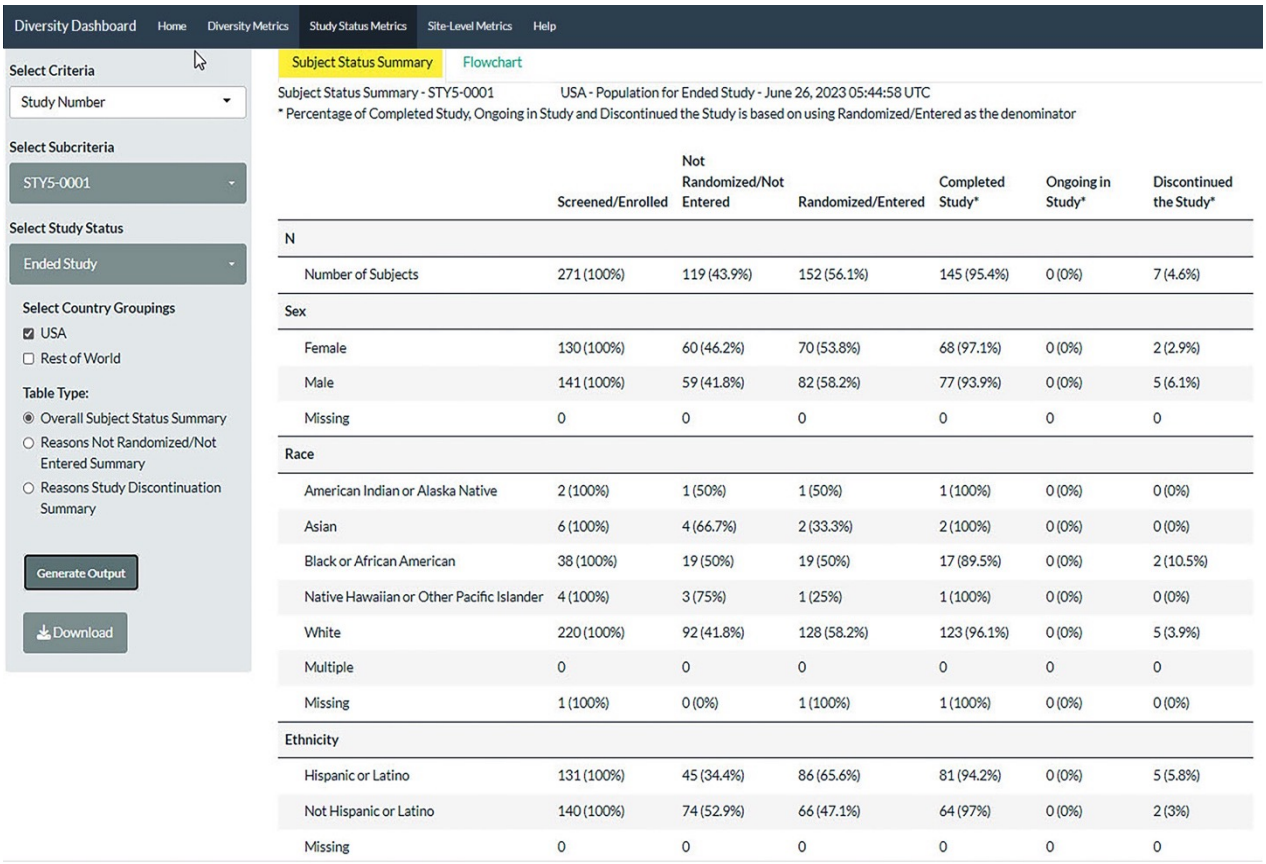


The Study Status Metrics contains two key features: Subject Status Summary information and the Flowchart. The Subject Status Summary has three table types: Overall Subject Status Summary, Reasons Not Randomized/Not Entered Summary and Reasons Study Discontinuation Summary. The Flowchart tab displays a disposition based on the specific diversity factors.

In the second example, Table 2 displays an overview of the subject status summary (partial display shown). Each diversity factor is tracked by various milestone groupings that occur during the trial: 'Screen/Enrolled', 'Not Randomized/Not Entered', 'Randomized/Entered', 'Completed Study', 'Ongoing in Study' and 'Discontinued the Study'. The 'Completed Study', 'Ongoing in Study' and the 'Discontinued the Study' group percentages are based on the number of subjects who were 'Randomized/Entered'. Observing a 10.5% loss of 'Blacks or African American' patients who discontinued the study may need additional discussions regarding the reason for this within the Clinical Operations team.

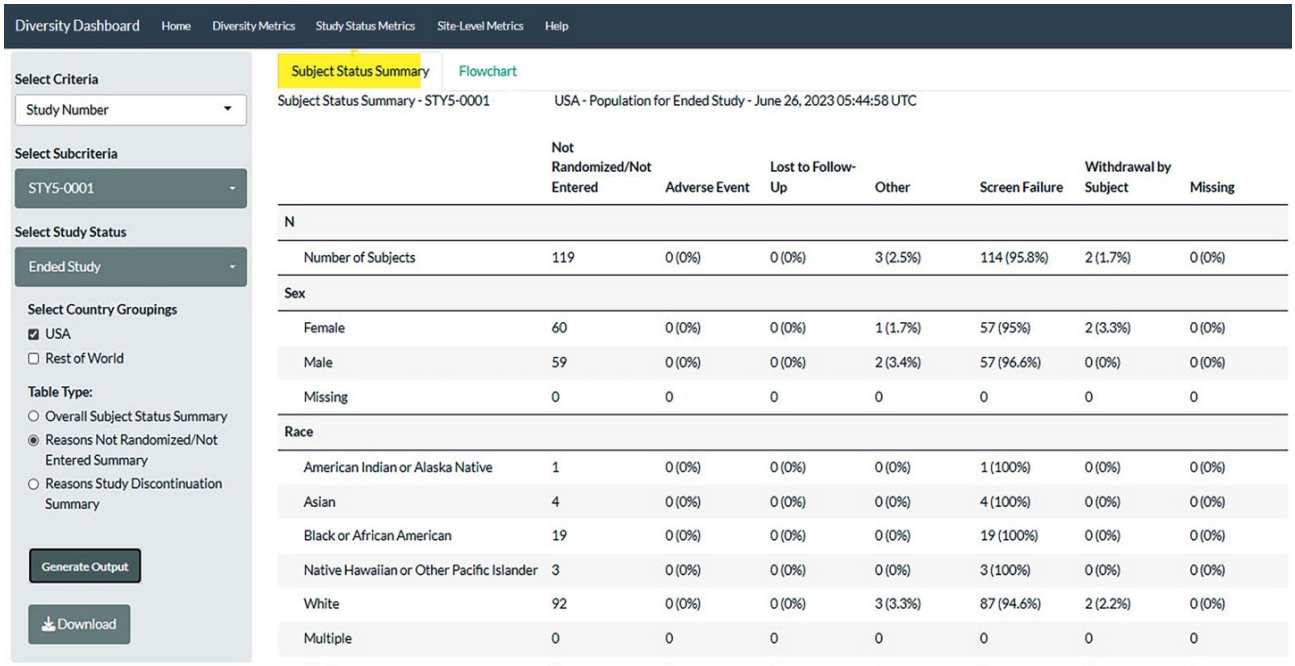
Continuing example 2, Table 3 displays the reasons USA subjects were not being randomized into the study STY5-001. This information would highlight to the Clinical Operation staff to discuss recruiting strategies and adjust the process proactively. In this case, 95.8% of the subjects who did not enroll in the study were due to screen failure reasons. This may lead to a thorough review of the specific screen failure reasons, identify the causes and possibly amend the protocol. For details on the specific screen failure reasons, the technical staff supporting the trial would be requested to provide the details.

Table 2. Subject Status Summary: Overall Subject Status Summary Table Type



	Screened/Enrolled	Not Randomized/Not Entered	Randomized/Entered	Completed Study*	Ongoing in Study*	Discontinued the Study*
N						
Number of Subjects	271 (100%)	119 (43.9%)	152 (56.1%)	145 (95.4%)	0 (0%)	7 (4.6%)
Sex						
Female	130 (100%)	60 (46.2%)	70 (53.8%)	68 (97.1%)	0 (0%)	2 (2.9%)
Male	141 (100%)	59 (41.8%)	82 (58.2%)	77 (93.9%)	0 (0%)	5 (6.1%)
Missing	0	0	0	0	0	0
Race						
American Indian or Alaska Native	2 (100%)	1 (50%)	1 (50%)	1 (100%)	0 (0%)	0 (0%)
Asian	6 (100%)	4 (66.7%)	2 (33.3%)	2 (100%)	0 (0%)	0 (0%)
Black or African American	38 (100%)	19 (50%)	19 (50%)	17 (89.5%)	0 (0%)	2 (10.5%)
Native Hawaiian or Other Pacific Islander	4 (100%)	3 (75%)	1 (25%)	1 (100%)	0 (0%)	0 (0%)
White	220 (100%)	92 (41.8%)	128 (58.2%)	123 (96.1%)	0 (0%)	5 (3.9%)
Multiple	0	0	0	0	0	0
Missing	1 (100%)	0 (0%)	1 (100%)	1 (100%)	0 (0%)	0 (0%)
Ethnicity						
Hispanic or Latino	131 (100%)	45 (34.4%)	86 (65.6%)	81 (94.2%)	0 (0%)	5 (5.8%)
Not Hispanic or Latino	140 (100%)	74 (52.9%)	66 (47.1%)	64 (97%)	0 (0%)	2 (3%)
Missing	0	0	0	0	0	0

Table 3. Subject Status Summary: Reasons Not Randomized/Not Entered Summary Table Type



	Not Randomized/Not Entered	Adverse Event	Lost to Follow-Up	Other	Screen Failure	Withdrawal by Subject	Missing
N							
Number of Subjects	119	0 (0%)	0 (0%)	3 (2.5%)	114 (95.8%)	2 (1.7%)	0 (0%)
Sex							
Female	60	0 (0%)	0 (0%)	1 (1.7%)	57 (95%)	2 (3.3%)	0 (0%)
Male	59	0 (0%)	0 (0%)	2 (3.4%)	57 (96.6%)	0 (0%)	0 (0%)
Missing	0	0	0	0	0	0	0
Race							
American Indian or Alaska Native	1	0 (0%)	0 (0%)	0 (0%)	1 (100%)	0 (0%)	0 (0%)
Asian	4	0 (0%)	0 (0%)	0 (0%)	4 (100%)	0 (0%)	0 (0%)
Black or African American	19	0 (0%)	0 (0%)	0 (0%)	19 (100%)	0 (0%)	0 (0%)
Native Hawaiian or Other Pacific Islander	3	0 (0%)	0 (0%)	0 (0%)	3 (100%)	0 (0%)	0 (0%)
White	92	0 (0%)	0 (0%)	3 (3.3%)	87 (94.6%)	2 (2.2%)	0 (0%)
Multiple	0	0	0	0	0	0	0
Missing	0	0	0	0	0	0	0
Ethnicity							
Hispanic or Latino	45	0 (0%)	0 (0%)	1 (2.2%)	43 (95.6%)	1 (2.2%)	0 (0%)
Not Hispanic or Latino	74	0 (0%)	0 (0%)	2 (2.7%)	71 (95.9%)	1 (1.4%)	0 (0%)
Missing	0	0	0	0	0	0	0

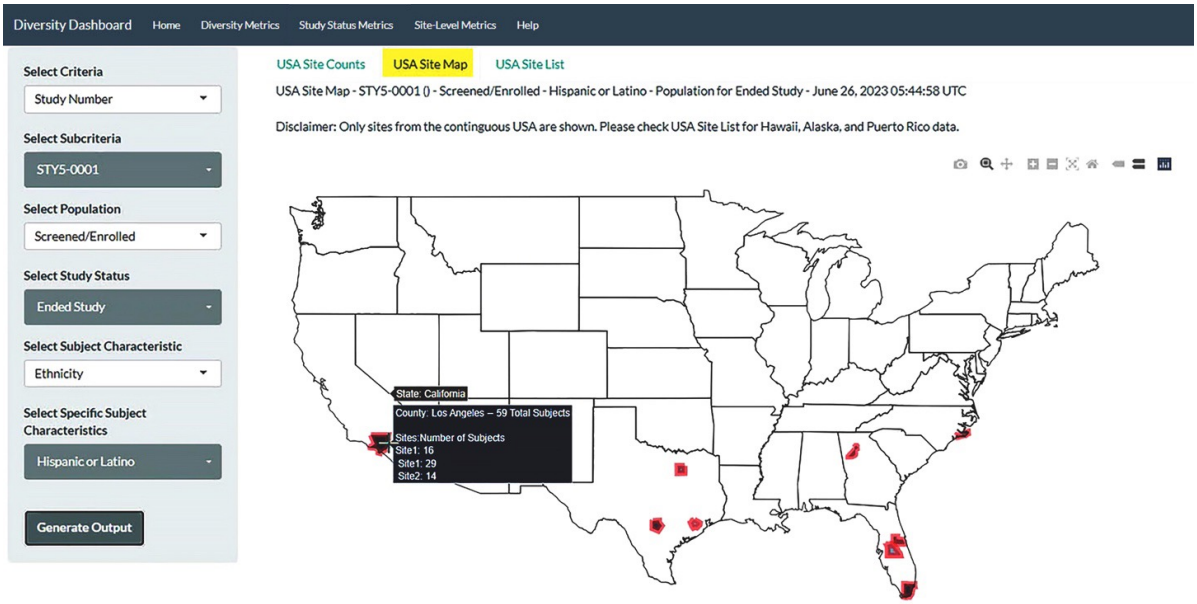
To dive deeper into a specific population, the Flowchart feature (Figure 2) provides a current status of the disposition of a desired population for both the USA and the Rest of the World. In this example, the Hispanic or Latino population is selected for Study STY5-0001. Of the 141 Hispanic or Latino population that were screened, 50 subjects did not make it into the study due to screen failures, subject withdrawal or other reasons, leaving only 91 Hispanic or Latino subjects entering the trial worldwide.

The third feature of the Diversity Dashboard application is the Site-Level Metrics tab.

The **Site-Level Metrics** tab has three key features developed specifically to support the Site Feasibility group. This tab includes the USA Site Counts, the USA Site Map and the USA Site List. These features drill down to the site and investigator and provide details by subject diversity categories.

In the third example, the USA Site Map (Figure 3) is created depicting the site location and primary investigator recruiting of the specific population selected. In this example, the user is interested in locating the sites and identifying the investigators that entered the Hispanic or Latino population for Study STY5-0001.

Figure 3. USA Site Map of the Hispanic or Latino Randomized Population for Study STY5-0001



Note: *All the visuals are dynamically created and are based on dummy data to show the functionality of the Diversity Dashboard. 'Subject' and 'Patient' & 'Trial' and 'Study' are used interchangeably throughout the paper.*

Conclusion

Data is the crucial part of clinical trials, particularly in monitoring diversity and representation. The Diversity Dashboard aids in enacting the FDA Diversity guideline by making the Clinical Operations teams acutely aware of what is occurring at the site level and what actions may need to be mitigated to enroll and retain diverse participants. By tracking trial recruitment data against benchmarks, the trial team can take proactive steps to ensure diversity and proper representation in clinical trials. The Diversity Dashboard tool permits real-time monitoring of diversity factors and helps identify sites and investigators from historical clinical trial data.

If the Diversity Dashboard can support your company's requirement of monitoring diversity metrics and need additional information about the tool, please contact us nancy.bauer@boehringer-ingelheim.com or julie.shah@boehringer-ingelheim.com. ■

Bibliography

[1] U.S. Department of Health and Human Services Food and Drug Administration: Oncology Center of Excellence (OCE), Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Devices and Radiological Health (CDRH), Office of Minority Health and Health Equity (OMHHE), 2020, *Diversity Plans to Improve Enrollment of Participants from Underrepresented Racial and Ethnic Populations in Clinical Trials Guidance for Industry (fda.gov)*

[2] European Medicines Agency Committee for Medicinal Products for Human Use (CHMP), 2018, *Geriatric Expert Group Reflection paper on physical frailty_adopted (europa.eu)*, EMA/CHMP/778709/201

STATISTICIANS UNDERPIN ASSURANCE OF PHARMACEUTICAL QUALITY FOR THE PATIENTS: SOME CONSIDERATIONS TO DO SO EFFECTIVELY

Ajaz S. Hussain, Ph.D., *Independent Advisor and Consultant,* Frederick, MD 21702. www.linkedin.com/in/ajazshussain

Introduction

I recently attended the 2023 American Statistical Association Biopharmaceutical Section Nonclinical Biostatistics Conference (NCB 2023), where I had the honor of delivering a keynote speech on "Statistical Thinking and Pharmaceutical [professional] Development" (1). The conference was a great learning experience, and this after-action report is to share my reflection and to learn from this experience.

The report is experiential and written with attention to "integrity of experience" to interpret *scientific training and experience* needed to ensure the *effectiveness* in developing new drug products *fairly and responsibly*; the phrases italicized are to invoke the letter and spirit of the Kefauver-Harris Amendments (1962) to the Food, Drug, and Cosmetic Act of 1938. It looks to bridge gaps between perceptions and reality we experience, individually and collectively, as a team, an association, and a society. It suggests some ways to confront variable perceptions of quality. It calls for us to do more to underpin the assurance of pharmaceutical quality patients need in the real world.

ACTION AFTER DOING: COMMON SENSE

In the familiar management tool, the Plan-Do-Check-Act (PDCA) cycle, action is separated from *doing to* check and study what should be done differently next time. It is to correct errors, prevent mistakes, and improve. For example, in reviewing, I noticed a few things that needed to be corrected in the content of the slides I had prepared (1). But how do I study what I didn't document, my experience (e.g., at event NCB 2023)? How can I verify its integrity?

A decade ago, I unexpectedly found myself embarking on a crucial mission to combat "*illusions of quality*" (2) and tackle *breaches in the assurance of data integrity* (BAD-I) in laboratory and factory environments across several corporations and regions. I chose to empower sidelined workers (3), and the experiences I gathered on this journey have significantly enriched my wisdom and understanding today.

Helping someone overcome feelings of helplessness and anxiety, such as feeling benched in a laboratory when one encounters an out-of-specification result or sweating on a factory floor when a procedural noncompliance occurs, needs to empower them to be more confident, to report, correct, and prevent *errors of commission* and be in control of their life, and able to claim their rights. This process should be practical and independent (without hired consultants). It should make intuitive sense, which I suggest is an aspect of making common sense, and emphasize experiential learning, drawing from both personal encounters and the shared experiences of others through reading and listening..

But often, it does not make common sense across an organizational hierachy, particularly in the "*FDA Approved*" and "*Validated*" process context. Furthermore, in the letter and spirit of the FD&C Act, for a community of knowledge with significant *scientific training and experience*, it is not *fair nor responsible* to ignore *errors of omission*, as argued in the paper "Pharmaceutical "New Prior Knowledge": Twenty-First Century Assurance of Therapeutic Equivalence" (4).

"Not doing something we should have" [to be fair and responsible] is an *error of omission*. Since pharmaceutical law and orders precede and evolve with each crisis at a rate faster than the evolution and maturity of pharmaceutical science and technology, preventing such errors is a constant struggle. Some corporations adopt a "*don't use and don't tell [FDA]*" policy on their use of process analytical technology to characterize pharmaceutical and control manufacturing processes (5). In contrast, others continue with outdated information and technology until the next crisis, resulting in a new amendment to the FD&C Act and a new FDA guidance finalized, which can take years, sometimes

many decades, as exemplified by the case of diethylene glycol used as an excipient at the root of the 1937 tragedy in a bottle of Elixir and now a continued risk of contamination and adulteration (6, 7); which on the FDA website is referred to as "*Sulfanilamide Disaster*" (8). Have you wondered why the tragedy is titled in a way that blames the drug substance and not the Elixir formulation? Extending the questions – why do we name regulatory applications New Drug Applications when FDA does not approve a drug substance, it authorizes application for pharmaceutical products? Might pharmaceutical formulations still be a blind spot?

Errors of commission, that is, doing something one is not supposed to do, are easier to detect and, when caught, should be punished [individually]. When an investigation on an out-of-specification resul ends in the "*root cause is unknown*," something still needs to be done in response to an FDA inspection. A typical corporate intervention is hiring consultants to tweak the SOPs, retrain the staff and move on; it is just the cost of doing business.

Life in pharmaceutical manufacturing can feel like that of a *caged hamster on a wheel*. Often a common challenge for cause, and the proposed solution must make common sense. But then, isn't "*common sense the collection of prejudices acquired by age eighteen*" - Albert Einstein (9)?

The nation and we can be wiser when we learn from errors, "*From the errors of other nations, let us learn wisdom*" — Thomas Paine, Common Sense (10). Common sense, to be reasonable, i.e., valid as in good manufacturing practice, needs *integrity of experience,* which is the purpose of writing this after-action report.

BEHAVIORAL ECONOMICS AND PROCRUSTEAN EXTENSION NEGLECT

A decade ago, my journey to empower workers was experiential. I was conducting training on the culture of pharmaceutical quality (11, 12, 13) in a monsoon season of a downpour of FDA Warning Letters. I named it "Schrödinger's Cat & My Journey From 2015 to 2020" in a blog on LinkedIn (14).

Schrödinger's Cat was not just a metaphor for pharmaceutical quality, good or not, depending on observation in Form 483 following an FDA inspection; it had a deeper meaning about what life is.

Schrödinger's book What is Life inspired me to renounce the *noblesse*, if any, and to be freed of the obligation on my journey to explore dimensions of quality by design which I had ignored while at the US FDA (15, 16).

What does *noblesse oblige* mean? It is a tradition, as Schrödinger notes, that *a scientist is supposed to have a complete and thorough knowledge, at first hand of some subjects and, therefore, is usually expected not to write on any topic of which he is not a master. This is regarded as a matter of nobleness* (17).

Renouncing the *noblesse*, if any, freed me to share insights by writing about subjects I did not hold a formal master's degree in, i.e., human behavior, behavioral economics, and why in the experience economy, the *integrity of experience* matters more than formal scientific training. With this freedom and from an experiential viewpoint, some challenges in pharmaceutical development and the variable interpretation of pharmaceutical quality can be seen from a different and nuanced perspective.

For instance, the representativeness heuristic can become a cognitive bias, *extension neglect*, which occurs when the sample size is ignored when its determination is relevant (18). It can still be evidenced in the continued need for the FDA (as in slide #36 of the 6 May 2011 presentation by Grace McNally (19)) to remind the industry not to confuse the role of pharmacopeial or "market standards" with quality control testing for batch release decisions; *at times, compendial standards take on the character of statistical procedures, with multiple units involved and perhaps a sequential procedural design to allow the user to determine that the tested article meets or does not meet the standard. The similarity to statistical procedures may seem to suggest an intent to make inferences to some larger group of units, but in all cases, statements about whether the compendial standard is met apply only to the units tested. Repeats, replicates, statistical rejection of outliers, or extrapolations of results to larger populations, as well as the necessity and appropriate frequency of batch testing, are neither specified nor proscribed by the compendia. First-party (manufacturer), second-party (buyer), or third-party (regulator) compliance testing may or may not require the examination of additional specimens in accordance with predetermined guidelines or sampling strategies (USP 33–NF 28 Reissue General Notices).*

Similarly, the behavioral influence and misuse of "defaults," such as a scale-up factor of 10X or 3-batches, as in Scale-Up and Post Approval Change Guidance documents, become apparent. Defaults can blind us and prevent us from correctly analyzing prior knowledge (i.e., its generalizability or applicability to

a particular situation) or verifying our assumptions. There are severe consequences of this misstep. Our knowledge pyramid can easily be toppled (20).

Our effort to investigate the root cause of out-of-specification is often a "streetlight effect," like searching for *keys lost somewhere in the darkness of a parking lot under a streetlight*. So we must be beware of (Regulatory) defaults, which behavioral economics informs us to mature the management of pharmaceutical QbD (21). With repeated efforts, many adopt the "if I don't. ok, there is no problem," it takes a lot of time and energy to overcome the "*immunity to change*" (22) to break the 2-3 sigma barrier and stop recycling "root cause unknown" excuse (23).

My journey to 2020 was to clear my foggy brain in a "Fog of War" that engulfs a corporation after a devastating Warning Letter. It soon focused on the *Procrustean* problem of a prescriptive or one-size-fits-all mindset (24).

As a grandfather, I realized *Procrustes* is an archetype, a pattern of behavior, and a stage in human development (25). The Covid-19 brain fog is now a new fog of war of mistrust. Not to get distracted, I continued my journey in and beyond COVID-19 to engage the US academic pharmaceutical science and engineering community to meet critical national needs (26).

My visit to NCB 2023 was another station in the journey. What was this experience like? What did I learn? How do I wish to remember this experience?

REMEMBERING AN EXPERIENCE WITH INTEGRITY

Indeed, it was a significant learning experience, even though I missed many goals I had intended to achieve. Perhaps noticing the slides projected were different from the final set I wanted to use threw me off my plan; it shouldn't have. Instead of taking a moment to check and correct, I went off-script; why? I forgot to take a moment to take a deep breath and refocus.

In deviating from my visualized routine, I risk returning to old habits. I am aware of the Procrustean tendencies within me, which I consider a lower level of consciousness (27). This mindset involves struggling to manage impulses and maintaining a need for control, which can lead to irrational anxiety and hyperactivity. It can also trigger a "fight- or- flight" response in situations where it is unnecessary. Despite this, it is only one aspect of my consciousness, including higher orders.

Consciousness orders are nested, with lower levels existing within higher ones. Is employing a nested experimental design for conducting N-of-1 thought experiments feasible to avoid potential pitfalls? I intended to pose this query at NCB 2023, but it slipped my mind.

This experience provides insights to appreciate more deeply than I had previously on the need to SMART in "*tiny details and the big picture*, " so I conclude this report with a few considerations to do so effectively in the memory of a former colleague at the US FDA, Dr. Stella Machado and acknowledging past collaborations with Dr. Meiyu Shen, who attended the NBC 2023. In remembrance of Dr. Stella Machado, Meiyu wrote (28), "*She was able to focus on the big picture and the tiny detail, which helped me advance my professional growth rapidly*."

I fondly remember a particular event in April 2002 at the US FDA when Dr. Machado and her team of biostatisticians, including Dr. Shen and Dr. Anello, worked in collaboration with the PAT team to distill information for a presentation dated 9 April 2002 by the Center Director (29); which I incorrectly attributed my presentation slides at NCB 2023 (in slide #11 (1)), on the core issue which now I describe as '*representativeness heuristics*' a short-cut in mind that can become a cognitive bias, *extension neglect*. It can trap a management system in barriers such as 2-3 sigma barriers inherited in legacy; individually, these are addictive and are self-imposed. When one thinks about it, the Center Director's presentation was titled "new challenges," a cognitive bias hidden in plain sight, a blog post (29).

Indeed, when tested, conformance to market standards is a legal requirement; shouldn't in-house standards be consistently aimed higher and expressed in ways that verify the assumption of normality, obtain robust estimates of variance without penalizing for an increase in sample size and express confidence instead of a check for a pass or fail? Furthermore, in the context of QbD, shouldn't regulatory specifications be a topic at the end of the Phase II discussion? A Procrustean approach to specification setting (e.g., imposing a black box "discriminating" dissolution test and limits) should be avoided after the Phase III trials are completed. When we do so, we lock in issues and stamp them "FDA Approved." Statisticians, as members of the pharmaceutical "team science," in poking and prodding similar questions, can do more to underpin pharmaceutical quality assurance for patients over the product life cycle. The opportunity to consider the "totality of the

evidence" instead of a "pivotal" trial is available, but it needs nurturing to make common sense.

At the heart of statistics is the core issue of uncertainty – allosteric and epistemic. In development, we work to reduce epistemic uncertainty, which naturally is amenable to Bayesian development, building prior knowledge, information, and data. Indeed I was thrilled to note at NCB 2023 that "Continuous Method Validation: Beyond One-Time Studies to Characterize Analytical Methods" (30) won first place in the Best Paper Awards. I was so excited that I took a photo to capture the moment (Figure 1), read the abstract online, and shared it on my LinkedIn feed.

Why was I so excited? The notion of "*Continuous Method Validation*" coupled with "*Continued Process Verification*," as in the current FDA guidance on "Process Validation: General Principles and Practices" (31), offers a solution to counter extension neglect and other cognitive biases and prevent the formation of *Procrustean trap* and *Popperian trap*.

What is the *Popperian trap*? In a different communication, I plan to elaborate on the Popperian *web* to discuss how statisticians can help to avoid the misuse of falsification standards for profiteering by discounting post-approval real-world observations and undermining evidence that demands action as in the PDCA cycle.

This after-action report was experiential. Writing in and of itself is SMART, *self-monitoring, analyzing, and reporting technology*, a tool to learn and use the power of intentional thinking in professional and personal life. Writing coupled with *sentiment monitoring analysis and report writing in real-time* is the other SMART dimension that helps me gauge my Order of Consciousness; drafts 2 and 3 break the procrustean trap to raise consciousness to be as close to being "*self-transforming*" as I could in completing this the final draft.

So, in the context of the fourth industrial revolution raging in an experience economy, consider how the integrity of experience matters and why it might matter more than you think in tiny details and the big picture. So then, "*a grain of wise subjectivity tells us more about the real world than any amount of objectivity*." — Judea Pearl (32) is a valuable reminder to conclude this experience. ■

Figure 1. The affect experienced at singular moments strongly influences retrospective evaluations of affective episodes. A moment at NCB 2023 to cherish.

REFERENCES

1. Hussain, A.S. "Statistical Thinking and Pharmaceutical [professional] Development," a keynote at the 2023 American Statistical Association Biopharmaceutical Section Nonclinical Biostatistics Conference. College Avenue Student Center at Rutgers University. 20 June 2023. Slides posted at: https://www.slideshare.net/a2zpharmsci/statistical-thinking-and-pharmaceutical-professional-development-a-keynote-by-ajaz-ncb-2023-20-june-2023 (Accessed 06.26.2023).

2. Cox, B. Confronting Illusions of Quality in Indian Generics Manufacturing. 28 March 2014. This article was originally published in The Gold Sheet. https://pink.pharmaintelligence.informa.com/PS000821/Confronting-Illusions-of-Quality-in-Indian-Generics-Manufacturing (Accessed 06.26.2023).

3. Cox, B. Hussain: Empowering Workers is the Key to Data Integrity. 28 March 2014. This article was initially published in The Gold Sheet. https://pink.pharmaintelligence.informa.com/PS000823/Hussain-Empowering-Workers-is-the-Key-to-Data-Integrity (Accessed 06.26.2023).

4. Hussain, A.S., Gurvich, V.J. & Morris, K. Pharmaceutical "New Prior Knowledge": Twenty-First Century Assurance of Therapeutic Equivalence. AAPS PharmSciTech 20, 140 (2019). https://doi.org/10.1208/s12249-019-1347-6 (Accessed 06.26.2023).

5. Hussain, A.S., Morris, K. & Gurvich, V.J. Pharmaceutical Quality, Team Science, and Education Themes: Observations and Commentary on a Remarkable AAPS PharmSciTech Theme Issue. AAPS PharmSciTech 22, 88 (2021). https://doi.org/10.1208/s12249-021-01970-7 (Accessed 06.26.2023).

6. Hussain, A.S. Product Quality & Patient Safety. USP Workshop, Mumbai, India, 12 June 2015. https://www.slideshare.net/a2zpharmsci/product-quality-patient-safety-usp-workshop-mumbai-12-june-2015 (Accessed 06.26.2023).

7. US FDA, Guidance Document. Testing of Glycerin, Propylene Glycol, Maltitol Solution, Hydrogenated Starch Hydrolysate, Sorbitol Solution, and Other High-Risk Drug Components for Diethylene Glycol and Ethylene Glycol. May 2023. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/testing-glycerin-propylene-glycol-maltitol-solution-hydrogenated-starch-hydrolysate-sorbitol (Accessed 06.26.2023).

8. Sulfanilamide Disaster: Taste of Raspberries, Taste of Death The 1937 Elixir Sulfanilamide Incident. FDA Consumer magazine. June 1981 Issue. https://www.fda.gov/about-fda/histories-product-regulation/sulfanilamide-disaster (Accessed 06.26.2023).

9. Horgan, J. Einstein and Science's Assault on Common Sense. Scientific American. 27 August 2015. https://blogs.scientificamerican.com/cross-check/einstein-and-science-s-assault-on-common-sense/ (Accessed 06.26.2023).

10. Pain, T. Common Sense. January 1776. Republished in America in Class. National Humanities Center. Text from 3d. Ed. (Bradford) of 14 February 1776, courtesy of Early American Imprints, American Antiquarian Society with Readex/NewsBank, #43116. http://americainclass.org/wp-content/uploads/2014/07/Common-Sense-_-Full-Text.pdf (Accessed 06.26.2023).

11. Hussain, A.S. The culture of Pharmaceutical Quality: Personnel Development. Biopharma Asia March/April 2016. https://biopharma-asia.com/magazine-articles/culture-of-pharmaceutical-quality-personnel-development/ (Accessed 06.26.2023).

12. Hussain, A.S. The culture of Pharmaceutical Quality Management System. Biopharma Asia November/December 2015. https://biopharma-asia.com/magazine-articles/culture-of-pharmaceutical-quality-improving-quality-management-system/ (Accessed 06.26.2023).

13. Hussain, A.S. The culture of Pharmaceutical Quality: Connecting the Dots. Biopharma Asia 34-38. Issue September/October 2015. https://biopharma-asia.com/magazine-articles/culture-of-pharmaceutical-quality-connecting-the-dots/ (Accessed 06.26.2023).

14. Hussain, A. S. Schrödinger's Cat & My Journey From 2015 to 2020. 23 December 2014. LinkedIn. https://www.linkedin.com/pulse/schr%25C3%25B6dingers-cat-my-journey-from-2015-2020-ajaz-s-hussain-ph-d-/?trackingId=23kXhWI6Tr%2BnWkAEh4BYHw%3D%3D (Accessed 06.26.2023).

15. Hussain, A.S. Chemometrics, Pharmacometrics, and Econometrics: Dimensions of Quality by Design. SWISS PHARMA 34 (2012) Nr. 6. Authors copy available at: https://www.slideshare.net/a2zpharmsci/chemometrics-pharmacometrics-econometrics-of-qb-d-swiss-pharma6-2012 (Accessed 06.26.2023).

16. Hussain, A.S. Chemometrics, Pharmacometrics, and Econometrics: Dimensions of Quality by Design. 25 May 2012 Basel, Switzerland. Presentation slides. https://www.slideshare.net/a2zpharmsci/chemometrics-pharmacometrics-andeconometrics-dimensionsofquality1 (Accessed 06.26.2023).

17. Schrödinger, E. What Is Life? The Physical Aspect of the Living Cell. Based on lectures delivered under the auspices of the Dublin Institute for Advanced Studies at Trinity College, Dublin, in February 1943. First published in 1944. https://www.academia.edu/10706644/What_is_Life_by_Erwin_Schrodinger?auto=download (Accessed 06.26.2023).

18. Kahneman, D. Evaluation by Moments: Past and Future. Chapter 38 in: D. Kahneman and A. Tversky (Eds.) Choices, Values and Frames, New York: Cambridge University Press, and the Russell Sage Foundation (2000). http://cdi.mecon.gov.ar/bases/docelec/mm1092.pdf (Accessed 06.26.2023).

19. Grace E. McNally. Process Validation A Lifecycle Approach. 6 May 2011. https://variation.com/wp-content/uploads/guidance/Presentation-Process-Validation-A-Lifecycle-Approach-Grace-McNally-2011.pdf (Accessed 06.26.2023)

20. Hussain, A.S. Toppled Knowledge Pyramid: Survival Mechanism. 3 March 2016. LinkedIn. https://www.linkedin.com/pulse/toppled-knowledge-pyramid-survival-mechanism-ajaz-s-hussain-ph-d-/?trackingId=BwdEIA5ySZ684qLt%2BH7XMg%3D%3D (Accessed 06.26.2023)

21. Hussain, A. S. Beware of (Regulatory) Defaults: Behavioral Economics & Management of Pharmaceutical QbD. 25 August 2016. LinkedIn. https://www.linkedin.com/pulse/beware-regulatory-defaults-behavioral-economics-qbd-hussain-ph-d-/?trackingId=atgT1KB6Q2%2BTpmb%2BcRMwDg%3D%3D (Accessed 06.26.2023)

22. Kegan, R. and Lahey, L.L. Immunity to change: How to overcome it and unlock potential in yourself and your organization." Harvard Business Press. 2009.

23. Hussain, A.S. How To Break the Pharmaceutical 2-3 Sigma Barrier (Like Amgen). Guest Column. Pharmaceutical Online. 18 September 2017. https://www.pharmaceuticalonline.com/doc/how-to-break-the-pharmaceutical-sigma-barrier-like-amgen-0001 (Accessed 06.26.2023)

24. Hussain, A. S. Procrustean Problem in Prescriptive Regulations: A journey to clear my foggy brain in a "Fog of War." 11 April 2016. LinkedIn. https://www.linkedin.com/pulse/procrustean-problem-prescriptive-regulations-journey-hussain-ph-d-/?trackingId=BwdEIA5ySZ684qLt%2BH7XMg%3D%3D (Accessed 06.26.2023)

25. Hussain, A. S. Six Blind Men, Regulatory Science, "Bob's Big Idea" and Gandhi's "a way out of hell": Goodbye 2015. 29 December 2015. LinkedIn. https://www.linkedin.com/pulse/six-blind-men-regulatory-science-bobs-big-idea-way-hussain-ph-d-/?trackingId=BwdEIA5ySZ684qLt%2BH7XMg%3D%3D (Accessed 06.26.2023)

26. Gurvich, V.J., Hussain, A.S. In and Beyond COVID-19: US Academic Pharmaceutical Science and Engineering Community Must Engage to Meet Critical National Needs. AAPS PharmSciTech 21, 153 (2020). https://doi.org/10.1208/s12249-020-01718-9 (Accessed 06.26.2023).

27. Kegan, R. In over our heads: The mental demands of modern life. Harvard University Press. 1995.

28. Memories and Condolences for Stella Machado. Published by Going Home Cremation & Funeral Care by Value Choice, P.A. on 31 December 2022. https://www.legacy.com/us/obituaries/name/stella-machado-obituary?id=38513657 (Accessed 06.26.2023).

29. Woodcock, J. FDA Regulation of Drug Quality: New Challenges. 9 April 2002. https://www.slideserve.com/lindsay/fda-regulation-of-drug-quality-new-challenges (Accessed 06.26.2023).

30. Faya, P., Novick, S.J., Seaman, J.W., Peterson, J.J., Pourmohamad, T., Banton, D., & Zheng, Y. (2022). Continuous Method Validation: Beyond One-Time Studies to Characterize Analytical Methods. Statistics in Biopharmaceutical Research, 15, 225 - 233.

31. FDA Guidance for Industry, Process Validation: General Principles and Practices. JANUARY 2011. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/process-validation-general-principles-and-practices (Accessed 06.26.2023).

32. Pearl, J., and Mackenzie, D. The Book of Why: The New Science of Cause and Effect. Penguin Books Limited, May 15, 2018.

ENHANCING TRANSLATIONAL RESEARCH SUCCESS: THE ROLE OF STATISTICAL INNOVATION

Xin Huang, Abbvie Inc.; Haiyan Zheng, University of Cambridge; and Steven Novick, Eli Lilly and Company

Translational research plays a crucial role in ensuring the success of drug development. Its effectiveness lies in facilitating the seamless translation of scientific discoveries and knowledge from basic research into human clinical trials. This process involves bridging the gap between laboratory findings and their practical implementation in clinical practice, while ensuring their relevance, safety, and effectiveness. Translational research thus must be firmly grounded in rigorous scientific methods and evidence-based practices, which is essential to warrant thorough evaluations of the experimental medicine through preclinical studies, clinical trials, and real-world evidence studies. The success of translational research heavily relies on robust empirical evidence. However, it is widely acknowledged that translating basic scientific findings from the lab to practical applications, potential disease treatments or biomarkers (FDA-NIH Biomarker Working Group, 2016) presents significant challenges in the pharmaceutical industry. Despite ongoing efforts in academic and industry settings to address this problem, attrition rates in drug development remain high and issues surrounding reproducibility and translatability of preclinical findings to human applications persist. This limits the clinical impact and return on investment. This article discusses the role of recent statistical innovation in facilitating the success of translational research. Specifically, our discussion focuses on the topic of statistical methods to ensure replicability and reproducibility of preclinical studies, recommendation of preclinical usage of p-values, recent development of Bayesian modeling for animal to first-in-human translation, and AI/ML based pharmacological modeling for human trial prediction.

1. Replicability and reproducibility of preclinical studies

Recently, translational medicine has focused its attention on improving *replicability* and *reproducibility* of preclinical experiments. While a study result is replicable when a repeated study leads to the same conclusion, experimental results are considered as reproducible if the data analysis produces the same outcome using the same data set and/or computer code . It readily follows that confidence in go/no-go decision-making study is higher when associated study results are both replicable and reproducible. The replicability of preclinical experiments has been shown to be as low as 10% (Begley & Ellis, 2012; Border et al., 2019; Prinz, Schlange, & Asadullah, 2011). Several publications have raised concerns regarding reproducibility issues, with difficulties in duplicating results from original source data. Baggerly et al (Keith & Kevin, 2009) highlighted concerns stemming from poor documentation. Ioannidis et al (Ioannidis et al., 2014) showed that deriving the same results from 16 out of 18 studies using the original raw data was not possible. Nekrutenko & Taylor (Nekrutenko & Taylor, 2012) pointed out a lack of access to the primary data or software in 26 out of 50 studies. Furthermore, Perkel (Perkel, 2020) and Hinsen (Hinsen, 2020) outlined a spectrum of reproducibility issues, including problems in reproducing one's own results .

In the traditional preclinical workflow (Hughes, Rees, Kalindjian, & Philpott, 2011), replicability is implicit rather than explicit. For example, a gene target may be identified by examining omics data derived from several phase III clinical trials. Rather than repeating the omics study, the target may be validated in a subsequent CRISPR in-vitro experiment in which the activity from a knocked-out gene of interest is compared to the control. Following non-replicated success, a drug compound may be developed to determine if the gene knock-out activity can be reversed. Eventually, the drug candidate will be tested in an animal model. Each success builds from the previous step and may be seen as a surrogate for direct replication with the thought: *if the results were not replicable, we could not have gotten this far.* While this recipe can certainly work for some drug

discovery campaigns, false positives and false negatives may result from small and underpowered preclinical studies. Indeed, replicability of animal studies and translatability to human trials can be challenging. Yet, most pharmaceutical companies would recoil at the cost and time required to repeat in-vivo studies. Alternative solutions must be explored.

2. Preclinical usage of *p*-values

One area that has recently received attention is the reliance of *p*-values for critical preclinical decision-making. In their seminal paper, Wasserstein and Lazar (Wasserstein & Lazar, 2016) stated that *p*-values do not reflect the probability that the null hypothesis or the alternative hypothesis is true and stress that scientific conclusions should not be based solely on the observation that a *p*-value is less than a pre-specified test size. Focused on the use of *p*-values in the nonclinical tests, Altan et al (Altan et al., 2023) generally agree with Wasserstein and Lazar, further citing *p*-value hacking, which occurs when an investigator attempts to manipulate a *p*-value below the test size by selectively including or excluding data (i.e., "cherry-picking data"). The authors discussed the loose, unregulated nature of preclinical studies, in which competing hypotheses and testing methods may not have been established until after examining the collected data. The authors also stated that some tests lacked multiplicity correction, which could yield false positive results. All of these can lead to lack of replicability in the preclinical laboratory and, of course, a lack of medicines translation to the human population. Altan et al (Altan et al., 2023) suggest common fixes for these issues, including strict compliance with *good statistical practice* as outlined in the Animal Research: Reporting of In Vivo Experiments (ARRIVE) guidelines (Percie du Sert et al., 2020) and false-discovery rate corrections with tools, such as Benjamini and Hochberg (Benjamini & Hochberg, 1995), Storey and Tibshirani (Storey & Tibshirani, 2003), and Edwards and Berry (Edwards & Berry, 1987).

In addition, Altan et al (Altan et al., 2023) point out that many preclinical hypotheses are to test the mean differences are equal to zero (H_0) or not equal to zero (H_1). Rejection of the null hypothesis is not earth-shattering and results with *p*-values below a pre-assigned threshold should be viewed in the broader context of effect size and other related measured results. Further, for a single preclinical study, an observation that *p*-value < α indicates that the null hypothesis is rejected for *this* study. That is, "acceptance" of H_1 does not provide evidence that the H_1 conclusion would be drawn with a replicate study.

A truly replicable study must be conducted several times to yield the same conclusion. For example, if an in-vivo study was reproduced five times by the same laboratory and scientists (but different time periods and different animals) and all the five resulting *p*-values are less than, say, 0.05, this would provide good evidence for accepting H_1. What if only three out of five *p*-values are less than 0.05? The decision becomes less clear and the translation of medicines is in doubt. A better approach might incorporate random study effects into a statistical model so that inference may be drawn from one instead of five separate hypothesis tests. For a test of means, Novick and Zhang (S. Novick & Zhang, 2021) considered the model $Y_{ijk} = \mu_i + S_j + \gamma_{ij} + e_{ijk}$, where $S_j \sim N(0, \sigma_S^2)$ denotes a j^{th} random study effect $(j = 1, \ldots, 5)$, $\gamma_{ij} \sim N(0, \sigma_\gamma^2)$ denotes the study by treatment effect, and $e_{ijk} \sim N(0, \sigma_e^2)$ denotes the subject-to-subject errors. With the mixed-effects model, the distribution of the sample mean difference is

$$\bar{Y}_{Trt} - \bar{Y}_C \sim N\left(\mu_{Trt} - \mu_C, 2\left(\sigma_\eta^2 + \frac{\sigma_e^2}{n}\right)\right),$$

where the study by treatment effect is prominent. Inference made from this distribution will capture the variability of a random future study so that a *p*-value < 0.05 that includes the σ_η^2 term would suggest replicability. The extra confidence from incorporating the σ_η^2 term may provide better translation from the in-vivo animal model to the clinic. Novick and Zhang (S. Novick & Zhang, 2021) propose that σ_η^2 may be estimated from historical data. This is practical according to our experience, since many Research & Development groups in the pharmaceutical industry typically store in-vivo data.

For reproducibility, the goal is to be able to recreate the reported results using the original data. For example, historical data may have gone through cleaning and normalization steps (e.g., with 'omics platform data), and data analysis. An investigator may wish to repurpose the cleaned and normalized historical data with the assurance that the quality of the data remains identical to that used in the original analysis. Perkel (Perkel, 2020) and Hinson (Hinsen, 2020) show a variety of situations in which ill conditions lead to a

failure of reproducibility. Good reproducibility centers on project organization, data management, computer coding standards, and report generation. Novick and Konings (Steven Novick & Konings, 2022) provide a summary of steps for reproducibility that include common file systems across projects, the use of a workflow management system to outline steps taken from data acquisition to final data analysis, the use of software audit trails (e.g., software log from SAS or R), and software version control. With proper systems in place, reproducibility of study analysis can be improved.

3. Bayesian modeling for animal to human translation

Next, key steps in successful translational research include how to improve the design and conduct of animal studies as well as how to bridge the information from animal to human studies. In this domain, Bayesian methodology is increasingly promoted and considered. Walley et al. (Walley et al., 2016) describe the use of a Bayesian meta-analytic predictive approach (Neuenschwander, Capkun-Niggli, Branson, & Spiegelhalter, 2010) to incorporate any historical control data made available prior to a new animal experiment for enhanced design, conduct and analysis. A general benefit comprises the reduction of the total number of animals required or the increase in the precision of key parameter estimation, since the historical control may supplement or replace the contemporary control group entirely. For synthesising findings from various animal experiments, Bayesian random-effects meta-analysis models are shown to be advantageous in estimating the between-study heterogeneity (Tanriver-Ayder, Faes, van de Casteele, McCann, & Macleod, 2021).

In the transition step from animal experiments to phase I first-in-man trials, one major concern is whether the accumulated animal data have good predictability of the toxicity in humans. In Bayesian frameworks, prior or posterior predictive distributions can be a powerful tool to aid decision making. Zheng and Hampson (Zheng & Hampson, 2020) develop a Bayesian decision-theoretic approach that represents animal data in a prior for the dose-toxicity parameters in humans; the weight attributed to this prior is determined dynamically, according to the measured (in)commensurability between the dose-toxicity relationships in animals and humans. To leverage animal data from multiple species, Zheng and colleagues (Zheng, Hampson, & Jaki, 2021; Zheng, Hampson, & Wandel, 2020) consider robust Bayesian hierarchical models to synthesize animal and human toxicity data; scaling factors are used to translate doses administered to different animal species onto an equivalent human scale for the dynamic incorporation. These Bayesian approaches to interspecies translation can improve the success rate of early drug development, for their advantages of formally combining all relevant data for informed decision making.

4. AI/ML based pharmacological modeling for human trial prediction

Finally, the pharmaceutical industry is undergoing a transformative phase in translational research, driven by the application of AI/ML built on recent advancements in data collection and generation tools, robust information management and exchange systems, and advanced computing capabilities. In identifying biological targets and understanding disease relationships, AI/ML plays a pivotal role by analyzing vast amounts of information from scientific research, publications, and diverse data sources. This could include genomic, transcriptomic, proteomic, and other clinical data from healthy and diseased individuals. Leveraging the complexity and diverse origins of these datasets, AI/ML approaches enable researchers to extract meaningful patterns and knowledge and provides researchers with an exceptional opportunity to inform the selection of biological targets and relevant biomarkers (Vamathevan et al., 2019; Weissler et al., 2021).

Preclinical in vivo studies typically involve complex objectives, thus may require sophisticated methods such as pharmacological modeling, experimental designs for toxicological studies in animals, mechanistic studies in animal models. AI/ML has contributed to the development of predictive models for human trials (Bulitta et al., 2019; Shroff et al., 2022) using preclinical in vivo data. Integration of pharmacokinetics (PK) and pharmacodynamics (PD) is a fundamental aspect of drug development, as it enables understanding of drug-body interactions over time and the body's response. More specifically, PK/PD modeling, extensively utilized in both preclinical and clinical stages, provides insights into dose-response relationships. The rise of computational tools and modeling platforms has led to increased utilization of physiologically-based pharmacokinetic (PBPK), physiologically-based PK/PD (PBPK-PD), and Quantitative Systems Pharmacology

(QSP) modeling (Brogi & Calderone, 2021). Novel AI/ML algorithms, including artificial neural networks, are also being explored in PK/PD modeling. For example, recurrent neural networks, widely used in time series analysis, show promise in enhancing the accuracy of preclinical to clinical translations for complex PK/PD data analysis (Liu et al., 2021).Additionally, Lu et al. (Lu, Bender, Jin, & Guan, 2021) combined crucial pharmacological rules with neural ordinary differential equations within deep neural networks to predict patients' drug concentration and response for dosing recommendations.

Conclusion

In conclusion, translational research embodies an ongoing and iterative journey that relies on continuous learning, adaptation, and feedback. This dynamic process, spanning pre-clinical, clinical, and post-marketing phases, greatly benefits from the statistical innovation that ensures repeatability and reproducibility of experiments while controlling false discovery rate and false positive rate, and benefits from recent innovations of Bayesian modeling for animal to human translation. By leveraging AI/ML models and simulations, researchers can integrate information from diverse sources and stages, enabling evidence-based decision-making throughout the drug development process. On this topic, Jiang *et al* (Jiang et al., 2022) reviewed statistical methods to integrate big data (cross-cohort data aggregation, cross-modality data integration, and knowledge transfer through data reuse) in basic and translational cancer research collected in the clinical stage. The success of translational research hinges upon effective communication between statisticians, research scientists, clinicians, and industry experts in the study team. This collaboration fosters the seamless integration of knowledge and expertise across various domains, fostering innovation and propelling discoveries from the laboratory to the clinical setting. ■

References

Altan, S., Amaratunga, D., Cabrera, J., Garren, J., Geys, H., Kolassa, J., . . . Vukovinsky, K. (2023). Survey and Recommendations on the Use of P-Values Driving Decisions in Nonclinical Pharmaceutical Applications. *Statistics in Biopharmaceutical Research*, *15*(2), 343-358. doi:10.1080/19466315.2022.2038258

Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, 483(7391), 531-533. doi:10.1038/483531a

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(1), 289-300. doi:https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F., & Keller, M. C. (2019). No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples. *Am J Psychiatry, 176*(5), 376-387. doi:10.1176/appi.ajp.2018.18070881

Brogi, S., & Calderone, V. (2021). Artificial Intelligence in Translational Medicine. *International Journal of Translational Medicine, 1*(3), 223-285.

Bulitta, J. B., Hope, W. W., Eakin, A. E., Guina, T., Tam, V. H., Louie, A., . . . Hoover, J. L. (2019). Generating Robust and Informative Nonclinical In Vitro and In Vivo Bacterial Infection Model Efficacy Data To Support Translation to Humans. *Antimicrob Agents Chemother, 63*(5). doi:10.1128/AAC.02307-18

Edwards, D., & Berry, J. J. (1987). The efficiency of simulation-based multiple comparisons. *Biometrics, 43*(4), 913-928.

FDA-NIH Biomarker Working Group. (2016). BEST (Biomarkers, EndpointS, and other Tools) Resource. Silver Spring (MD): Food and Drug Administration (US).

Hinsen, K. (2020). [¬Rp] Stokes drag on conglomerates of spheres. *ReScience C, 6*(1), \#7. doi:10.5281/zenodo.3889694

Hughes, J. P., Rees, S., Kalindjian, S. B., & Philpott, K. L. (2011). Principles of early drug discovery. *Br J Pharmacol, 162*(6), 1239-1249. doi:10.1111/j.1476-5381.2010.01127.x

Ioannidis, J. P., Greenland, S., Hlatky, M. A., Khoury, M. J., Macleod, M. R., Moher, D., . . . Tibshirani, R. (2014). Increasing value and reducing waste in research design, conduct, and analysis. *Lancet, 383*(9912), 166-175. doi:10.1016/S0140-6736(13)62227-8

Jiang, P., Sinha, S., Aldape, K., Hannenhalli, S., Sahinalp, C., & Ruppin, E. (2022). Big data in basic and translational cancer research. *Nat Rev Cancer, 22*(11), 625-639. doi:10.1038/s41568-022-00502-0

Keith, A. B., & Kevin, R. C. (2009). Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology. *The Annals of Applied Statistics, 3*(4), 1309-1334. doi:10.1214/09-AOAS291

Liu, X., Liu, C., Huang, R., Zhu, H., Liu, Q., Mitra, S., & Wang, Y. (2021). Long short-term memory recurrent neural network for pharmacokinetic-pharmacodynamic modeling. *Int J Clin Pharmacol Ther, 59*(2), 138-146. doi:10.5414/CP203800

Lu, J., Bender, B., Jin, J. Y., & Guan, Y. (2021). Deep learning prediction of patient response time course from early data via neural-pharmacokinetic/pharmacodynamic modelling. *Nature Machine Intelligence, 3*(8), 696-704. doi:10.1038/s42256-021-00357-4

Nekrutenko, A., & Taylor, J. (2012). Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. *Nat Rev Genet, 13*(9), 667-672. doi:10.1038/nrg3305

Neuenschwander, B., Capkun-Niggli, G., Branson, M., & Spiegelhalter, D. J. (2010). Summarizing historical information on controls in clinical trials. *Clin Trials, 7*(1), 5-18. doi:10.1177/1740774509356002

Novick, S., & Konings, P. (2022). 2.27 - Replicability and Reproducibility in Drug Discovery. In T. Kenakin (Ed.), *Comprehensive Pharmacology* (pp. 617-626). Oxford: Elsevier.

Novick, S., & Zhang, T. (2021). Mean comparisons and power calculations to ensure reproducibility in preclinical drug discovery. *Stat Med, 40*(6), 1414-1428. doi:10.1002/sim.8848

Percie du Sert, N., Hurst, V., Ahluwalia, A., Alam, S., Avey, M. T., Baker, M., . . . Wurbel, H. (2020). The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. *PLoS Biol, 18*(7), e3000410. doi:10.1371/journal.pbio.3000410

Perkel, J. M. (2020). Challenge to scientists: does your ten-year-old code still run? *Nature, 584*(7822), 656-658. doi:10.1038/d41586-020-02462-7

Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov, 10*(9), 712. doi:10.1038/nrd3439-c1

Shroff, T., Aina, K., Maass, C., Cipriano, M., Lambrecht, J., Tacke, F., . . . Loskill, P. (2022). Studying metabolism with multi-organ chips: new tools for disease modelling, pharmacokinetics and pharmacodynamics. *Open Biol, 12*(3), 210333. doi:10.1098/rsob.210333

Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A, 100*(16), 9440-9445. doi:10.1073/pnas.1530509100

Tanriver-Ayder, E., Faes, C., van de Casteele, T., McCann, S. K., & Macleod, M. R. (2021). Comparison of commonly used methods in random effects meta-analysis: application to preclinical data in drug discovery research. *BMJ Open Sci, 5*(1), e100074. doi:10.1136/bmjos-2020-100074

Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., . . . Zhao, S. (2019). Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov, 18*(6), 463-477. doi:10.1038/s41573-019-0024-5

Walley, R., Sherington, J., Rastrick, J., Detrait, E., Hanon, E., & Watt, G. (2016). Using Bayesian analysis in repeated preclinical in vivo studies for a more effective use of animals. *Pharm Stat, 15*(3), 277-285. doi:10.1002/pst.1748

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician, 70*(2), 129-133. doi:10.1080/00031305.2016.1154108

Weissler, E. H., Naumann, T., Andersson, T., Ranganath, R., Elemento, O., Luo, Y., . . . Ghassemi, M. (2021). The role of machine learning in clinical research: transforming the future of evidence generation. *Trials, 22*(1), 537. doi:10.1186/s13063-021-05489-x

Zheng, H., & Hampson, L. V. (2020). A Bayesian decision-theoretic approach to incorporate preclinical information into phase I oncology trials. *Biom J, 62*(6), 1408-1427. doi:10.1002/bimj.201900161

Zheng, H., Hampson, L. V., & Jaki, T. (2021). Bridging across patient subgroups in phase I oncology trials that incorporate animal data. *Stat Methods Med Res, 30*(4), 1057-1071. doi:10.1177/0962280220986580

Zheng, H., Hampson, L. V., & Wandel, S. (2020). A robust Bayesian meta-analytic approach to incorporate animal data into phase I oncology trials. *Stat Methods Med Res, 29*(1), 94-110. doi:10.1177/0962280218820040

ASA BIOP SECTION NEWSLETTER ON ASA-DAHSHU IDSWG MULTIDISCIPLINARY MASTER PROTOCOL WORKING GROUP

Robert Beckman (*Georgetown University*), Deborah Collyar (*Patient Advocates in Research (PAIR)*), Freda Cooner (*Eli Lilly*), Jonas Haggstrom (*Cytel*), Julia Niewczas (*Johnson & Johnson*), Chengxing (Cindy) Lu (*AstraZeneca*), May Mo (*Amgen*), Melissa E Spann (*Eli Lilly*), Kyle Wathen (*Cytel*), Jingjing Ye (*BeiGene*)

On Behalf of ASA-DahShu IDSWG Multidisciplinary Master Protocol Working Group

Introduction

Designing and implementing a complex and innovative clinical trial, such as a master protocol, often requires multidisciplinary collaborative efforts, including clinical development, biostatistics, regulatory science, clinical operation, and patient engagement. However, in reality, practical barriers exist during this immense collaborative effort. With this consideration in mind, the American Statistical Association (ASA) DahShu IDSWG Multidisciplinary Master Protocol Working Group (MPWG) was chartered in 2021 with joint effort from multiple working groups: the ASA BIOP Oncology Methods Scientific Working Group-Master Protocol subteam, the DahShu Innovative Design Scientific Working Group (IDSWG) and the Bayesian Scientific working group (BSWG). The objective is to overcome potential barriers to designing and executing master protocols and to promote multidisciplinary collaboration. Within this joint WG, five subteams were assembled focusing on 1) concept, framework and methodology; 2) confirmatory basket trials; 3) regulatory considerations; 4) implementation standards and 5) people and patient engagement (PE).

Concept, Methodology, and Framework

The goals of the Concept, Methodology, and Framework Sub Team include establishing general definitions/concepts on master protocols that provide multidisciplinary value including 1) an overview on statistical methods and general considerations/recommendations; 2) master protocol multidisciplinary evolving practices; 3) other aspects that relate to a master protocol framework. The workstreams aim to define the optimal conditions for utilizing a master protocol, create a comprehensive framework to assist sponsors in making decisions regarding master protocols and their functional support, evaluate the impact of using concurrent control versus nonconcurrent control in decision-making processes, and investigate the utility of modeling and simulation tools for effective master protocol design.

A master protocol quick start tool is currently under development by the team, which proves to be valuable in specifying objectives, indication specific design elements of known existing studies, and necessary design elements for potential individual protocol appendices (IPAs). This tool facilitates identifying essential elements for the indication of interest and how those elements align with asset needs for single indication umbrella/platform master protocols and for basket trials. Once design elements are specified, the master protocol scorecard that is also under development by the team, can be used to assess general considerations such as primary objective, pipeline/portfolio, patient population, visit schedule, endpoints/assessments, randomization, blinding, study duration, sites, screening/enrollment and features such as shared control, perpetual nature, and borrowing. The utilization of these tools helps in summarizing information, enabling a better assessment of whether the conditions are favorable for employing a master protocol. Principles of these tools plus a comprehensive guide for sponsors when assessing feasibility and planning/executing a master protocol will be the foundation of a presentation at 2023 ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop. The "Myth busting: Master Protocol Edition" presentation will explore and "bust" myths associated with master protocols and illustrate how many concerns are not unique to master protocols.

Confirmatory Basket Trial

Since the largest sample sizes and greatest costs and timelines are associated with the confirmatory phase of development, general application of basket trials in the confirmatory phase has the potential to significantly increase the efficiency of drug development. This may in turn lead to cost-effective development pathways for rare diseases as well as small biomarker defined subsets. However, up to this point, basket trials have primarily been utilized in the exploratory setting. Approvals and confirmatory studies have been restricted to exceptional drugs supported by robust scientific evidence, and in cases where they represent the last line of therapy with significant unmet medical needs. The vast majority of drugs cannot take advantage of this paradigm, and it has limited application to earlier lines of therapy.

The confirmatory basket trial sub-team, consisting of statisticians, a clinician, and several former health authority professionals from the United States and Europe, was formed with the goal of making a consensus recommendation concerning a more generalizable approach to confirmatory basket trials. Our current focus is on ensuring type I error control by conducting a thorough evaluation of available approaches in randomized settings. Additionally, we will address other potential concerns, such as assessing the adequacy of the safety database for the marketing application. Our plan involves creating an original consensus white paper, a book chapter, and delivering one or more presentations. If necessary, we will develop modifications to existing methods or introduce new methods to establish potentially viable approaches.

Regulatory Considerations

Master Protocols require more frequent and earlier interactions with regulatory authorities due to their complexity compared to fixed designs. The goal of the regulatory consideration sub-team is to bridge between sponsors and regulators that may help smooth future interactions. The subteam strives to take a deep dive into the current regulatory landscape and provide considerations for sponsors during future interactions, and on the other hand to identify multidisciplinary challenges and aspects that may need regulator's guidance. The team also aims at identifying and consolidating similarities and differences among the global agencies. The regulatory sub-team compares the feedback and considerations from FDA and EMA regarding multiplicity adjustments, benefit-risk assessments, use of non-concurrent controls, and estimands. In addition, very few articles are currently available by health technology assessments (HTAs) on master protocol from a reimbursement perspective. The sub-team discusses several considerations regarding the designs of master protocols, taking into account how HTAs would require the evaluation of the evidence. These considerations include a preference using clinically relevant endpoints (e.g., time to event based, quality of life), assessment of heterogeneity between subtypes in basket trial, and not in favor of the inter-dependency of results when a shared control is use in platform trial.

Implementation Standards

The Implementation Standards subteam includes members from industry and academic institutions with multi-disciplinary subject matter expertise in design and execution of master protocols, along with patient voice and tool developer representatives. The subteam strives to learn from real-life experience in master protocol implementation; collect data from survey and literature review; and develop general standards (e.g., templates and tools) and good practice recommendations. The goal is to maximize efficiency and reduce cost and time by streamlining trial logistics, improving data quality and collection, and reducing unnecessary complexity and redundancy. The recommendations will be summarized in a white paper focusing on the implementation of master protocols.

Implementation areas of focus for the sub-team are further detailed below:

- Protocol structure and amendments – The structure of master protocols, i.e., which contents should be in the common protocol sections, and which contents should be in IPAs I, is critical to master protocols. General guidance of the protocol structure and good practices to streamline protocol amendments will be shared.
- Trial integrity and oversight - Master protocols increase procedural complexity which may result in more errors in data collection, trial conduct and impact trial integrity. Good practices to ensure proper and sufficient level of

oversight (e.g., governance, steering committee) will be provided.

- Study startup – Best practices on the initial start-up for the master protocol and its first few sub-studies with the sites and vendor, contracts; patient engagement, database build, etc to enable future savings will be highlighted.
- Study conduct – With increased procedural complexity in site management, data monitoring, drug supply, etc, master protocols will require a higher level of coordination among operational functions, vendors, and clinical sites. Good practices to improve such coordination will be highlighted.
- Result reporting - Master protocols are required to clearly describe how trial subjects are allocated to the individual sub-studies; when are the interim and final analyses planned; what are the successful criteria and decision rules. Good practices in these areas and in clinical trial disclosure (clinicaltrials.gov and EU CT registries) per regulation will be provided.
- Safety monitoring –Regulatory requirements and industry experience with safety monitoring planning, reporting procedures, and internal or external monitoring committees such as iDMC and DSMB, will be shared.

People & Patient engagement (PE)

Patients should be at the center of drug development as experts in multiple aspects of their disease, including how it relates to: the signs and symptoms of disease, the burden of living with or managing a medical condition, barriers and difficulties of treatment (and the degree of unmet need), and barriers and burden of participating in clinical trials (US Food and Drug Administration (FDA), 2020). Patient engagement (PE) has been recognized by the FDA through the patient-focused drug development (PFDD) initiative and by the European Medicines Agency (EMA) with a statement on its website indicating that it has been actively interacting with patients since creation of the agency in 1995.

Although the advancement in clinical trial methodology and design and MPs have evolved over time, very little has been written about the important role that PE can play in the development and ultimate success of MPs. The People & Patient Engagement Sub-team (PE) of the MPWG is tasked to incorporate guidelines, standards and recommendations on how to include patients into clinical trials and MPs and to establish education and communication channels to patient advocacy groups (PAGs).

In this regard, the team has conducted a survey between October and December 2022 encompassing 43 patient advocates including individual patient advocates, and leaders of PAGs, spanning 26 organizations in the US and Canada. The goal of the survey is to gauge the overall understanding and awareness of MPs among PAGs that help educate and encourage participation in clinical trials. Of the 43 responses, only a small percentage (2%) participated in Master Protocols (MPs). Also measured in the survey, over 71% respondents are interested to be involved in planning new MPs. The participants are interested in co-creating educational programs with the drug developers and would like to include information of MPs on patient advocacy organization websites. The survey to PAGs has confirmed recommendations made by the PE sub team in the manuscript, "Aiding Adoption of Master Protocols by Optimizing Patient Engagement". A Plain Language Summary for publication (PLSP) of this manuscript is also planned for wide distribution.

Members in the five areas are working closely and interactively together within the WG to deliver multidisciplinary views, guidance, methods and tools for the future of master protocol usage. The team is passionate in sharing part of the ongoing work above at the "Multidisciplinary Considerations in Master Protocol Framework" session at 2023 ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop and in selected chapters of a master protocol book to be authored and edited by the MPWG and hear further feedback from the community. ■

ASA BIOPHARMACEUTICAL SECTION SCIENTIFIC WORKING GROUP PROPOSAL COMMITTEE UPDATE

The ASA BIOPHARMACEUTICAL SECTION (BIOP) has a well-defined process for establishing new scientific working groups (SWGs).

The process starts with the submission of a proposal, which will be reviewed by the SWG Proposal Committee. The proposal has minimum requirements that are specified on the ASA BIOP website (Scientific Working Group Proposal Committee - Biopharmaceutical Section (amstat.org)) and serves as the foundation for developing a charter. The proposal should be submitted to the chair of SWG Proposal Committee, who will then forward it to the SWG Proposal Committee for evaluation.

During the review, the committee will (1) assess whether the proposed SWG would benefit the ASA BIOP community, (2) ask for clarifications, and (3) suggest enhancements which usually involve establishing connections between the proposed SWG and other SWGs where there may be overlaps or synergies. After the review and comments have been addressed, the SWG Proposal Committee chair will forward the proposal to the ASA BIOP Executive Committee chair. The Executive Committee will then vote on whether to approve the proposed SWG. Either the Executive Committee chair or the SWG Proposal Committee chair will inform the applicant on the Executive Committee's decision.

Once an SWG is approved and established, there is an annual check-in at the beginning of each year when the annual health check form is filled out (link for annual health check). The SWG Proposal Committee chair (with help from other committee members as needed) collects the annual health check forms from the SWG chairs, and then forwards these forms to the Executive Committee chair to share as pre-reads for the Executive Committee spring meeting. During this meeting, the SWG Proposal Committee chair will present a summary of the health checks. A recent summary was concluded in April 2023.

In 2022, there were a total of nine active SWGs that had been previously established, along with the addition of one new SWG. As for 2023, two new SWGs have already been formed, and there are three additional potential SWGs currently undergoing review. On the following page is a summary of the existing SWGs and highlights of their recent accomplishments.

If any readers are interested in any of the following SWGs, please feel free to reach out to the respective chairs. Furthermore, if there is a specific area of statistical research that is not currently covered by the existing SWGs and readers are interested in establishing a new SWG, please reach out to the SWG Proposal Committee. They will be happy to assist you in initiating the process. ■

Scientific Working Group	Chairs	Recent highlights
Alzheimer's Disease	Paul Delmar	Members from pharma and academia, representing US, Europe, and Japan meet and discuss statistical methods for the analysis of Alzheimer's Disease clinical trials. An abstract recently submitted to the CTAD congress presents a method comparison based on re-analysis of recent phase 2 and 3 clinical trials.
Biostatistical Software Engineering	Daniel Sabanes Bove, Ya Wang	Grew to almost 40 members; built and rolled out R package 'mmrm'; authored a full day workshop on "Good Software Engineering Practice for R Packages" presented in Basel, Switzerland and Shanghai, China; started a video series on the ASA BIOP YouTube channel "Statistical Software Engineering 101".
Cell and Gene Therapy	Daniel Li, Alan Chiang	New SWG! March kick-off and SWG meets monthly. Three near term deliverables 1) white paper on challenges and lessons learned from CGT; 2) presentation and round table discussion at various conferences, e.g., RISW; 3) CGT short courses/tutorials in conferences.
Centralized Statistical Monitoring and Quality Tolerance Limits	Sue Talbot, Tim Rolfe	2023 activities: CSM methodology paper submitted (Feb); Duke Industry Stats Symposium 2023 Session: CSM tools in trial conduct (Mar); session at PSI conference on CSM/QTL (Jun); QTL methodology paper drafted and aiming for submission (May/Jun)
Estimands in Oncology	Kaspar Rufiback, Evgeny Degtyarev	SWG webpage (www.oncoestimand.org) is a resource to foster sharing of information and generated content such as presentations, event recordings (if available), and publications including a December 2022 webinar with the ASA NJ chapter entitled "Getting the question right – Applying the Estimand and Target Trial Frameworks with External Controls."
Health Technology Assessment (HTA)	Weili He, Min-Hua Jen, Cornelia Dunger-Baldauf	New SWG! Kick-off in the June timeframe with two workstreams: 1) RWE usage in HTA submissions, and 2) Innovative clinical trial study designs incorporating HTA requirements for regulatory purposes with or without use of RWD.
Nonclinical Biostatistics	Eve Pickering	Supports a bi-annual conference and twice-yearly leadership forum. Completed a white paper on the assessment of the p-value in the nonclinical pharmaceutical space
Real World Evidence	Jie Chen, Hana Lee, Yixin Fang	Completed three manuscripts in 2022 with two accepted by SBR website and the other under revision. Delivered presentations at JSM and Deming conference in 2022. Formed a public private partnership (PPP) with CDER/FDA in 2023 to support regulatory science.
Re-randomization of subjects	Yeh-Fong Chen	Working on a special issue of a journal for 2024. Major contents of the papers have been outlined and will be sent to the journal soon.
Safety	Jim Buchanan, Mengchun Li, Juergen Kubler	In 2022 made >35 presentations at the DIA, US/EU/China, JSM, the FDA/DIA statistics forum, Global Safety Congress, BASS, IISA, MBSW, Nonclinical Biostat (NCB), 25th PV conference and other conferences. Delivered 4 scientific webinars with participation by top experts from regulatory agencies
Statistical Methods in Oncology	Qi Jiang, Olga Marchenko	2022 accomplishments: 11 publications (2 online), 4 sessions at RISW, and 8 Open Forums in coordination with the Oncology Center of Excellence, US FDA, and other global regulatory agencies
Statistics in Pediatric Drug Development	Margaret Gamalo	Completed a special Issue within the Journal of Biopharmaceutical Statistics that will be published in June 2023

SUMMARY OF ASA BIOP SECTION'S VIRTUAL DISCUSSION WITH REGULATORS ON ASSESSING BIAS IN CANCER STUDIES WITH REAL-WORLD DATA ELEMENTS

Rajeshwari Sridhara (FDA), Olga Marchenko (Bayer), Qi Jiang (Seagen), Elizabeth Barksdale (LUNGevity Foundation), Yiyi Chen (Seagen), Paul Kluetz(FDA)

On October 13th of 2022, the American Statistical Association (ASA) Biopharmaceutical Section (BIOP) and LUNGevity Foundation hosted a virtual forum to discuss approaches for assessing bias in cancer studies with real-world data elements. This forum was part of a series conducted under the guidance of the U.S. FDA Oncology Center of Excellence's Project SignifiCanT (Statistics in Cancer Trials). The goal of Project SignifiCanT is to advance cancer drug development through collaboration and engagement among various stakeholders in the design and analysis of cancer clinical trials. The discussion was organized jointly by the ASA BIOP Statistical Methods in Oncology Scientific Working Group, the FDA Oncology Center of Excellence (OCE), and LUNGevity Foundation.

Randomized clinical trials remain the preferred clinical trial design to assess the benefits and risks of cancer treatments. However, in some clinical settings (or indications) it may be infeasible to conduct randomized clinical trials in a timely manner, particularly for the treatment of rare cancers. To accelerate cancer drug development, it may be necessary to use innovative clinical trial designs that may integrate use of real-world data elements. This open forum discussion session among various stakeholders focused on assessing bias in the estimation of treatment effect when real-world data (RWD) are used in the evaluation of treatment effect.

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 (Food and Drug Administration (FDA), Therapeutic Goods Administration (TGA), Health Canada (HC), European Medicines Agency (EMA), Medicines and Healthcare products Regulatory Agency (MHRA), Brazilian Health Regulatory Agency (ANVISA), Swissmedic (SMC)), clinical investigators, academicians, patient advocacy groups, and expert statisticians in industry. In addition, over 100 participants attended the virtual meeting, including representatives

Speakers/ Panelists:

Dr. Jie Chen (Overland Pharmaceuticals), Dr. Michael Coory (TGA, AU), Dr. Leonardo Costa (ANVISA, BR), Dr. Theodor Framke (EMA, EU), Dr. Boris Freidlin (NCI), Dr. Elizabeth Garrett-Mayer (ASCO), Mr. Lorenzo Hess (Swissmedic, CH), Dr. Qi Jiang (Seagen), Dr. Paul Kluetz (FDA), Dr. Nicole Li (Beigene), Dr. Olga Marchenko (Bayer), Dr. Pallavi Mishra-Kalyani (FDA), Dr. Richard Pazdur (FDA), Prof. Martin Posch (Center for Medical Statistics, Informatics, and Intelligent Systems at the Medical University of Vienna), Dr. Khadija Rantell (MHRA, UK), Mr. Andrew Raven (Health Canada), Dr. Donna Rivera (FDA), Dr. Sunhee Ro (Sierra Oncology), Dr. Rajeshwari Sridhara (FDA), Dr. Mark Stewart (Friends of Cancer Research), Mr. Andrew Thomson (EMA, EU), Dr. Jonathon Vallejo (FDA)

from other international regulatory agencies (Health Sciences Authority (HAS), Pharmaceutical Division Israel Ministry of Health)). 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, consultant from OCE, FDA.

The OCE leadership gave an introductory presentation and remarks that defined RWD and real-world evidence (RWE) to align the participants for discussion, and highlighted situations where selection bias, information bias, and immortal time bias may arise with clinical trials integrating RWD. The presentation concluded with approaches for panelists from academia, industry, and regulatory agencies to consider for the topic. This presentation was followed by two others representing patient advocates and industry.

The first speaker discussed Friends of Cancer Research RWE pilots 1.0 and 2.0. These studies aimed to establish methodologies for RWD evaluation and shared protocols across real-world data groups, assess the feasibility of identifying real-world endpoints for oncology, and to evaluate the consistency of these endpoints in characterizing the differences between therapies. Using data from different vendors on non-small cell lung cancer patients treated with checkpoint inhibitors, Pilot 1.0 concluded that it is possible to reach high-level alignment on important data elements and definitions for real-world endpoints in the context of a focused research question. Additionally, real-world endpoints may be effectively used in assessing treatment effect, although additional studies should be conducted to further characterize performance of the real-world endpoints and potential sources of variability, which are included in the ongoing real-world-response pilot effort.

The speaker from industry gave an overview of how bias can be introduced when RWD are used for oncology studies and how to conduct quantitative bias analyses using a four-step approach. This approach includes 1) identification of potential sources of bias; 2) selection of bias model; 3) quantification of bias, and 4) interpretation of bias analysis. The presenter also discussed the use of estimands and E-value for sensitivity analysis in assessing the bias. To conclude, examples of potential bias assessment on tumor assessment bias, immortal time bias, and treatment effect heterogeneity bias were provided.

The key points raised in the panel discussion following these presentations were:

- Potential efficiency gains from the use of hybrid randomized clinical trials integrating RWD are based on the assumption that external data are similar to those from the trial population. This assumption may not be accurate or assessable for small/rare patient populations. Use of adaptive randomization to adjust for dynamic borrowing may impact efficiency.
- Estimand framework is useful in assessing the bias and in better understanding the analysis assumption. Endpoints based on RWD should be as objective as possible.
- It is important to understand the RWD source(s) (e.g., electronic health records, claims data, etc.) as the types and quality of data they provide can be heterogenous and difficult to harmonize.
- RWD can be prospective and prospective data collection may be preferred to reduce methodological concerns for bias.
- E-value method can be useful and may be considered as a tipping point analysis for bias assessment.
- The appropriateness and utility of integrating RWD with clinical trials depends on data relevance and reliability and overall data quality as well as the clinical setting. The selection of high-quality fit-for-purpose RWD that appropriately answers the scientific questions of interest is challenging. Continued collaborative efforts are needed on this topic to better assess bias and its multifaceted effects on estimation of the treatment effect.

This forum provided an opportunity to have open scientific discussion among a diverse multidisciplinary stakeholder group – clinicians, epidemiologists, statisticians, patient advocates, international regulators, and representatives from pharmaceutical companies--focused on emerging statistical issues in cancer drug development. We plan to continue with similar multidisciplinary open forum discussions on a variety of important topics.

Acknowledgement

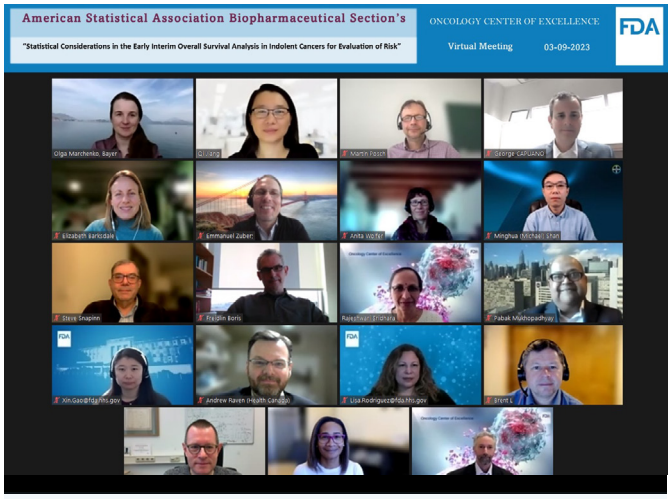
Authors thank Joan Todd (FDA) and Syed Shah (FDA) for technical support. ■

SUMMARY OF ASA BIOP SECTION'S VIRTUAL DISCUSSION WITH REGULATORS ON STATISTICAL CONSIDERATIONS IN THE EARLY INTERIM OVERALL SURVIVAL ANALYSIS IN INDOLENT CANCERS FOR EVALUATION OF RISK

Rajeshwari Sridhara (FDA), Olga Marchenko (Bayer), Qi Jiang (Seagen), Elizabeth Barksdale (LUNGevity Foundation), Yiyi Chen (Seagen), Marc Theoret (FDA)

On March 9, 2023, the American Statistical Association (ASA) Biopharmaceutical Section (BIOP) and LUNGevity Foundation hosted a virtual forum to discuss statistical considerations in the early interim overall survival analysis in indolent cancers for evaluation of risk. This forum was part of a series conducted under the guidance of the U.S. FDA Oncology Center of Excellence's Project SignifiCanT (Statistics in Cancer Trials). The goal of Project SignifiCanT is to advance cancer drug development through collaboration and engagement among various stakeholders in the design and analysis of cancer clinical trials. The discussion was organized jointly by the ASA BIOP Statistical Methods in Oncology Scientific Working Group, the FDA Oncology Center of Excellence (OCE), and LUNGevity Foundation.

Overall survival (OS) is an important endpoint in all randomized cancer clinical trials as it measures both efficacy and safety. However, progression-free survival (PFS) is generally considered as the primary endpoint for regulatory decision making in indolent/chronic cancers such as hematologic malignancies, and OS is evaluated as a secondary endpoint. Given the long course of such chronic diseases, the number of events for the OS analysis is often limited at the time of the final analysis of PFS. In recent times, in trials evaluating PI3K inhibitors, early OS analyses have suggested possibility of a detrimental effect of the investigational drugs (Oncologic Drugs Advisory Committee



Speakers/ Panelists:

Dr. Elizabeth Barksdale (LUNGevity Foundation), Dr. George Capuano (Johnson & Johnson), Dr. Michael Coory (TGA, AU), Dr. Leonardo Filho (ANVIS, BR), Dr. Boris Freidlin (NCI), Prof. Tim Friede (University Medical Center Göttingen), Dr. Xin (Cindy) Gao (FDA), Dr. Nicole Gormley (FDA), Dr. Qi Jiang (Seagen), Prof. Brent Logan (Medical College of Wisconsin), Dr. Olga Marchenko (Bayer), Dr. Pabak Mukhopadhyay (AstraZeneca), Dr. Richard Pazdur (FDA), Prof. Martin Posch (Center for Medical Statistics, Informatics, and Intelligent Systems at the Medical University of Vienna), Dr. Khadija Rerhou Rantell (MHRA, UK), Mr. Andrew Raven (Health Canada), Dr. Lisa Rodriguez (FDA), Dr. Minghua (Michael) Shan (Bayer), Dr. Steven Snapinn (Statistical Consultant), Dr. Rajeshwari Sridhara (FDA), Dr. Marc Theoret (FDA), Dr. Anita Wolfer (Swissmedic), Dr. Emmanuel Zuber (Novartis).

(ODAC) meeting, April 21, 2022; *https://www.fda.gov/advisory-committees/advisory-committee-calendar/updated-information-april-21-22-2022-meeting-oncologic-drugs-advisory-committee-meeting-announcement*). Thus, it is important to include a pre-specified plan to conduct early OS analysis at the time of PFS analysis, understanding the uncertainty of the early results, in future indolent cancer trials irrespective of the drug class. This open forum discussion among multi-disciplinary experts focused on exploring the possibility of pre-specifying criteria for unacceptable risk for early OS analysis in indolent cancer trials.

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 (Food and Drug Administration (FDA), Health Canada (HC), Medicines and Healthcare products Regulatory Agency (MHRA), Swissmedic (SMC), and Brazilian Health Regulatory Agency (ANVISA)), clinical investigators, academicians, patient advocacy groups, and expert statisticians in industry. In addition, over 100 participants attended the virtual meeting, including representatives from other international regulatory agencies (European Medicines Agency (EMA), Pharmaceuticals and Medical Devices Agency (PMDA), Health Sciences Authority (HAS), Israel Ministry of Health (MOH)). The discussions were moderated by the BIOP Statistical Methods in Oncology Scientific Working Group co-chairs, Dr. Olga Marchenko from Bayer and Dr. Qi Jiang from Seagen; Dr. Elizabeth Barksdale from LUNGevity Foundation; and Dr. Rajeshwari Sridhara, consultant from OCE, FDA.

In the introductory presentation, the OCE leadership reviewed multiple randomized PI3K inhibitor trials with potential detrimental OS effects at the time of primary analysis of PFS (*ODAC meeting on April 21, 2022*), and highlighted criteria specified in 2008 FDA guidance for evaluating cardiovascular risk in new antidiabetic therapies to treat type 2 diabetes. Panelists were requested to discuss the feasibility, benefits, and challenges of setting similar criteria for detrimental OS effect at early analysis for indolent cancer trials.

The first speaker, from industry, reviewed two approaches for early interim OS assessment in trials with PFS as the primary efficacy endpoint in indolent cancers. The 2-outcome procedure takes a non-inferiority trial approach, specifying a required number of OS events with pre-set criteria to determine if there is an OS concern at the early analysis. The 3-outcome procedure reduces the required sample size by adding a "gray zone" to the potential outcome. The trade-off is that the trial will require more OS events before a detrimental OS effect is determined if the outcome falls in the gray zone.

The second speaker, also from industry, pointed out that OS represents the primary clinical interest for oncology trials, therefore, a detrimental effect in OS should negate an efficacy effect in PFS. The potential of having non-proportional hazard at pre-progression and post-progression adds challenges in evaluating OS detrimental effect at an early timepoint.

The key points raised in the panel discussion following these presentations were:

- Pre-set criteria for detrimental OS effect at early analysis timepoint is feasible and desired. This will provide more rigor to the assessment and clarify the expectation, including the necessary amount and relevant sources of information.
- A pre-planned early OS analysis at the time of PFS analysis by allocating a small fraction of alpha for the OS analysis is recommended.
- Statistical modelling and comprehensive simulation studies may be utilized in setting up the criteria. Criteria should allow to quantify the uncertainty of clinically relevant effects. Transparency regarding the assumptions used in the simulation studies is required.
- The existing guidance on the diabetes is interesting and relevant, but unique challenges with indolent oncologic disease trials should be considered when setting up the criteria (e.g., small sample size, low mortality rate, OS is the ultimate clinical benefit endpoint, OS hazard ratio may vary over time, patients might be allowed to switch treatment at the time of progression).

- The definition of "detrimental effect" depends on factors such as type of cancer, length of follow-up, length of expected survival, line of therapy, and the patient population. There may not be one-size-fits-all criteria. Multidisciplinary team including clinicians, statisticians, regulators and patients should be involved in setting the criteria. Close collaboration and communication with regulatory agencies are critical and should start early at the design stage.
- Unless the experimental agent is approved for use in the subsequent line of therapy, designs that incorporate crossover from the control arm to the experimental agent (at progression) confound OS effect and compromise the study ability to assess safety.
- Defined criteria should aim at supporting a better-informed benefit-risk assessment.
- The regulatory decision based on the early analysis generally considers a comprehensive evaluation of available information such as treatment effect in PFS, quality of life, crossover effect, toxicity profile including individual case reviews of causes of death, etc.

This forum provided an opportunity to have open scientific discussion among a diverse, multidisciplinary stakeholder group – clinicians, statisticians, patient advocates, international regulators, and representatives from pharmaceutical companies-- focused on emerging statistical issues in cancer drug development.

Acknowledgement

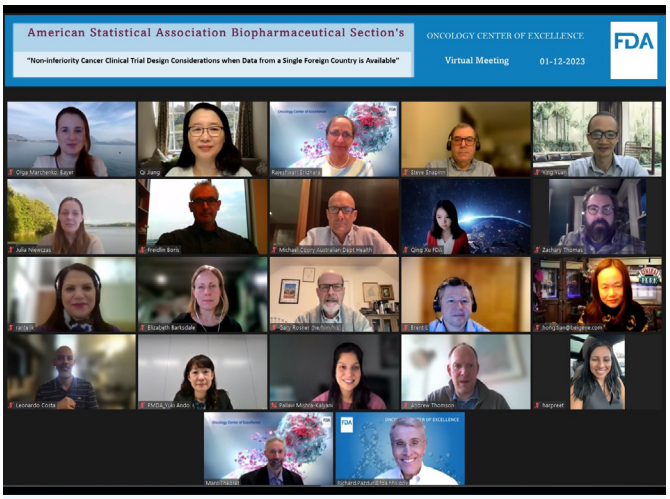
Authors thank Joan Todd (FDA) and Syed Shah (FDA) for technical support. ■

SUMMARY OF ASA BIOP SECTION'S VIRTUAL DISCUSSION WITH REGULATORS ON NON-INFERIORITY CANCER CLINICAL TRIAL DESIGN CONSIDERATIONS WHEN DATA FROM A SINGLE FOREIGN COUNTRY IS AVAILABLE

Rajeshwari Sridhara (FDA), Olga Marchenko (Bayer), Qi Jiang (Seagen), Elizabeth Barksdale (LUNGevity Foundation), Yiyi Chen (Seagen), Marc Theoret (FDA)

On January 12, 2023, the American Statistical Association (ASA) Biopharmaceutical Section (BIOP) and LUNGevity Foundation hosted a virtual forum to discuss non-inferiority cancer clinical trial design considerations when data from a single foreign country is available. This forum was part of a series conducted under the guidance of the U.S. FDA Oncology Center of Excellence's Project SignifiCanT (Statistics in Cancer Trials). The goal of Project SignifiCanT is to advance cancer drug development through collaboration and engagement among various stakeholders in the design and analysis of cancer clinical trials. The discussion was organized jointly by the ASA BIOP Statistical Methods in Oncology Scientific Working Group, the FDA Oncology Center of Excellence (OCE), and LUNGevity Foundation.

This open forum discussion among multi-disciplinary experts focused on potential non-inferiority (NI) cancer clinical trial designs, specifically innovative designs including Bayesian methodology and use of available information from clinical trials conducted completely outside of the regulatory region for marketing authorization. This discussion was a continuation of the discussion held in October 2021, on the acceptability of data from a single foreign country for regulatory decision making. In the prior discussion it was recognized that multi regional clinical trials (MRCT) can increase efficiency of drug development and aide in extrapolation of treatment effect in diverse populations and regions by providing information on the intrinsic and /or extrinsic factors that may impact responses to the drug (ICH E17). Burden of proof on the sponsor when seeking regulatory authorization using data from a randomized clinical trial conducted in a single country is very high. For example, the randomized cancer trials (RCT) conducted in a single country may have used a control arm that is unlikely to be used in another country where the sponsor is seeking marketing authorization.



Speakers/ Panelists:

Dr. Yuki Ando (PMDA, JP), Dr. Elizabeth Barksdale (LUNGevity Foundation), Dr. Michael Coory (TGA, AU), Dr. Leonardo Filho (ANVIS, BR), Dr. Boris Freidlin (NCI), Dr. Qi Jiang (Seagen), Prof. Brent Logan (MSW), Dr. Olga Marchenko (Bayer), Dr. Pallavi Mishra-Kalyani (FDA), Dr. Julia Niewczas (Janssen Pharmaceutical Company of J&J), Dr. Richard Pazdur (FDA), Dr. Khadija Rerhou Rantell (MHRA, UK), Mr. Andrew Raven (Health Canada), Prof. Gary Rosner (Johns Hopkins University, School of Medicine), Dr. Harpreet Singh (FDA), Dr. Steven Snapinn (Statistical Consultant), Dr. Rajeshwari Sridhara (FDA), Dr. Marc Theoret (FDA), Dr. Zachary Thomas (Eli Lilly), Mr. Andrew Thomson (EMA), Dr. Hong Tian (BeiGene), Dr. Qing Xu (FDA), Prof. Ying Yuan (MD Anderson Cancer Center).

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 (Food and Drug Administration (FDA), Therapeutic Goods Administration (TGA), Health Canada (HC), European Medicines Agency (EMA), Medicines and Healthcare products Regulatory Agency (MHRA), Brazilian Health Regulatory Agency (ANVISA), and Pharmaceuticals and Medical Devices Agency (PMDA)), clinical investigators, academicians, patient advocacy groups, and expert statisticians in industry. In addition, over 100 participants attended the virtual meeting, including representatives from other international regulatory agencies (Health Sciences Authority (HAS), Swissmedic (SMC), Ministry of Health, Israel). 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, consultant from OCE, FDA.

The OCE leadership gave an introductory presentation that reviewed the October 2021 discussion and provided a hypothetical scenario where a NI trial needs to be conducted to support regulatory decision in the country where marketing authorization is being sought for a drug demonstrated efficacy only in a single country RCT. Panelists were requested to discuss what is considered a MRCT and various statistical considerations in designing NI trials.

The first speaker, from industry, reviewed the FDA NI trial guidance document and discussed the choice of non-inferiority margins. The three key assumptions (assay sensitivity, constancy of effect, and historical evidence of sensitivity to drug effects) in NI trials were highlighted and factors that impact these assumptions were elaborated. The presenter concluded with some comments on the hypothetical scenario presented by the FDA.

The second speaker, also from industry, focused on Bayesian methods and considerations for borrowing on control arm or treatment arm, and borrowing of treatment effect for bridging of single region studies. Potential methods for borrowing and the corresponding assumptions and potential problems were discussed under the hypothetical scenario. The presenter concluded that borrowing on the control could be straightforward for well-established standard of care treatments (SoCs) in the US without compromising the safety profile of the investigational drug.

The third speaker, from academia, further illustrated approaches (for example, regression model approach, informative prior approach) and challenges to account for intrinsic and extrinsic factors in the hypothetical scenario provided by OCE. Dynamic borrowing was recommended to allow more flexibility in determining the level of borrowing.

The key points raised in the panel discussion following these presentations were:

- Single-country trials are challenging to generalize to the patient population of interest due to differences in population, SoC/available control treatment, and endpoints. It is important to understand potential regional differences in these factors to support the design of the NI trial.
- Borrowing relevant information from outside trials may boost the power of a trial if exchangeability assumption holds. Otherwise, borrowing may inflate the type I error rate; if inflation of the type I error is considered acceptable then instead of borrowing information one can use a regular frequentist design with a higher type I error rate.
- The use of PK/PD models may help inform the design of bridging studies.
- "MRCT" is not just defined by geography. It also depends on other factors such as socioeconomic conditions, underlying biology, disease prevalence, and SoC. A MRCT may include heterogeneous patient populations from multiple regions with different biology and other factors (such as socio-economic conditions, standard clinical practice of the country).
- When lacking confidence in the generalizability of single-country trials and the assumptions (e.g., constancy assumption, assay sensitivity assumption) made in the NI trial, specifying a very conservative NI margin or even a superiority trial may be considered. A clinically meaningful and adequately justified NI margin is critical to ensure success of the NI trial.

This forum provided an opportunity to have open scientific discussion among a diverse multidisciplinary stakeholder group – clinicians, epidemiologists, statisticians, patient advocates, international regulators, and representatives from pharmaceutical companies--focused on emerging statistical issues in cancer drug development.

Acknowledgement:

Authors thank Joan Todd (FDA) and Syed Shah (FDA) for technical support. ■

SUMMARY OF ASA BIOP SECTION'S VIRTUAL DISCUSSION WITH REGULATORS ON CONSIDERATIONS OF INNOVATIVE CANCER CLINICAL TRIAL DESIGNS FOR POST-MARKET DOSE-OPTIMIZATION STUDIES

Rajeshwari Sridhara (FDA), Olga Marchenko (Bayer), Qi Jiang (Seagen), Elizabeth Barksdale (LUNGevity Foundation), Yiyi Chen (Seagen), Marc Theoret (FDA)

On November 10th of 2022, the American Statistical Association (ASA) Biopharmaceutical Section (BIOP) and LUNGevity Foundation hosted a virtual forum to discuss considerations of innovative cancer clinical trial designs for post-market dose-optimization studies. This forum was part of a series conducted under the guidance of the U.S. Food and Drug Administration (FDA) Oncology Center of Excellence's Project SignifiCanT (Statistics in Cancer Trials). The goal of Project SignifiCanT is to advance cancer drug development through collaboration and engagement among various stakeholders in the design and analysis of cancer clinical trials. The discussion was organized jointly by the ASA BIOP Statistical Methods in Oncology Scientific Working Group, the FDA Oncology Center of Excellence (OCE), and LUNGevity Foundation.

This discussion was a continuation of three earlier discussions on trial designs for dose optimization. It has been acknowledged that dose-finding studies typically assess dose-limiting toxicity in a small cohort of patients in the first cycle of treatment and not in subsequent cycles, and do not adequately assess tolerability and adherence to drug product. Ideally, dose optimization studies should be conducted in the pre-market setting so that the recommended dosage of the marketed product maximizes efficacy and minimizes toxicity. However, it is not uncommon that a dose has not been optimized at the time of marketing approval, and dose-optimization studies are needed after the drug

Speakers/ Panelists:

Dr. Elizabeth Barksdale (LUNGevity Foundation), Dr. Neby Bekele (Exelixis), Dr. Mark van Bussel (Medicines Evaluation Board, Netherlands), Dr. Arunava Chakravartty (Novartis), Dr. Joyce Chen (FDA), Dr. Michael Coory (TGA, AU), Dr. Leonardo Filho (ANVISA, BR), Dr. Theodor Framke (EMA.EU), Dr. Boris Freidlin (NCI), Dr. Elizabeth Garrett-Mayer (ASCO), Dr. Nahor Haddish-Berhane (JNJ), Dr. Brian Heiss (FDA), Dr. Larissa Higgins (HPRA, Ireland), Dr. Qi Jiang (Seagen), Dr. Olga Marchenko (Bayer), Dr. Richard Pazdur (FDA), Dr. Atik Rahman (FDA), Dr. Khadija Rantell (MHRA, UK), Mr. Andrew Raven (Health Canada), Prof. Gary Rosner (Johns Hopkins University, School of Medicine), Dr. Mirat Shah (FDA), Dr. Rajeshwari Sridhara (FDA), Dr. Marc Theoret (FDA), Dr. Jonathon Vallejo (FDA), Dr. Jing Xu (Takeda), Prof. Ying Yuan (MD Anderson Cancer Center).

has been approved. This open forum discussion among multi-disciplinary experts explored innovative clinical trial design options for post-market dose-optimization studies for cancer therapies.

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, Therapeutic Goods Administration (TGA), Medicines and Healthcare products Regulatory Agency (MHRA), Health Canada (HC), European Medicines Agency (EMA), Brazilian Health Regulatory Agency (ANVISA), Medical Evaluation Board (MEB, Netherlands), Health products Regulatory Authority (HPRA, Ireland)), clinical investigators, academicians, patient advocacy groups, and expert statisticians in industry. In addition, over 100 participants attended the virtual meeting, including representatives from other international regulatory agencies (Pharmaceuticals and Medical Devices Agency (PMDA), Health Sciences Authority (HAS), Pharmaceutical Division Israel Ministry of Health). The discussions were moderated by the BIOP Statistical Methods in Oncology Scientific Working Group co-chairs, Dr. Olga Marchenko from Bayer and Dr. Qi Jiang from Seagen; Dr. Elizabeth Barksdale from LUNGevity Foundation; and Dr. Rajeshwari Sridhara, consultant from OCE, FDA.

The OCE leadership provided background on the topic and considerations for post-marketing dose optimization studies for approved drugs of which efficacy has been established with acceptable safety. The studies should allow comprehensive assessment of tolerability and efficacy to show superior tolerability and non-inferior efficacy at the optimal dose compared to the approved dose. Potential designs for such studies include randomized clinical trials (RCTs) comparing two or more doses with approved dose as the control arm; single-arm trials of a lower dose compared to external control; and other designs such as pragmatic randomized trials, platform trials, etc. Bayesian methods can be incorporated into some designs as appropriate.

A speaker representing the FDA discussed how to improve dose optimization by considering the totality of PK, PD, efficacy, safety and tolerability. Two post-market dose optimization trial designs were presented as examples (PROSELICA trial evaluating optimal dose of cabazitaxel and CA180-034 trial evaluating optimal dose of dasatinib) that resulted in dosage change in the product label. The first presenter from industry used case study of ribociclib from a post-marketing requirement study of MONALESSA 2's approval, evaluating if a lower dose of the drug could reduce the dose dependent toxicities and demonstrate a non-inferior efficacy compared to the approved dose. The presenter also discussed estimand considerations and the strategies for intercurrent events in post marketing non-inferiority studies. The second presenter from industry discussed the OPTIC trial, a post marketing dose optimization trial for ponatinib for treatment of chronic myeloid leukemia. The drug was reapproved with the recommended dose changed to start with a 45 mg QD and reduce to 15 mg QD after achieving a molecular response.

The key points raised in the panel discussion following these presentations were:

- Use of a single endpoint combined with a composite endpoint may be considered for post-market dose optimization studies. A composite endpoint provides more information for decision making whereas a single endpoint provides information of interest (i.e., risks) to patients.
- A comprehensive measure of tolerability may include dose reduction, dose interruption, dose discontinuation, reduced compliance and all relevant adverse events. Patient reported outcomes, assessed using validated tools, are important endpoints in evaluating tolerability.
- Time-to-treatment discontinuation is another endpoint that was discussed for dose optimization studies as it could potentially reflect important factors such as tolerability, adverse events, quality of life and clinical efficacy.
- "Optimal" dose or schedule may vary based on patient population and disease stage.
- PD/PK modeling can be useful tools to identify the potential optimal dose.

- PROSELICA trial provides a cautionary tale for the use of biomarkers in deciding whether a reduced dose preserves efficacy: PSA-based endpoints suggested inferiority of 20-mg dose, yet the study conclusively established OS noninferiority for 20-mg vs 25-mg dose.
- Innovative adaptive designs should be considered for post-market dose optimization studies. Bayesian methods may be used to provide estimation on the treatment effect of different dose levels.
- A non-inferiority trial may be appropriate in certain circumstances. Although a non-inferiority trial tends to focus on efficacy endpoints, a mechanism to formally incorporate evidence on safety and tolerability is desired.
- Discussions with regulatory bodies early in the drug development process is important in designing and conducting adequate dose-optimization studies, with particular consideration for the type of compound under investigation.

This forum provided an opportunity to have open scientific discussion among a diverse multidisciplinary stakeholder group – clinicians, statisticians, patient advocates, international regulators, and representatives from pharmaceutical companies-- focused on emerging statistical issues in cancer drug development.

Acknowledgement

Authors thank Joan Todd (FDA) and Syed Shah (FDA) for supporting the forum. ■

THE ASA BIOPHARMACEUTICAL SECTION REGULATORY-INDUSTRY STATISTICS WORKSHOP

September 27–29, 2023 • Rockville, MD

Statistical Thinking and Innovation with Global Impact

Dear BIOP colleagues,

We are delighted to invite and welcome you to the 2023 ASA BIOP Regulatory Industry Statistics Workshop (RISW), to be held on September 27-29 in Rockville, Maryland. Our 2023 RISW Theme will be *Statistical Thinking and Innovation with Global Impact.* We are thrilled of the many great events that highlight the workshop's theme and demonstrate the close collaboration as well as strength of the BIOP community.

Since its modest inception in the 1990s as a workshop organized by the FDA Statistical Association specially for the FDA statisticians, the RISW has grown and evolved into the preeminent forum for meaningful discussions between regulatory and industry statisticians. Today, many statisticians in these areas consider the RISW as a must-see annual event, and look forward to the sharing of ideas and pushing the boundaries of statistical science to enhance medical product development.

Building from the successful in-person 2022 workshop, the 2023 workshop has added new initiatives, expanded on the number of parallel sessions, and organized two plenary sessions that reflects the importance of critical thinking and innovative work that we do at a global scale.

This year we are expecting over 1,000 statisticians from across regulatory, industry and academics to join. The venue will be at the Bethesda North Marriott Hotel and Conference Center. Early registration has been open since June 15, 2023, and we have already seen a steady stream of registrations. Based upon early results, as well as a record number of parallel sessions, short courses, and poster proposals, we expect the workshop to sell out and enthusiasm to be high. Below are some highlights:

- 10 short courses
- 2 plenary sessions
- 49 Parallel sessions
- 52 Round Table sessions
- 45 Poster sessions

Short Courses

On day 1, we will kick off the workshop with 10 outstanding half-day short courses. These courses will be ticketed events that can be added to your registration for in-depth learning on hot topics such as causal inference, dose optimization, master protocols, RWD, Bayesian methods, covariate adjustments, diagnostic devices, and pediatric extrapolation. One important highlight this year are the 3 courses taught by academic and NCI statisticians: a course on Bayesian sample size estimation by Joe Ibrahim at UNC, a course on Absolute Risk Prediction by NCI statistician Ruth Pfeiffer, and the final one on oncology dose finding by Yuan Ji at the University of Chicago. Registration for each short course is $115 during early registration, which is a bargain for the insights provided by these world-class instructors.

Plenary Sessions

On the morning of the second day of the workshop, we will offer two plenary sessions where we bring in diverse representation from academic, industry, and regulatory panelist and speakers across the globe. The first session titled "Statistical Influence and Opportunities on the International Harmonization of Drug Development" will feature Frank Bretz (Novartis) and John Scott (FDA) as keynote speakers. Amy Xia (Amgen) will moderate this plenary session and will be joined by three esteemed panelists Yuki Ando (PMDA), Lisa LaVange (UNC), and Frank Petavy (EMA). In this session, speakers and panelists will highlight the need of harmonization on technical topics such as innovative designs as well as global initiatives such as simultaneous drug development and harmonization of regulatory feedback. Adoption of innovative clinical trials will be limited without a harmonized perspective from drug regulatory agencies, especially for confirmatory studies in multi-regional drug development programs.

The second plenary session titled "Digital Innovation and Artificial Intelligence: Outlook and Trends" will feature Jennifer Goldsack (Digital Medicine Society) and Troy Tazbaz (FDA) as keynote speakers. Kelly Zou (Viatris) will be moderating this session. Yuki Ando (PMDA), Sandeep Menon (Pfizer), Vinay Pai (FDA), Frank Petavy (EMA) will join as panelists to provide their thoughts on the global challenges and opportunities for digital health through patient-centricity. In this session, several key topics will be discussed: (1) a proliferation of artificial intelligence (AI) applications; (2) digital innovations and examples in the healthcare and pharma industries, particularly for regulatory purposes; (3) digital tools to accelerate, complement and be integrated into trials.

Parallel Sessions

The parallel sessions are organized to cover topics at high priorities and across all regulated areas such as clinical trial conduct/analyses, clinical trial design (e.g. complex/innovative design), digital health, big data, estimands and missing data, real-world evidence, multiplicity, causal inference, and time to event analysis. They are also organized to cover topics of specific interests such as decision analyses, vaccines, gene and cellular therapies, diagnostics, oncology products, small populations, and leadership. The parallel sessions will bring both FDA and industry speakers into each session, so that the highly valued original intent of the workshop is maintained.

Roundtables

Roundtable discussions have consistently been a highlight of the workshop, and this year we have enhanced the experience by expanding and scheduling the roundtables on both days 2 and 3 of the event. By doing this, we aim to provide participants with increased flexibility and additional networking opportunities. We also encourage roundtable organizers to "keep the conversation going" with the participants and perhaps consider submitting a parallel session to future workshops.

Poster Sessions

The poster sessions will take place on days 2 and 3 of the workshop. Poster sessions have been popular as they provide an opportunity for papers to be presented in greater visual detail and allow presenters to engage in lively discussions with attendees. The workshop will display 45 posters. Ten student travel awards will be granted, and three best posters will be recognized and selected by the steering committee.

Thank You

A successful workshop relies heavily on careful planning and hard-working of many people involved. The RISW Steering Committee members and advisors has participated in multiple meetings and various tasks since the kick-off meeting in September of 2022. We are indebted to the 2023 Steering Committee members and advisors. We are also grateful for the support and guidance from the ASA Meetings staff as well as ASA Biopharmaceutical Section Executive Committee.

For more information on the workshop including registration details, please visit the conference website at *https://ww2.amstat.org/meetings/biop/2023/workshopinfo.cfm*.

We hope to see you there!

Sincerely,

Erik Bloomquist, *FDA*

Fanni Natanegara, *Eli Lilly and Company*

2023 ASA BIOP RISW Co-Chairs

UPCOMING CONFERENCES

2023 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 will be held from September 27 to 29, 2023 in Rockville Maryland, with invited sessions co-chaired by statisticians from industry, academia, and the FDA. Short courses on related topics will be offered on the first day of the workshop. To find out more details, please visit: *https://ww2.amstat.org/meetings/biop/2023/*.

- **Early Registration Opens:** June 15, 2023
- **Early Registration Closes:** August 16, 2023

2023 Women in Statistics and Data Science

The Women in Statistics and Data Science (WSDS) Conference is sponsored by the ASA. It will be held from October 25 to 27 in Bellevue, Washington. WSDS will gather professionals and students from academia, industry and the government working in statistics and data science. For more details, please visit: *https://ww2.amstat.org/meetings/wsds/2023/*.

- **Early Registration Deadline:** August 24, 2023

The 79th Annual Deming Conference on Applied Statistics

The Deming conference is sponsored by the ASA Biopharm Section. It will be held on December 4-8, 2023 in Philadelphia, PA. It consists of 3 days of Tutorials and 2 days of Short Courses on Applied Statistics, aimed at providing a learning experience on recent developments in statistical methodologies in biopharmaceutical applications. The first 3 days of the conference is composed of twelve three-hour tutorials on current topics in applied biopharmaceutical statistics and FDA regulations, and a one-hour distinguished keynote speaker on each of the 3 days of the conference. The last 2 days of the conference consist of short courses on special topics that will offer in-depth review of theory and practical considerations. For more details, please visit https://demingconference.org/.

- **Registration Opens:** August 14, 2023