Volume 30, No. 1 • Spring 2023

BIOPHARMACEUTICAL REPORT

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

CONTENTS

Notes from the editors

Warmest greetings from us and hope everyone is doing well! With spring finally here, and life returning to even more normal now as compared to last year, most of us have had a chance to spend more time with friends and family, go to in-person gatherings and meetings such as JSM and/or RISW last year, or travel around on vacations. These connections are important to everyone. For this spring issue of the Biopharmaceutical Report, we have prepared some very interesting topics. There have been discussions in the last few years on areas such as clinical trial diversity, and recently regulatory agencies have published guidance on this topic. In the meantime, there is continued interest in artificial intelligence and machine learning (AI/ML)'s role in clinical trials. The theme of this spring issue is "*Reaching for the top in clinical trial excellence*", with highlights on clinical trial diversity and use of AI/ML to advance clinical trials operations or methodology.

In this issue, we open with a joint article from our current Session Chair, **Brian Millen**, and the 2022 BIOP Chair, **Alan Hartford**, discussing achievements of the section from last year and laying out plans for 2023. For the featured articles, we start with two articles from Tufts Center for the Study of Drug Development on the topic of clinical trial diversity. The first one, written by **Zachary Smith** (Tufts), **Emily Botto** (Tufts) and **Kenneth Getz** (Tufts), examined racial and ethnic disparities in registration-enabling trials for FDA or EMEA approved drugs. On a similar topic, **Kenneth Getz** (Tufts) looks into the diversity of investigators site personnel and how that correlates with the diversity in patient enrollment in clinical trials. The findings could help clinical trialists to think about better trial implementations and may spark ideas among statisticians and quantitative researchers to find better trial designs to increase clinical trial diversity. Next featured article is by **Jianmin Chen** (University of Connecticut), **Zhaowei Hua** (Servier), **Qian Meng** (Servier), **Lucie Truffaut-Chalet** (Servier), **Zhaoyang Teng** (Servier) and **Rui Tang** (Servier). They introduced a new digital tool using AI/ML that can be used to predict patient recruitment in multi-center clinical trials. Our fourth featured article reviewed machine learning based population selection and enrichment in drug development, contributed by **Natalia Muehlemann** (Cytel). We close the featured articles section with the non-clinical paper by **Stan Altan** (Janssen R&D LLC), **Dave LeBlond** (Robert Singer Consulting) and **Tim Schofield** (CMC Sciences LLC). They brought us a new framework to achieve quality in drug and vaccine development. On the topic of leadership development, we have the pleasure of inviting the Biopharmaceutical Statistics Leaders Consortium (BSLC) to contribute a discussion article on the topic of effective collaboration between statisticians and data scientists in biopharmaceutical research and development, with authors **Erik Pulkstenis** (Abbvie), **Xun Chen** (Sanofi), **Simon Davies** (Takeda), **Pameljit Kalra** (ZS), **Becky Maksimovic** (ZS), **Christopher Miller** (Astrazeneca), **Jonathan Rowe** (ZS) and **Amy Xia** (Amgen). Many statisticians and data scientists would be able to benefit from their perspectives. Later in the issue, you can see summary of five 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 discussions are: 1) Cancer clinical trial design considerations when accepting foreign data from a single country, 2) Design considerations in evaluating treatment effect in marker negative population, 3) Impact on type I error with unplanned analyses in cancer clinical trials, 4) Evaluation and interpretation of interim overall survival results from randomized cancer clinical trials in chronic diseases, 5) Considerations in the evaluation of progression free survival with informative censoring in cancer clinical trials. In the last section, we first would like to congratulate the BIOP members who have been elected as ASA fellows in 2022! We are very proud of your great contributions to the statistical community. 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 first issue of BIOP report possible!

TRANSITION REPORT

Alan Hartford (BIOP Chair, 2022) and Brian Millen (BIOP Chair, 2023)

Hello BIOP Section! We hope you and your families are well and happy as we move rapidly along in 2023. Thank you for allowing us the opportunity to serve the membership of the Biopharmaceutical Section (BIOP). In this report, we'd like to update you on BIOP's progress over the past year and fill you in on plans for 2023.

We finally broke through the fog, pain, and isolation of the pandemic in 2022! It was great to see everyone that was able to attend JSM, our ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop (RISW), and other statistical meetings. These events felt a bit like a family reunion after some time being unable to meet in person.

Because our virtual meetings were successful prior to 2022, we investigated the possibility of making BIOP functions hybrid. However, the cost of hosting many sessions, each with its own A/V and tech support, is still prohibitive at the current time. We also need to protect the natural advantages of meeting in person. But we hope we can someday make use of further advanced tech to test a new approach for our meetings.

In 2022, the BIOP Executive Committee (EC) met virtually in the Spring and in person at JSM and at RISW. Our EC is made up of 12 elected members and 6 appointed at large members, all 18 of whom are voting members. The EC also has 17 subcommittees (each with its own chair and members) and 12 scientific working groups. We also have had a BIOP task force, the 40th Anniversary Planning Committee. More on their effort and accomplishments coming up.

Biopharmaceutical Section Scholarship

The Biopharmaceutical Section scholarships were awarded again in 2022. Consideration for the awards is based primarily on notable academic achievement or applied project work related to the area of biopharmaceutical statistics. General academic performance, leadership, volunteering, and service are also considered. The 2022 recipients were:

- Larry Han (Harvard University)
- Sasha Kravets (University of Illinois at Chicago)
- Chenqi "Stacy" Fu (College of Medicine Penn State Hershey)
- Lianlian Du (University of Wisconsin-Madison)

Thank you to Guochen Song (Chair of Committee 2022), Jared Lunceford (Chair of Committee 2023) and Wenting Cheng for facilitating this scholarship selection process. For more information, please go to: Scholarship Award - Biopharmaceutical Section (amstat.org).

Student Paper Award at Joint Statistical Meetings

First: **Junyi Zhou**, Indiana University on *A New Clustering Method for Longitudinal Data*

Second: **Tian Gu**, Harvard University on *Regression Inference for Multiple Populations by Integrating Summary-Level Data Using Stacked Imputations*

Third: **Haixu Ma**, University of North Carolina on *Learning Optimal Group-Structured Individualized Treatment Rules with Many Treatments*

Honorable Mention: **Lillian Haine**, University of Minnesota on *Semi-Supervised Mixture Multi-Source Exchangeability Model Approach for Incorporating Real World Data into Randomized Controlled Trial Analyses*

Honorable Mention: **Xiaohan Chi**, Shanghai Jiao Tong University on *BOB: Bayesian Optimal Design for Biosimilar Trials with Co-Primary Endpoints*

Thank you to Lanju Zhang (Chair of Committee), Yang Chen, Brenda Kurland, Jianchang Lin, Meijing Wu, and Yu Du for facilitating this award selection process.

Other Awards

Generally, we would have acknowledged Best Contributed Paper and Best Contributed Poster awards based on 2021 presentations at JSM. Because JSM was not held in person in 2021, there are now 2022 awards to announce. Stay tuned for announcements at the 2023 JSM BIOP business meeting.

Biopharmaceutical Section 40th Anniversary

The Biopharmaceutical Section celebrated its 40th anniversary as a section in person in 2022 at both JSM and RISW. The celebration was originally planned for 2021, the true year for the 40th. In 2021, we were able to celebrate with a wonderful virtual panel session with past BIOP chairs as part of the online 2021 JSM program. There were also several articles included in 2021 Biopharmaceutical reports (Summer & Winter issues) on the BIOP section journey over the years and reflection from several past BIOP section chairs on their experience. For 2022, we celebrated with a photo booth, superb refreshments, and some first class take-a-way gifts.

Much effort went into the planning and execution of these events. Initial plans had to be changed due to postponing the in-person nature of some of the celebrations. Sincere gratitude and kudos go to the 40th Anniversary Celebration Committee: Jennifer Gauvin (Co-Chair), Meg Gamalo (Co-Chair), Veronica Bubb, Lisa Lupinacci, Meijing Wu, Richard Zink, and their family members and friends who assisted.

Joint Statistical Meetings

BIOP bounced back from COVID at JSM. Our 2022 program was spectacular thanks to Freda Cooner, our BIOP 2022 Program Chair, and Elena Polverejan, Program Chair-Elect. There were 29 invited session submissions listing BIOP as the first sponsor, compared to 12 last year. We were allotted only 4 invited session slots and our submissions won through competition 2 more slots, allowing us a total of 6 invited sessions at JSM 2022. We were also co-sponsor for several others. Topics included Causal Inference, RWE, Bayesian Design and Analysis, Innovations Caused by COVID-19, ML/AI, and External Control Arms.

There were 25 submissions for BIOP contributed sessions in 2022 compared to 14 in 2021. We were able to sponsor 16 in 2022 based on our allotment. There were 85 contributed paper abstracts submitted, 21 posters, and 8 speed abstracts, all of which were included in the program. In addition, BIOP had a student paper session for our 5 award recipients (announced above) plus one competition submission. It was a great way to learn and appreciate the excellent work students have been doing.

New Committee in 2022: Statisticians in Small Biotech (Current Working Committee Name)

A new BIOP subcommittee was approved in 2022. This group is focusing on how to reach out and support statisticians working in small companies, as the only statistician or one of a few. More and more small companies are hiring statisticians in our specialized area and we often find ourselves isolated. Where should we reach out for help with writing an SOP, to learn software, to ask technical questions, or how to survive in a company that doesn't understand what we do?

It was noted that this committee's proposed name may not sound inclusive enough so we are open to changing it. Please stay tuned as the group moves forward with their ideas. Information will be posted on our BIOP website.

Leadership in Practice Committee (LiPCom)

BIOP's LiPCom held its first Leadership Mixer event at JSM. The target audience was early career professionals to build and strengthen their statistical leadership skillset. Several stations were set up in the mixer room, each targeting different important skills and strategies necessary for impactful career development.

Thank you to Rakhi Kilaru (LiPCoM Chair), Lisa Lupinacci, Veronica Bubb, Abie Ekangaki, Emily Butler, Shanthi Sethuraman, Claude Petit, and Simon Davies. It was wonderful to see the early career professionals engaging with BIOP professionals in this constructive, hands-on, example driven approach. Great job!

ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop (RISW)

In 2022 we held RISW at a new venue for us, the Bethesda North Marriott Hotel and Conference Center. By moving to this new venue, we were able to increase the size with additional flexibility for activities. We had 1,020 attendees in person this 2022 which is the largest number we have on record. There were 10 short courses on topics including estimands, using external data, vaccine clinical trials, real-world data, causal inference, and cell and gene therapy statistical issues. There were 43 round table discussions, 2 plenary sessions, 42 parallel sessions, and a poster session where 38 posters were presented.

We are also very happy to report that the 2022 Workshop sponsorship was highly successful! There were 10 principal sponsors and 8 supporting sponsors, raising $72000 to support Workshop and BIOP activities. Sponsorships for the 2023 Workshop are already underway, with numerous opportunities remaining. For more information on Workshop sponsorships, please see the BIOP website.

Thank you to Workshop co-chairs Chia-Wen (Kiki) Ko and Hope Knuckles and the entire RISW organizing committee for bringing us back to an in-person venue and providing and excellent program.

We look forward to the exciting program that our 2023 co-chairs, Fanni Natanegara and Erik Bloomquist, and their steering committee will design and implement.

ASA Biopharmaceutical Section Nonclinical Biostatistics Conference

The Nonclinical Biostatistical Conference is held every other year. It was held with much success in 2021 and we can look forward to the same quality in 2023. Xin Huang will chair the 2023 conference. Stay tuned for announcements as information becomes available.

Fellows Committee

Our BIOP Fellows Committee identifies potential fellows from within BIOP and provides guidance in submitting nomination packages. Please utilize this resource if you are considering applying for an ASA Fellow or if sponsoring someone. A wonderful article was produced by BIOP members in 2020 providing Advice for ASA Fellow Nominations. If you haven't read it, check it out here: *https://magazine.amstat.org/blog/2020/06/01/advice-for-asa-fellow-nomination/*

You can find additional information here: Fellows Nomination Committee - Biopharmaceutical Section (amstat.org)

Scientific Working Groups

The process of establishing new working groups is overseen by the Scientific Working Group (SWG) Committee. Through the SWG Committee, section members can submit research topics that contribute to the goals of advancing science, enabling innovation, and leveraging membership expertise. The establishment of the working group must be approved by the Section Executive Committee and each scientific working group must provide a yearly update report to the Executive Committee.

Individuals interested in forming a new SWG can review the BIOP guidelines for more information.

In 2022, one new SWG joined the ranks, the Software Engineering Working Group led by Daniel Sabanes Bove and Ya Wang.

Thank you to Jennifer Gauvin and Brian Waterhouse for facilitating the SWG process and governance and to all the SWG chairs and members. We've seen a lot of great output from these groups, much of which was shown at our Webinars last year.

Outreach

We are always looking for new ways to build synergies with other groups and share best practices within our own membership! In 2022, we renewed our focus on outreach with through the Outreach and Collaboration Committee and through broader investment proposals initiated by the Next Generation Stewardship Committee.

Our Outreach and Collaboration Committee (OCC) was assigned new members to help give these efforts the attention it needs. Many of the objectives in our BIOP Charter focus on these activities. Successes of the OCC include partnering with ENAR and WNAR to have representatives from our BIOP on their program committees. We're also engaging with the Society of Clinical Trials to determine how we can support them and achieve synergies. We are also supporting external conferences, both those that we've already cemented relationships with and new. We are reaching out to ASA Chapters in regions where our BIOP industries have footprints. We expect our efforts to flower more in 2023.

Additionally, standing budget was allotted to other committees of BIOP which have outreach as part of their remit. This includes the Leadership in Practice Committee (LiPCom) which launched its first outreach initiative in 2022, the Mentoring Committee, and the Membership Committee. With this strategic investment, the BIOP EC has enabled the committees to proactively increase and sustain their outreach efforts with reduced bureaucracy.

Publications

We had 7 BIOP sponsored webinars and 3 issues of the Biopharmaceutical Report.

Education

A huge thanks to Richard Zink for all his work establishing and implementing the BIOP podcast series. Richard has passed the torch after 100 podcasts! Wow! Welcome to Amy LaLonde and Christina Nurse who are continuing this series. There was a total of 5 podcasts in 2022 and we already have the first for 2023. All the Podcast Episodes can be found here: *https://www.buzzsprout.com/16296*.

Biopharmaceutical Section Newly Elected Officers

We would like to welcome the following elected officers and At-large members to the Biopharmaceutical Section Executive Committee for 2023:

- **Chair-Elect 2023:** Ted Lystig
- **Program Chair-Elect 2023:** Bo Huang
- **Treasurer 2023-2025:** Emily Butler
- **Council of Sections Representative 2023-2025:** Meg Gamalo
- **At-large member 2023-2025:** Adrian Coles
- **At-large member 2023-2025:** Xiang Zhang

We know these individuals will do a fantastic job representing the Section in these roles. We thank them for serving and look forward to all they will accomplish!

Outgoing Officers at end of 2022

We'd also like to extend huge thanks to the following outgoing officers:

- Weili He (2021 Chair)
- Freda Cooner (2022 Program Chair)
- Jane Qian (2020-2022 Treasurer)
- Ted Lystig (2020-2022 CoS Representative)

2023 Plan and Final Thoughts

After 40+1 years of BIOP, our Section is thriving with a healthy, engaged membership, established programs that provide value both to the members and the profession, and multiple recent initiatives aimed at increasing our impact. Thus, in our 2023 plan, we've chosen to emphasize maintaining momentum in recent and higher-impact initiatives while nurturing and supporting newer groups for rapid success. The 2023 RISW is poised to grow in several meaningful ways this year, offering more sessions and incorporating regulatory statistics perspectives from FDA and other agencies around the world. The theme of this year's workshop is "Statistical Thinking and Innovation with Global Impact." The Mentoring Program Committee is planning ways to increase engagement at JSM to increase the impact of this successful program. Scientific Working Groups have detailed plans for advancing methods and practice, including the Software Engineering SWG (which is in its first year) and at least one other SWG currently forming which we hope to formally announce soon. And outreach remains a key focus for the Section, as already outlined, through OCC and several Section committees which continuously look for ways to broaden our reach, impact, and membership base.

Throughout 2023, expect to remain informed via social media channels, thanks to our Publications Officer, Hiya Banerjee, and Social Media Chair, Yu Du. Please follow *ASA Biopharmaceutical Section* on LinkedIn and *@ASABiopharm* on Twitter and interact with content (i.e., 'like,' comment, or repost) for additional reach. In addition, expect continued high-quality educational content through our webinar and podcast series.

Thank you for your support of BIOP. As we close this message, we want to thank all volunteers (officers, committee members) who keep this very active Section running. Your work is impactful and enormously appreciated. Lastly, we'd like to remind you that we want to hear your voice. We welcome ideas, suggestions, and any feedback which would help make BIOP a more relevant and impactful home for statisticians in our profession for the years to come. Please reach out to us at any time!

With best regards,
Alan and Brian ■

RACIAL AND ETHNIC DISPARITIES IN PIVOTAL TRIALS SUPPORTING FDA-APPROVED AND EUROPEAN COMMISSION-APPROVED DRUGS

Zachary Smith MA, Emily Botto, Kenneth Getz, MBA
Tufts Center for the Study of Drug Development, Tufts University School of Medicine, Boston, MA. USA
Author address and email: 145 Harrison Avenue, Boston, MA 02111; *Zachary.smith605922@tufts.edu*

Introduction

Interest in clinical trial enrollment diversity has intensified in recent years, in part due to its importance for equity in treatment outcomes among different demographic groups. Clinical trials that are representative of the intended patient population are better able to identify variations in disease biology, social determinants of health, and treatment outcomes. Despite efforts to address the lack of diversity among clinical trial participants, racial and ethnic minorities and women remain proportionally underrepresented in clinical trials when compared to the distribution of these populations by disease prevalence and census estimates[,]. Quantifying and monitoring the magnitude of underrepresentation of demographic groups in clinical trials is an essential step in beginning to improve it.

Over the last several years, the Tufts Center for the Study of Drug development (Tufts CSDD) has undertaken multiple efforts to quantify the depth and magnitude of underrepresentation in pivotal clinical trials. While several studies have examined diversity and representation in individual disease conditions or in a single therapeutic area [, , ,] the Tufts CSDD studies assess a much wider scope of medical conditions that have been the target of all recently approved new drug and biologic applications (NDAs and BLAs) for marketing authorization. The Tufts CSDD studies collect data on pivotal trials because they gather and report the most applicable data assessing the safety and efficacy of an investigational treatment among patient subgroups. As pivotal trials are a required component of regulatory submissions, they are also most likely to report demographic information.

Studies conducted by Tufts CSDD include data on regulatory approvals over longer time horizons, thereby allowing for the evaluation of trends in the data. Because these studies have looked at both FDA-approved and European Commission-approved drugs (two of the largest global regions in terms of regulatory decision making bodies), and were conducted with similar methodologies, some comparisons regarding global patient enrollment diversity and representation can be made.

Methods

Detailed descriptions of the methodology used to examine FDA and EU Commission approvals are provided in peer-reviewed manuscripts and are summarized here [1, 2]. In each study, a list of drugs approved within the timeframes (2007 to 2017 for US, 2007 to 2019 for EU) was compiled from the relevant government website (FDA for US, EMA for EU). Data collected on each drug included trade and generic names, date of approval, special designations (i.e., Breakthrough, Fast Track, Orphan, PRIME, or ATMP), sponsor company, and approved indication.

For each of the approved products, a list of pivotal trials was compiled using the approved labels and drug applications. For FDA approvals, the primary sources of data were the medical reviews or inter-discipline reviews (available on the FDA website); for products approved in the EU, the primary source of data was the European Public Assessment Report (EPAR) for each product. Trial demographic data (average age of participants, number of male participants, number of female participants, number of white participants, number of Black participants, number of Asian participants, number of participants of other racial identities, and number of Hispanic or Latino participants) were collected for each pivotal trial where the data were available. When the data were not available in the applications, publicly available databases were consulted (clinicaltrials.gov for US approvals, EUDRACT for products approved in the EU).

Indication-specific demographic data -were collected for each of the approved indications, again using publicly available data sources. These sources included CDC Wonder, USCS Data Visualizations, and PubMed for the US data. Because the EU does not collect racial

demographic data for a census, much of the indication-specific demographic data was based on rates in the UK, world-wide rates, or in some cases, US rates calculated to match the populations of EMA-covered countries.

Both studies calculated disparities as the percent difference between the actual trial demographics and the estimated patient populations for the approved indications. In cases where no patient demographic estimates existed, comparisons were made to relevant population data (US Census data for the US, and a population estimate calculated for the EU data). Studies that under- or over-enrolled a demographic by 20% or more were considered unrepresentative, either under or over, of that patient demographic.

Results and Discussion

One of the first areas of comparison between pivotal trials for new approvals in the EU and the US is the availability of demographic data. Although the availability of demographic data is generally higher for pivotal trials on US-approved products, overall it is quite low. The most frequently available demographic data is gender of participants – available for around 90% of pivotal trials supporting approvals in the US and less than 80% of trials supporting approvals in the EU (Table 1). Racial demographic data was reported for less than three-quarters of pivotal trials in the US and less than two-thirds of pivotal trials in the EU. Ethnicity demographic data is reported even less often: just over one-third of pivotal trials in the US, and just over one-quarter of pivotal trials in the EU. While the reporting of demographic data has increased significantly over the observed timeframes, there is still room for improvement in this regard.

The reporting of demographic data is important for multiple reasons, however the most important is that it can help identify areas where representation is most lacking. Without the reporting of trial demographic data, areas of inconsistency that are most in need of improvement may not receive the necessary attention or assistance. With diversity apparently increasing both in the US and in Europe, these areas risk falling even further behind if disparities between trials and patient populations are not measured and tracked.

In general, trials supporting European Commission and FDA approval have similar demographic distributions, with only slight differences between them (Figure 1). Despite the similarities, trials underrepresenting non-white racial identities and Hispanic or Latino ethnic identities make up a higher percentage of the pivotal trials overall in the US than in the EU (Table 2). This is likely due to the increased diversity in the US

Table 1. Pivotal Trial Demographic Data Reporting

	Pivotal Trials Supporting New Approvals in the EU (2007 – 2017)		Pivotal Trials Supporting New Approvals in the US (2007 – 2017)	
	n	Percent	*n*	Percent
Gender Demographics Reported	649	78.8%	679	89.7%
Ethnicity Demographics Reported	510	61.9%	551	72.8%
Racial Demographics Reported	229	27.8%	278	36.7%

**n* = number of pivotal trials reporting demographic data.

population, meaning more diverse trials are required in order to properly represent patient populations. A higher percentage of white participants in a trial means the participant demographics are more representative of the EU patient population than the US patient population. Regardless, in both the EU and the US, trials underrepresenting non-white identities are extremely common, with little improvement over the observed timeframe. Between 2007 and 2012, all non-white and Hispanic or Latino identities were underrepresented in roughly half or more of pivotal trials for both European Commission and FDA approvals. While Asian and Hispanic or Latino identities did see some improvement during the 2013 to 2017 timeframe, both identities are still underrepresented in more than 1-in-3 pivotal trials. Black and other racial identities saw little to no improvement in

Figure 1. Demographic Distribution of Pivotal Trial Participants (2007 – 2017)

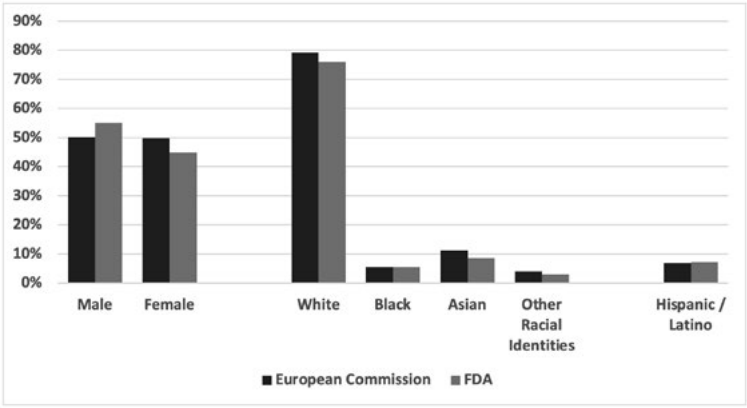


Table 2. Percent of Pivotal Trials Underrepresenting Demographics by More than 20%

	EU Commission Approvals (2007 – 2012)	EU Commission Approvals (2013 – 2017)	FDA Approvals (2007 – 2012)	FDA Approvals (2013 – 2017)
Gender				
Male	12.9%	12.7%	24.3%	13.6%
Female	30.2%	33.1%	33.1%	32.0%
Racial Identity				
White	22.3%	23.0%	13.6%	10.4%
Black	50.9%	51.1%	78.4%	80.7%
Asian	60.0%	42.7%	71.6%	40.3%
Other Racial Identities	66.1%	62.1%	81.1%	76.2%
Ethnic Identity				
Hispanic / Latino	47.8%	36.6%	66.9%	42.8%

the frequency with which they were underrepresented in pivotal trials.

Some therapeutic areas underrepresented minorities at a higher rate than others. For example, among FDA approvals, Black participants were underrepresented in 100% of approvals across 5 therapeutic areas (pulmonary/respiratory, rheumatology, nephrology, hepatology, and pediatrics/neonatology) while Asian participants were underrepresented in 100% of approvals across 4 therapeutic areas (OBG/YN, nephrology, hepatology, and gastroenterology). Additionally, Black participants were underrepresented in over 50% of drug approvals across all therapeutic areas except for psychiatry. Among EU Commission approvals, respiratory (61.5%), cardiovascular (57.4%), and oncology (57.0%) underrepresented non-white populations in the highest proportions compared to other therapeutic areas.

Although both the US and the EU have made efforts during the past decade to improve diversity and representation in clinical trials, it is clear that there is still room for improvement. Stricter requirements for collecting and reporting of demographic data will allow for more accurate tracking of diversity and representation in clinical trials, and the inclusion of more trial sites outside the US and the EU could help increase the diversity of the trials. Tufts CSDD is currently working to update the FDA dataset to include more recent approvals and will continue to periodically perform updates to track the success of any additional efforts.

Bibliography

[1] Getz, K.A., Smith, Z.P., Peña, Y., 2020, Quantifying Patient Subpopulation Disparities in New Drugs and Biologics Approved Between 2007 and 2017, Ther Innov Regul Sci, 54, 1541–1550.

[2] Smith, Z., Botto, E., Getz, K., 2022, Quantifying Diversity and Representation in Pivotal Trials Leading to Marketing Authorization in Europe, Ther Innov Regul Sci, 56, 795–804.

[3] Charrow A., Xia FD., Joyce C., Mostaghimi A., 2016, Diversity in Dermatology Clinical Trials: A Systematic Review, JAMA Dermatol ,153(2), 193–198.

[4] Kaakour A., Hua H., Rachitskaya A., 2022, Representation of Race and Ethnicity in Randomized Clinical Trials of Diabetic Macular Edema and Retinal Vein Occlusion Compared to 2010 US Census Data, JAMA Ophthalmol, 2022, 140(11), 1096–1102.

[5] Kim J., Kester R., Blumenthal G., 2022, Clinical Trial Diversity in Oncology: FDA Takes Action with Post-Marketing Requirements or Commitments, Oncologist, 27(12), 993-997.

[6] Pinsky P., Ford M., Gamito E., Higgins D., Jenkins V., Lamerato L., et al., 2008 Enrollment of Racial and Ethnic Minorities in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial, Journal of the National Medical Association, 100(3), 291-298.

ASSESSING INVESTIGATIVE SITE PERSONNEL DIVERSITY AND ITS RELATIONSHIP WITH PATIENT ENROLLMENT DIVERSITY

Kenneth Getz, MBA
Tufts Center for the Study of Drug Development, Tufts University School of Medicine, Boston, MA. USA
Author address and email: 145 Harrison Avenue, Boston, MA 02111; *Kenneth.getz@tufts.edu*

Introduction

For several decades, the clinical research enterprise has believed that the geographic proximity of investigative sites is associated with the diversity of patients enrolled in clinical trials. It stood to reason that research centers in urban settings or areas with a high concentration of minority communities would be more likely to attract and retain patients from those communities. This belief is unfounded and empirical research shows that, during the past 30 years, placing clinical trials based on geographic location has had limited impact on improving minority disparities in clinical trials [1].

More recent empirical research has examined many other factors that may impact the underrepresentation of racial and ethnic demographic subgroups in clinical trials including strategies and tactics that improve: transparency and disclosure; public awareness, education and trust; patient recruitment communication effectiveness; and patient willingness to enroll and remain in clinical trials [2]. Recent research also suggests that the underrepresentation of minority clinical investigators who receive National Institute of Health (NIH) awards is associated with patient recruitment disparities [3].

Until recently, however, there has been virtually no empirical evidence examining the relationship between the race and ethnicity of clinical research professionals conducting pharmaceutical industry-funded clinical trials and the race and ethnicity of patients enrolled in these trials. Despite the fact that 85% of all clinical trials are funded by industry, there has been no empirical baseline data on the racial and ethnic distribution of investigative site personnel performing clinical trials [4].

In 2021, my team at the Tufts Center for the Study of Drug Development (Tufts CSDD), Tufts University School of Medicine (Tufts CSDD), conducted a robust study to gather these baseline measures and to better characterize the relationship between investigative site personnel diversity and patient enrollment diversity.

Capturing Personnel and Patient Diversity

A global online survey was conducted between May and July 2021. Twenty-four organizations provided financial support and assisted Tufts CSDD in inviting investigative site professionals to complete the survey. These organizations included AbbVie, Amgen, Association of Clinical Research Professionals (ACRP), AstraZeneca, Biogen, BMS, Covance, CSL Behring, Eli Lilly, EMD Serono, ICON, IQVIA, Janssen, Merck, Otsuka, Parexel, Pfizer, Roche/Genentech, Sanofi, Seagen, Society of Clinical Research Sites (SCRS), Syneos Health, Takeda and UCB. The survey instrument and methodology were reviewed and approved by an ethical review committee, by Tufts University's independent data review committee, and deemed compliant with European General Data Protection Regulation (GDPR). The online survey asked respondents to provide 2020 data on total headcount, race and ethnicity of all personnel employed, site operating characteristics, and the demographic characteristics of patients enrolled.

Table 1 summarizes the definitions of race and ethnicity -- drawn from internationally recognized sources -- that were provided to assist respondents in completing the online survey [5] [6].

Table 1: Definitions of Race & Ethnicity

Asian	A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.
Black or African descent	A person having origins in any of the black racial groups of Africa.
Hispanic/Latino (Spanish Origin, Hispanic, or Latino)	A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race.
Other	A person having origins in any of the original peoples of the Americas (including North America, Central America, and South America), who maintain tribal affiliation, have origins in any of the original peoples of the Middle East or North Africa or have origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.
White	A person having origins in any of the original peoples of Europe.

Descriptive data (e.g., geographic location, site type, clinical trial volume) and data from several arbitrarily selected site-specific operating characteristics (e.g., type of communication and technology solutions used) were assessed to identify and remove duplicates and derive estimates for the total number of distinct investigative sites participating in the survey.

Survey response data was stored in Microsoft Excel and data cleaning and analysis was conducted in SAS version 9.4. Descriptive statistics, frequency comparisons, coefficients of variation (defined as the ratio of standard deviation to the mean), comparisons of mean overall and subgroup response values, significance testing, correlations and multi-variate regression analyses were performed.

Investigative Site Personnel Diversity

In total, we gathered complete responses from 3,187 distinct investigative sites representing 40% of the total number of unique investigative sites worldwide involved in FDA-regulated clinical trials in 2020 [7]. One-third of these sites are operating in Europe; 17% in South/Central America, Asia Pacific and other parts of the world; and half in North America. Survey respondents had a mean tenure of 12 years at their respective investigative site. The role and demographic characteristics of respondents are summarized in Table 2.

Table 2: Respondent Characteristics

	N	Percent of Total
Total Respondents Consenting	3,462	100%
Role*		
Study coordinators and research nurses	466	25%
Principal investigators	756	40%
Site directors, managers and other administrative staff	676	36%
Sex*		
Female	1,206	61%
Male	770	39%
Race & Ethnicity*		
White	1,344	69%
Hispanic/Latino	231	12%
Asian	146	8%
Black	69	4%
Other	159	8%

** Note – reflects only those respondents that chose to divulge this information.*

The Investigative sites captured in this study participate in a wide variety of therapeutic area specialties with the most common being cardiovascular disease, oncology, gastroenterology, neurology and pulmonary/respiratory diseases. Eight-out-of-ten (82%) global investigative sites are operating in urban settings, 15% in suburban and 3% in rural settings where these sites serve a mix of communities by income level: 49% of sites are located in communities classified as middle income, 29% as high income and 22% as low income. Half of the investigative sites represented in this study are based within academic medical centers, health systems, regional and community hospitals; the other half are based in private settings as independent investigative sites or as part of a site network.

Overall, nearly two-thirds (65%) of global investigative site personnel identify as White; 6% identify as Black or of African Descent; 19% identify as LatinX; 7% identify as Asian and 4% identify as Other races and ethnicities. Nearly three-out-of-four (71%) investigative site personnel identify as female. Table 3 shows the racial and ethnic diversity of investigative site personnel overall and by setting where industry-funded clinical trials are conducted.

Table 3: Distribution of Global Investigative Site Personnel by Race & Ethnicity Overall and by Site Type

All Investigative Sites	
Asian	7.3%
Black	6.2%
Hispanic/Latino	19.1%
Other	3.5%
White	65.1%
Sites based in academic, health system and hospital settings	
Asian	9.2%
Black	4.5%
Hispanic/Latino	10.5%
Other	3.1%
White	73.6%
Sites based in private sites and site networks	
Asian	5.3%
Black	7.8%
Hispanic/Latino	27.6%
Other	3.8%
White	56.5%

The mix of investigative site personnel by race and ethnicity varies considerably by therapeutic area. The most diverse personnel are in investigative sites conducting clinical trials focusing on infectious diseases, vaccines and

endocrine disorders. Oncology and dermatology have significantly lower relative representation of Black investigative site personnel. Table 4 presents site personnel diversity for the most common therapeutic areas.

Patient Diversity and its Relationship with Site Personnel Diversity

Global investigative sites report that 64.2% of enrolled patients identify as White; 9.9% as Black or of African Descent; 19.3% as LatinX; 5.2% as Asian and 2.3% identify as Other races and ethnicities. In clinical trials performed in academic, health system and hospital settings, 73.9% of patients enrolled are White. This compares to 57.9% of enrolled patients in private settings. Private sites and site networks enroll significantly higher proportions of Black and LatinX patients. Table 5 shows the distribution of enrolled patient race and ethnicity overall and by research setting.

Worldwide, site personnel diversity is correlated and predictive of patient enrollment diversity. As personnel diversity in global sites increases, the proportional races and ethnicities of patients enrolled at those sites also increases. This association is significant for all races and ethnicities, except for Asian patients who were enrolled in clinical trials conducted by sites that tended to have substantially higher relative representation of Asian personnel. In addition, a significant relationship was observed between the races and ethnicities of site personnel and the corresponding races and ethnicities of patients enrolled in clinical trials at that investigative site. As the proportion of a given race or ethnicity increases among site personnel, so does the corresponding race and ethnicity of patients enrolled. Tables 6 presents the results of linear regression analyses.

A Critical Opportunity to Address Racial and Ethnic Disparities in Clinical Trials

This study establishes an important baseline measure of the distribution of investigative site personnel conducting industry-funded clinical trials world-wide by race and ethnicity. The study also presents baseline measures of site personnel diversity by therapeutic area. And the study empirically demonstrates a strong relationship between the race and ethnicity of investigative site personnel and that of their enrolled patients.

A major opportunity exists for research sponsors and CROs to address racial and ethnic under-representation in clinical trials by selecting and engaging

Table 4: Distribution of Global Investigative Site Personnel by Race, Ethnicity and Therapeutic Area

	Asian	Black	Hispanic/ Latino	Other	White
Cardiology	8.5%	6.9%	22.3%	4.0%	59.7%
Dermatology	7.6%	5.3%	24.6%	3.0%	59.5%
Endocrinology	8.4%	7.7%	24.9%	4.2%	55.4%
Hepatology	9.8%	7.2%	23.5%	4.2%	57.2%
Infectious Diseases and Immunology	9.7%	9.1%	20.3%	4.4%	58.2%
Nephrology	12.1%	6.4%	25.5%	4.8%	53.4%
Oncology	8.7%	5.2%	19.1%	4.5%	63.2%
Ophthalmology	14.8%	6.5%	24.2%	4.6%	52.8%
Pediatrics/ Neonatology	10.4%	8.4%	17.5%	4.7%	59.6%
Psychiatry	8.5%	7.1%	21.6%	3.9%	59.5%
Rheumatology	7.7%	5.9%	25.1%	3.6%	57.1%
Vaccines	7.4%	10.0%	24.6%	4.6%	54.4%

Table 5: Site Reported Distribution of Patients Enrolled Overall and by Site Type

	All Global Investigative Sites	AMC/HS	Private Sites and Site Networks
Asian	5.2%	7.7%	2.5%
Black	9.9%	6.6%	12.2%
Hispanic/Latino	19.3%	10.4%	25.4%
Other	2.3%	2.3%	2.4%
White	64.2%	73.9%	57.9%

Table 6. Linear Regression Showing the Relationship between Site Personnel and Enrolled Patient Diversity

Site Personnel Diversity and Enrolled Patient Diversity	Coefficient	Adjusted r-square	*p*-value
All Investigative Sites Worldwide			
Asian	-.004	-.002	.862
Black	.273	.173	<.001
Hispanic/Latino	.162	.018	<.001
Other	.053	.051	<.001
White	-.499	.138	<.001

Site Personnel Diversity and Corresponding Enrolled Patient Diversity	Coefficient	Adjusted r-square	*p*-value
All Investigative Sites Worldwide			
Asian	.692	.647	<.001
Black	.857	.443	<.001
Hispanic/Latino	.873	.866	<.001
Other	.393	.746	<.001
White	.826	.246	<.001

with investigative sites whose personnel best reflect the under-represented patient communities of interest. As clinical trials increasingly transition to remote and virtual approaches (e.g., telehealth, wearable devices, smartphones, and home visits), where the geographic location of the investigative site becomes less defining, the diversity and cultural competency of the clinical research workforce and its ability to connect with a diverse patient population will become even more important. Machine-learning and other AI-enabled planning and decision support are expected to assist sponsors and CROs in interrogating rich structured and unstructured datasets to identify patients from diverse demographic communities.

Given high interest in addressing racial and ethnic disparities in clinical trials, the results suggest a strategic opportunity for investigative sites to differentiate themselves by hiring diverse investigative site personnel who share and reflect the perspectives, cultural views and experiences with the corresponding race and ethnicity of their patients.

The study methods had several notable limitations. Data on personnel and patient enrollment diversity are based on self-report from respondents. And bias may have been introduced by limiting responses to an online survey instrument. The Tufts CSDD team is planning several future studies including research that examines investigative site personnel diversity and clinical trial operating economics and performance as well as patient-reported demographic and socio-economic characteristics and their relationship with clinical trial recruitment and retention effectiveness.

Acknowledgments

The author thanks members of the Tufts CSDD team - Zak Smith, Maria Florez, and Emily Botto -- for their invaluable assistance in this research. The author also thanks the following organizations -- AbbVie, Amgen, Association of Clinical Research Professionals, AstraZeneca, Biogen, BMS, Covance, CSL Behring, Eli Lilly, EMD Serono, ICON, IQVIA, J&J, Merck, Otsuka Pharmaceuticals, Parexel, Pfizer, Roche/Genentech, Sanofi, Seagen, the Society of Clinical Research Sites, Syneos Health, Takeda, UCB -- for providing input, assistance and grant funding.

References

[1] Getz K., Smith Z., Pena Y., 2020, Quantifying Patient Subgroup Disparities in New Drugs and Biologics Approved Between 2007 – 2017, TIRS, 54(6), 1541-1550.

[2] Clark L., Watkins L., Pina I., Elmer M., ASkinbobooye O., Gotham M., Jameson B., McCullough C., Pierre C., Polis A., Puckrein G., Regnante J., 2019, Increasing Diversity in Clinical Trials: Overcoming Critical Barriers, *Curr Probl Cardio*, 44, 148-172.

[3] Ginther D., Schaffer W., Schnell J., Masimore B., Liu F., Haak L., Kington R., 2011, Race, Ethnicity, and NIH Research Awards, *Science*, 333(6045), 1015-1019.

[4] Getz K., 2018, Trends driving clinical trials into large clinical care setting, Nature Review Drug Discovery, 17, 703-704.

[5] World Health Organization, 2020, International Clinical Trials Registry Platform.

[6] Racial and Ethnic Categories and Definitions for NIH Diversity Programs and for Other Reporting Purposes, 2015.

[7] Food and Drug Administration's Bioresearch Monitoring Information System (BMIS), 2020, *https://www.fda.gov/drugs/drug-approvals-and-databases/bioresearch-monitoring-information-system-bmis*

A DIGITAL TOOL FOR DYNAMIC PATIENT RECRUITMENT PREDICTION IN MULTICENTER CLINICAL TRIALS

Jianmin Chen (University of Connecticut), Zhaowei Hua (Servier), Qian Meng (Servier), Lucie Truffaut-Chalet (Servier), Zhaoyang Teng (Servier), Rui Tang (Servier)

Clinical trial is the crucial step to investigate the safety and efficacy profile of a new drug with years of costly research and development. The time-consuming and expensive patient recruitment marks the start of the clinical trial. Optimal recruitment planning and monitoring plays a vital role in successful trials. As reported in the review paper [1], poor recruitment is the principal reason for the premature discontinuation of clinical trials, leading to various negative impacts, including time delay, extra expenses, and missed marketing opportunities. Growing digital technology advances patient screening and enrollment with robust real-time data collection footprint and support. As such, a dynamic predictive digital tool can be utilized to monitor the recruitment process over time, which supports decision-making and risk management in drug development process.

Managing patient screening and enrollment is complicated. One important reason is heterogeneous patient population, which can include multiple regions, different disease subtypes etc. Sometimes enrollment plan changes in the middle of the study can add another layer of complexity. The challenges to forecast patient screening and enrollment involve simultaneous screening prediction and enrollment prediction, addressing heterogeneity in patient population, capturing changes on enrollment curve, and randomness in the process (e.g. from centers, patients). Here we propose a Bayesian model-based, data-driven, and dynamic prediction framework. It includes multi-stage modeling, captures heterogeneity in patient population, and continuously provides multi-view inference to support risk management.

Specifically, we consider the case when patients with distinct disease subtypes or biomarker outcomes are to be recruited [2] [3]. The multicenter trial with centers distributed in different regions is commonly conducted to ensure an adequate number of patients. An example with three centers is shown in Figure 1, where patients diagnosed with two different disease subtypes are recruited, and the target numbers of patients are set separately for each subtype. As illustrated, centers can have different opening and stopping times, and different subtypes may also have specific enrollment completion dates.

Furthermore, each center may have a different status at the data cutoff point (stopped, opened, opened with no screened patients, or still in planning). Under such a circumstance, the complexity increases as the randomness in the process

Figure 1: A multicenter recruitment with two disease subtypes

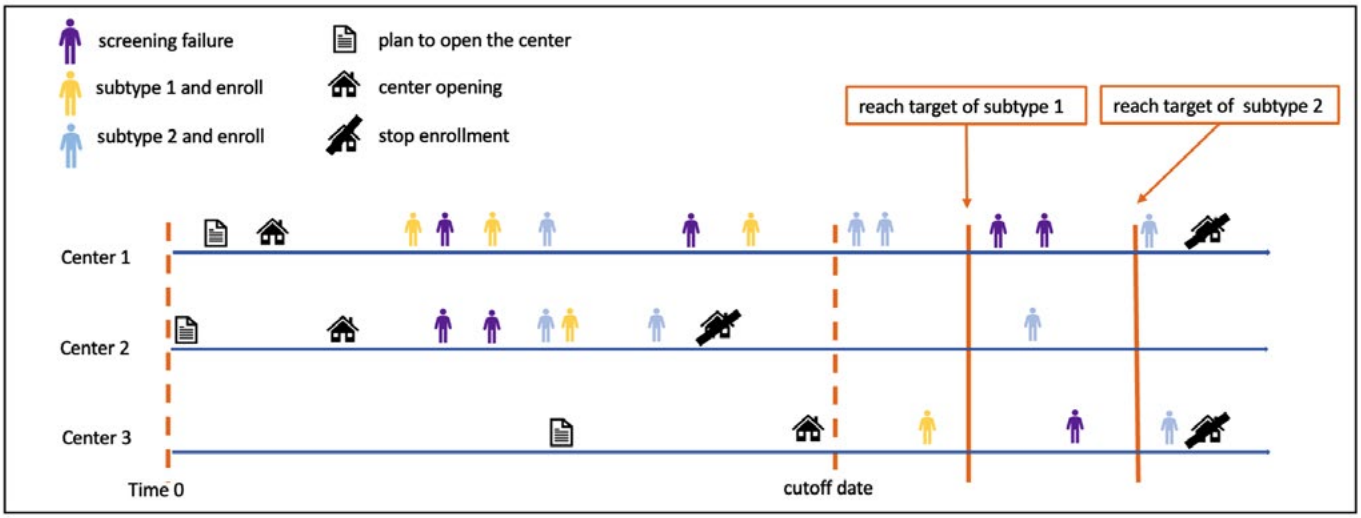


Figure 2: Four periods in the patient recruitment process. IAT stands for "inter-arrival time".

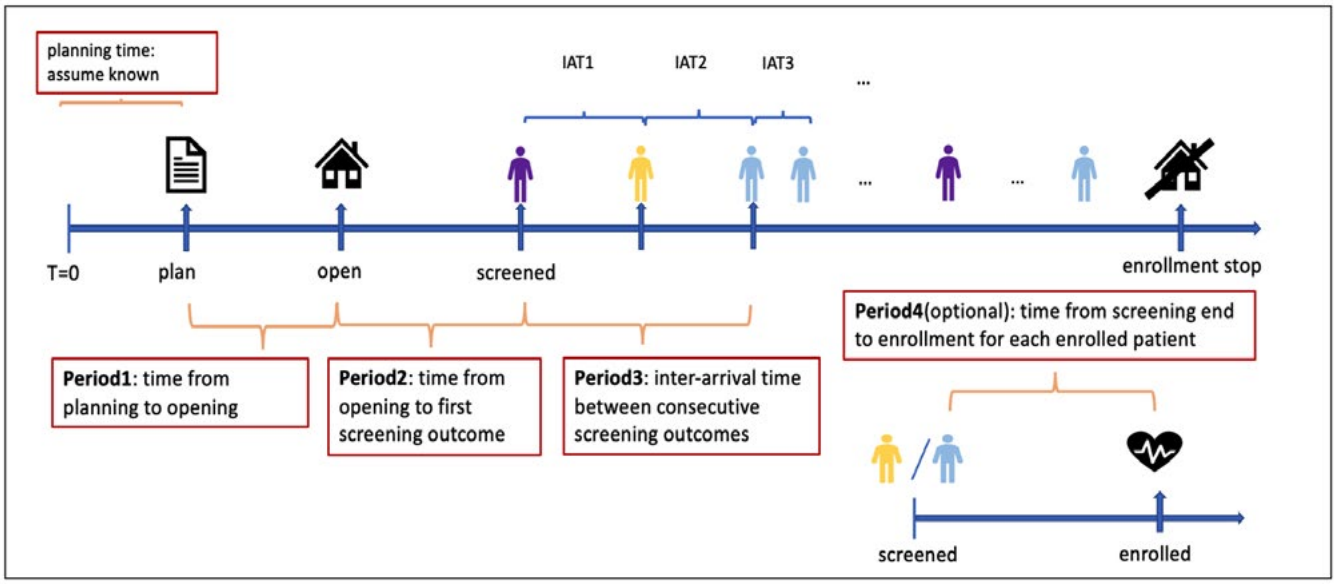
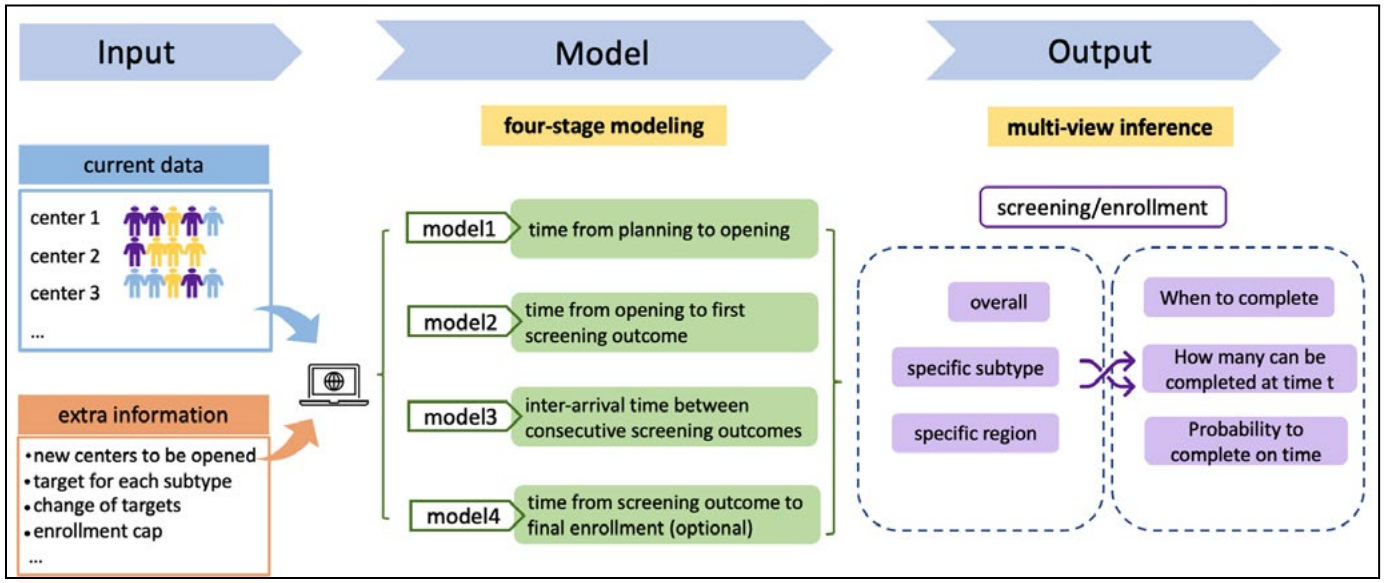


Figure 3: Patient recruitment prediction framework

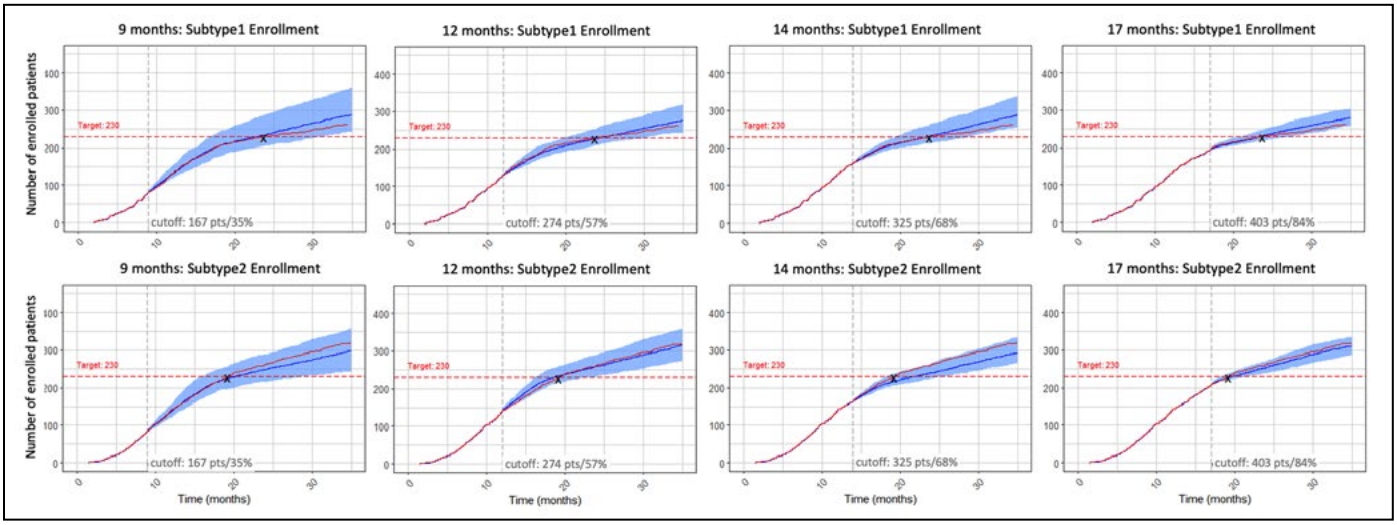


can arise from centers, disease subtypes, and patients. To address this, we consider a Bayesian model-based approach to capture the uncertainty, and simultaneously predict the number of patients screened/enrolled at a future time point, at different center(s), or for different disease subtype(s).

Consider the recruitment of the ith patient at the jth center. As described in Figure 2, the recruitment can be viewed as a four-stage process. Assuming the planning time is known we focus on modeling the time gaps between different stages: time from planning to opening (T_{1j}), time from opening to the first screening outcome (T_{2j}), the inter-arrival time between consecutive screening outcomes (T_{3ji}) deliveries, and the time between screening outcome and the final enrollment (T_{4ji}) if the patient does not fail the screening. We consider four separate models for the four periods and the final predicted enrollment time of the ith patient at the jth center can be formulated as $T_{1j} + T_{2j} + \sum_{k \le i} T_{3jk} + T_{4ji}$ (Figure 3).

At each data cutoff point, we collect both center-level data and patient- level data from centers that have already been opened. The information is then shared between similar centers to help predict the future enrollment process. By working through this four-stage framework, we can predict the screening results of a specific subtype and the final screening/enrollment date for each upcoming patient. By combining these individual prediction results, we can generate predictions on the number of screened/enrolled

Figure 4: Dynamic prediction example: The grey dashed line indicates the cutoff date, where the number of cumulative patients (pts) is calculated based on overall enrollment. The truth is marked with the red curve and the prediction is marked with blue curve. The blue region is the 95% prediction interval for prediction based on repeated sampling. For each subtype, the enrollment target is 230.



patients at a future time point for different centers or subtypes. To assess the probability that the recruitment can be completed on time, we obtain repeated samples from the posterior distribution. To further enhance accuracy, the final prediction is adjusted to accommodate other prior information, including but not limited to the enrollment cap for a single center/region.

To provide continuous support for recruitment planning, a dynamic prediction framework is proposed. The motivation is to keep the predictive model adaptive to the latest recruitment behavior which can be influenced by unexpected changes that may occur during the long recruiting period, while taking into account all patient screenings from the beginning of the study. Such changes are commonly seen. For example, the Covid-19 pandemic has significantly impacted the recruitment process for many trials. As a result, models built at the early time of recruitment are less precise in generating accurate predictions. Therefore, regular model updates are suggested to ensure effectiveness and reliability for prediction. A prediction example is provided in Figure 4 with simulated data, where the model is fitted at four cutoff points: 9 months, 12 months, 14 months, and 17 months. The study has two disease subtypes, and we demonstrate the enrollment prediction for each subtype. As one can tell from the blue curves in the plots, our model works very well in predicting the nonlinear trend in the true data curves, and is able to adapt to the change points in the recruitment process. Additionally, as more data becomes available, the prediction interval gets narrower, indicating increased precision and reliability.

To summarize, this model-based, data-driven, and dynamic prediction tool offers great flexibility and accuracy in predicting patient recruitment in various real-world scenarios, while highlighting potential risks in the future and supporting decision-making. It can capture heterogeneity in patient population, enable dynamic recruitment prediction, make multi-view inference, and provide accurate and timely forecast. It is essentially a digital tool for dynamic recruitment prediction. Ultimately, the use of digital technology can lead to more efficient and successful clinical trials, benefiting both patients and the pharmaceutical industry.

References

[1] Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review. Contemporary clinical trials communications. 2018 Sep 1;11:156-64.

[2] Glynn C, Tokdar ST, Zaman A, Caruso VC, Mohl JT, Willett SM, Groh JM. Analyzing second order stochasticity of neural spiking under stimuli-bundle exposure. The annals of applied statistics. 2021 Mar;15(1):41.

[3] Horwitz S, O'Connor OA, Pro B, Illidge T, Fanale M, Advani R, Bartlett NL, Christensen JH, Morschhauser F, Domingo-Domenech E, Rossi G. Brentuximab vedotin with chemotherapy for CD30-positive peripheral T-cell lymphoma (ECHELON-2): a global, double-blind, randomised, phase 3 trial. The Lancet. 2019 Jan 19;393(10168):229-40

ML-BASED POPULATION SELECTION AND ENRICHMENT IN DRUG DEVELOPMENT

Natalia Muehlemann, MD, MBA (Cytel)

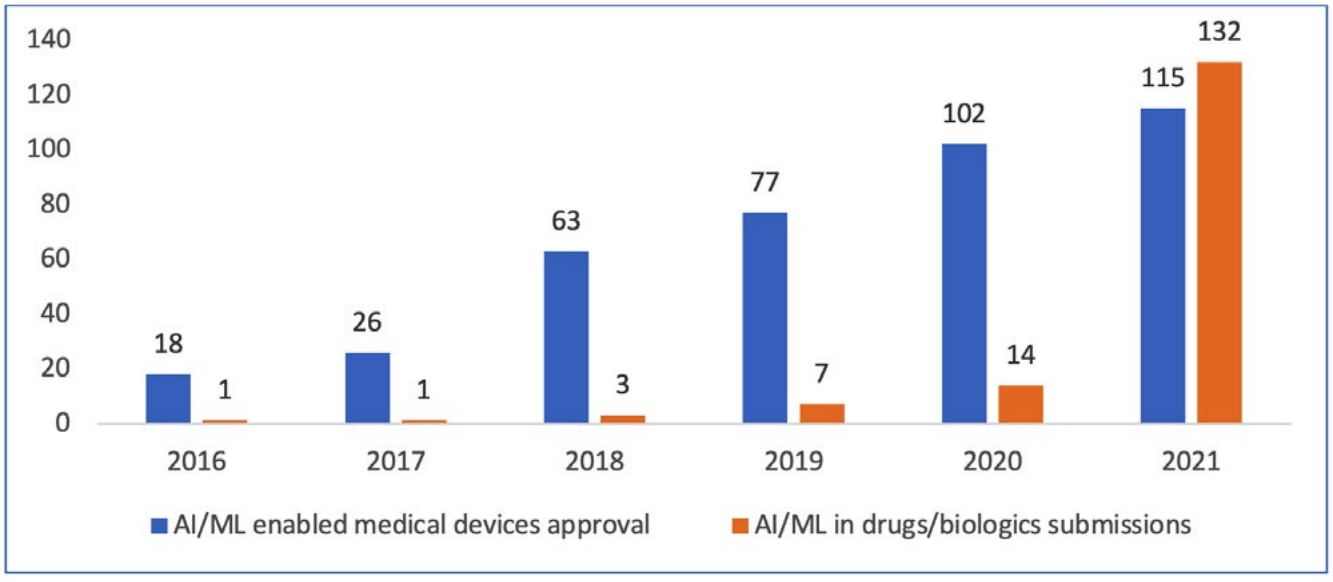
Introduction

Artificial Intelligence (AI) and Machine Learning (ML) techniques have been used to develop software algorithms for imaging diagnostic systems, smart sensor devices and other clinical applications[1,2]. FDA's CDRH has established a regulatory framework for ML-based devices and software as medical devices (SaMD) and continues to lead the U.S. regulatory framework for AI/ML software development. Currently there are more than 500 AI/ML-enabled medical devices / algorithms approved by FDA[1]. Approximately 75% of applications are in radiology and 11% in cardiology[1].

ML has a potential in many areas of drug development, from drug discovery to clinical development. Drug discovery, including drug design and AI target selection to better characterize disease states and identify novel 'druggable' targets, accounts for the majority of AI/ML applications in industry research and development pipeline[3]. However, the application of AI/ML to clinical development have been increasing, especially recently. Based on an analysis of AI/ML use in regulatory submissions of drugs and biologics to FDA, 2021 seems to be a break-through year with 132 submissions as compared to a growing yet single to low double-digit numbers in previous years[4]. About 88% of theses AI/ML applications were in the clinical development stage[4]. Figure 1 below illustrates the evolution of AI/ML-enabled medical devices approvals by FDA's CDRH and submission for drugs and biologics with AI/ML components to FDA's CDER and CBER.

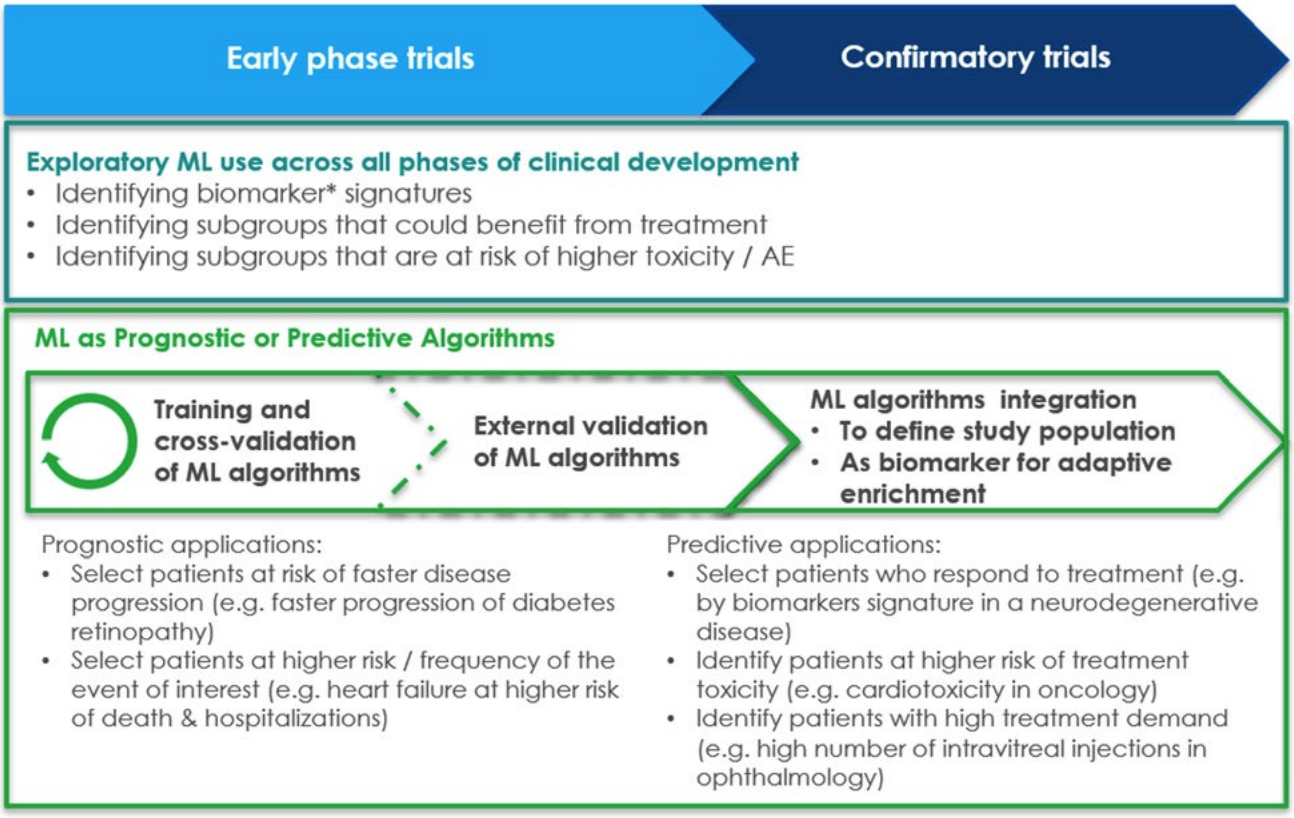
Personalized medicine, also known as precision medicine, shifts the direction of drug development from traditional trials to biomarker-based approaches. Traditional clinical trials are based on broad eligibility criteria and measure average outcome for all patients recruited. The goal of personalized medicine is to tailor medical treatment to the individual characteristics, needs, and preferences of a patient during all stages of care, including prevention, diagnosis, treatment, and follow-up[5]. The use of diagnostic and prognostic biomarkers for disease classification and risk prediction in clinical development is not a new idea. For example, in oncology, predictive biomarkers have been associated with more than 50% of oncology clinical trials[6]. Biomarker includes not only molecular biomarkers

Figure 1. **AI/ML-enabled medical devices approvals and Drugs submissions with AI/ML components**



Sources: based on references [1] and [4]

Figure 2. **High-level illustration of ML applications in clinical development**



but also radiological and physiological biomarkers that could be based on imaging, lab or measurements from wearable devices[7]. With increasing availability of good quality data, different types of measurements and clinical characteristics could be combined into AI/ML-based algorithms to help identify patients who are more likely to develop a particular outcome (prognostic algorithms) or who are more likely to respond to a particular drug (predictive algorithms). Figure 2 below illustrates potential AI/ML applications in clinical development. The following section outlines selected examples of developing AI/ML algorithms that could help patient selection for clinical trials based on published literature and our group's experience.

AI/ML-enabled patient selection in drug development

AE prediction

The ability to predict who will experience an adverse events using AI/ML algorithm has the potential to identify patient populations at lower risk of treatment toxicity and more favourable risk-benefit profile for an investigational therapy. Different ML-based approaches have been applied to predict treatment toxicity. One of the more frequently used approaches is safety risk prediction based on drug's structure, physiochemical properties or affinity for targets. For an example of using this approach to predict safety risks of drug in regulatory submission, readers could refer to "In silico Analyses on the Potential Association of Remdesivir with Renal and Hepatic Events"[8]. Other approaches include a ML prediction model based on clinical and/or genetic factors. The example below describes a case study where ML approaches can help to identity clinical and genetic factors associated with anthracycline cardiotoxicity in pediatric cancer survivors[9].

Anthracycline chemotherapy is frequently used in pediatric cancers. However, it is major cause of cardiac morbidity: approximately 60% of cancer survivors develop echocardiographic cardiac dysfunction (measured by left ventricular ejection fraction (LVEF)), and approximately 10% develop delayed toxicity manifested by symptomatic cardiomyopathy[9]. A number of clinical factors, such as demographic characteristics, cumulative dose, radiotherapy involving heart and pre-existing cardiac dysfunction, were associated with cardiotoxicity, however these clinical factors have limited predictive ability. The goal of one project was to identify the contribution of rare and low-frequency single-nucleotide variants (SNV) that influence the susceptibility

to cardiotoxicity, to validate the functional role of the affected genes, and to develop a prediction model for cardiotoxicity that combines clinical and genetic factors. As the first step, gene-level statistical tests were applied to gene-level rare variants in order to identify significant genes associated with cardiotoxicity. As the second step, ML approaches - Random Forest and Penalized Regression Models – were applied to genes to select predictive biomarkers[9].

Three machine learning algorithms were developed using 1) clinical factors alone, 2) genetic factors alone and 3) a combination of clinical and genetic factors. The stratified bootstrapping sampling scheme was applied to generate 1,000 replicates by using the discovery cohort (i.e. set of subjects used to identify factors and develop the algorithm). Each replicate was a pair of independent randomly selected training and testing sets with the testing set corresponding to samples not selected in the training set. A random forest (RF) classifier was applied to each replicate to predict cases with cardiotoxicity based on the 3 models (clinical, genetic, and combined). RF aggregates the votes from different decision trees to determine the prediction[9].

There were 289 patients included in the discovery cohort for exome sequencing, including 183 case patients (with reduced LVEF) and 106 control patients (with preserved LVEF). Three RF prediction models were developed in a training set and evaluated using a random resampling approach in the absence of an external validation cohort. The AUC for the clinical, genetic and the combined models were 0.59 (95% CI: 0.51 to 0.67), 0.71 (95% CI: 0.63 to 0.80) and 0.72 (95% CI: 0.63 to 0.80) respectively. The genetic and the combined clinical and genetic models outperformed the clinical model with a higher AUC, higher specificity, higher positive predictive value, and a lower misclassification rate[9].

Prognostic applications

ML can be used to develop algorithms to identify subpopulations at higher risk of developing a clinical event of interest or being at a risk of faster disease progression. Focusing clinical development on a subpopulation enriched for the events of interest can help reduce sample size and duration of clinical trials. Depending on indication and clinical rationale as well as the mechanism of action for the investigational treatment, such population enrichment does not necessarily lead to indication restrictions.

Prognostic algorithms have been developed and widely used in clinical practice. For example, in Intensive Care, several scoring systems (e.g. APACHE, SAPS, SOFA) have been developed and applied for over four decades. Score can be used for outcome prediction, risk stratification and monitoring response to treatment. It is important to note that in order to remain valid predictors, the scores need to be updated from earlier versions because of updates in clinical science and practice (e.g. APACHE IV, SAPS-3). A large quantity of measurements from multiple monitoring systems, lab measurements and clinical factors enable novel ML application in Intensive Care Units (ICU). The recent publication by Hyland and colleagues [10] describes the use of ML to develop an early-warning system to predict circulatory failure events in critically ill patient in ICUs. Monitoring of circulatory function is an essential part if ICU patient management, and early detection of patient at risk is important because the effects of circulatory failure are initially reversible in many ICU patients[10]. The researcher developed a comprehensive analysis framework including data pre-processing and cleaning, feature extraction and interpretation, and selection of large-scale supervised ML techniques to construct the early-warning systems that predicted 90% of circulatory-failure events in the test set, with 82% being identified more than 2 hours in advance. The researchers reported achieving an area under the receiver operating characteristic curve of 0.94 and an area under the precision-recall curve of 0.63. Readers who are interested in details on methods and results can refer to the publication [10].

Cardiovascular pivotal trials are more and more challenging due to increasingly large sample size and duration. For example, in drug or device development for Heart Failure, clinical outcomes such as a combination of mortality and Hospitalizations for Heart Failure (HFH) remains a requirement for registrational trials. This is another clinical situation where prognostic modelling could be potentially useful to identify patient population at higher risk for the event of interest i.e. death and HFH. An example of recent developments in this field is the research by Bradley and colleagues on development and external validation of a Cox proportional hazards model to predict hospitalisation for heart failure and death in patients with, or at risk of, heart failure before first hospitalisation [11].

Drug development in some therapeutic areas is fac-

ing a challenge of designing clinical trials in populations with heterogeneous rates of disease progression. For example, in ophthalmology, patients with mild and moderate diabetes retinopathy (DR) could benefit from novel therapeutic approaches that could slow down progression, delay the development of vision-threatening complications and eventually postpose the need for invasive procedures. Patients with mild and moderate DR are currently managed through a "watch and wait" approach of regular follow up monitoring and observation by ophthalmologist. However, evidence suggests that some of these patients can progress quickly to vision-threatening complications, and therefore may benefit from early treatment. ML can help to identify sub-population of patients with mild and moderate DR who are at high risk of progression. This approach helps to ensure that only patients who can benefit are eligible to receive the treatment, and clinical trial is feasible from the sample size / power perspective. Several researchers aimed at developing risk predictors in DR have already published their algorithms. For example, a recent publication by Arcadu and colleagues [12] reported the development of a deep learning algorithm used as input color fundus photographs to predict DR progression. Ting and colleagues [13] described the use of deep learning systems as a grading tool to determine the prevalence and the systemic cardiovascular risk factors on DR. Other projects to identify ML-based algorithms to select DR population at risk of fast progression as enrichment strategy for ophthalmology drugs development are ongoing.

Predicting Response to Treatment

ML can be used to identify patient population(s) with better response to an investigational treatment.

Similar to the previously described example of cardiotoxicity prediction, ML can be used to predict response to therapies. Immunotherapies are known to yield a strong a sustained response but only in subset of patients. Therefore, a major challenge of precision medicine in immunotherapy is identifying methods that could predict drug responses across multiple cancer patient cohorts. In one of the recent publication in this field, Kong and colleagues described a Network-based ML to predict immunotherapy response in cancer patients [14].

An interesting example from the field of Intensive Care is a retrospective analysis by Seymour and colleagues [15] using statistical, machine learning, and simulation tools in 20 189 total patients (16 552 unique patients) who met Sepsis-3 criteria within 6 hours of hospital presentation at 12 Pennsylvania hospitals. Applying consensus k-means-clustering algorithm to 29 variables, researchers identified 4 phenotypes which were correlated with host-response patterns and clinical outcomes. Simulations suggested it may help in understanding heterogeneity of treatment effects[15].

ML can also be used to help establish patient-specific treatment plan. The current treatment for diabetes macular edema (DME), neovascular age-related macular degeneration (nADM) and retinal verin occlusion (RVO) is based on intravitreal injections of anti-vascular endothelial growth factor agents (anti-VEGF). The current optimization of treatment regimens relies on retinal imaging to monitor the disease activity and treatment efficacy. Some patients would need no more than 5 injections (low demand) but some would need more than 16 injections (high demand) over the 24 months period. Several research groups have been working on predicting the response to anti-VEGF treatment. In the recently published study, Gallardo and colleagues used a random forest classifier with 100 trees and maximum tree depth of 100 to predict the probability of the long-term treatment demand [16]. The models used morphological features automatically extracted from the imaging at baseline and after 2 consecutive visits, as well as patient demographic information. The nAMD-trained models yielded mean AUCs of 0.79 for both low and high demand using 10-fold cross-validation.- Models for RVO and DME showed similar results, with a mean AUC of 0.76 and 0.78 for low and high demand, respectively [16].

Regulatory considerations

There is a high potential for application of prognostic and predictive ML-based algorithms for population enrichment in clinical development. It is important to consider the regulatory requirements as the ML applications move from exploratory to confirmatory regulatory setting. The trade-offs between the range of potential intended use scenarios and clinical development strategies need to be considered in the selection

of AI/ML approaches. In order to define a population for clinical trials, one could use predictors identified by a model as inclusion / exclusion criteria. As long as selected predictors are consistent with current clinical knowledge of the disease, the use of this enrichment approach may be straightforward from the regulatory perspective. Another approach is to select patients based on an algorithms' classification or scores. In this case, a rigorous validation plan needs to be developed before this algorithm could be applied in drugs' registrational trials. This is even more critical if there are aspects of novelty in predicting factors or their combination (compared to current clinical knowledge).

Similar to "traditional" biomarkers in clinical development, "digital" biomarkers could be included in IND/NDA/BLA process if they are specific to individual drug development program, or submitted through a biomarker qualification process, e.g. if they are independent of an individual drug development program. Readers interested in an example of Qualification pathways for digital biomarkers could refer to the EMA approval of stride velocity 95th centile measured at the ankle as an acceptable secondary endpoint in pivotal or exploratory drug studies for regulatory purposes and for quantification of a patient's baseline performance in such studies[17]. Readers interested in FDA biomarker qualification framework could refer to FDA guidelines [18] and the publication by Amur et co-authors [19].

Similar to clinical development approaches with "traditional" biomarkers, adaptive population enrichment design could be considered to leverage ML-based population enrichment as de-risking strategy if the efficacy in the overall population is below expectations (see Figure 4 below).

Different approaches to cross-validation are used during development and testing. This could be sufficient in exploratory setting to inform internal decisions and clinical development strategies. Moving to confirmatory setting and market launch preparation, an external validation would be required, i.e. a "frozen" algorithm should be validated against pre-defined performance targets in a new dataset. In order to reduce risk of failures related to an algorithms' performance in Phase 3 during trials and in post-market setting which could jeopardize the drug success, a number of considerations can be given to the selection of the validation dataset. For example, single site or single country data are often used for ML algorithms development, however using multi-sites and multi-country data for the validation could ensure its generalizability for Phase 3 trials and commercialization. It is also important to ensure that the demographics and clinical practice are in line with intended use. If ML-based algorithm is intended for stand-alone approval, a sponsor could purse SaMD pathway regulated by FDA CDRH. It should be noted that if particular devices are used to generate outputs used in an ML algorithm, regulators will want to understand if the results need to be restricted to only those particular devices. Readers could refer to respective FDA CDRH framework and guidance [20,21]. To date, FDA approved AI/ML-based SaMD that are based on "locked" ("frozen") algorithms. However, a new regulatory framework is designed to update traditional paradigm to Total Product Lifecycle (TPL) approach is under development [21]. It is important to consider that a documentation and quality system according to SaMD requirements needs to be established by a pharmaceutical / biotech sponsor engaged in ML-based algorithms development. In Europe, SaMD are reviewed by designated Notified Bodies. The European regulatory framework is evolving, and it is expected that most SaMD would require regulatory approval. In February 2023, the European Parliament adopted a proposal by the European Commission to delay the transition to the Medical Devices Regulation (MDR)[22,23].

The selection of ML methods should also be considered in view of regulatory strategy and future acceptance by clinical practice. Deep learning (DL) is a very attractive approach to develop a potentially better performing models. DL is well established in imaging applications, however its use in applications such as clinical risk prediction or classification is challenging because the "black box" nature limits regulatory and clinical acceptance of DL. ML methods that result in "transparent" algorithms with clinically explainable and clinically meaningful features & thresholds have lower bar for regulatory and clinical acceptance. It is also common that the initial or iterative selection of potential predictors among multiple potential factors is done based on clinical considerations (literature, exerts opinions etc). Likewise, pruning of a decision tree model (a type of model simplification) can be done based on clinical considerations.

Conclusion

Personalized medicine shifts clinical developments strategies from clinical trials with broad eligibility criteria to approaches that allow to tailor novel treatments to the individual characteristics of patients. ML has a potential in many areas of drug development, from drug discovery to clinical trials. With increasing availability of good quality data, different types of measurements and clinical characteristics could be combined into AI/ML-based algorithms to help to identify patients who are more likely to develop a particular outcome or who are more likely to respond to a particular drug. The trade-offs between the range of potential intended use scenarios and clinical development and regulatory strategies need to be considered in the selection of AI/ML approaches.

References

1. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices. Assessed 25.02.2023

2. Topol E. High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, Vol. 25, 2019: 45-46.

3. https://www.iqvia.com/insights/the-iqvia-institute/reports/global-trends-in-r-and-d-2023. Assessed 25.02.2023

4. Liu Q, Huang R, Hsieh J et al. Landscape Analysis of the Application of Artificial Intelligence and Machine Learning in Regulatory Submissions for Drug Development From 2016 to 2021. Clin Pharmacol Ther, 2022 Jun 16, doi: 10.1002/cpt.2668.

5. https://www.fdanews.com/ext/resources/files/10/10-28-13-Personalized-Medicine.pdf. Assessed 25.02.2023

6. https://www.iqvia.com/library/infographics/biomarker-testing-in-oncology Assessed 25.02.2023

7. BEST (Biomarkers, EndpointS, and other Tools) Resource FDA-NIH Biomarker Working Group. Silver Spring (MD): Food and Drug Administration (US); Bethesda (MD): National Institutes of Health (US); 2016.

8. *https://www.accessdata.fda.gov/drugsatfda_docs/nda/2020/214787Orig 1s000 ClinpharmR.pdf.* Assessed25.02.2023

9. Chaix M, Parmar N, Kinnear et al. Machine Learning Identifies Clinical and Genetic Factors Associated With Anthracycline Cardiotoxicity in Pediatric Cancer Survivors. JACC: CardioOncology, vol. 2, no. 5, 2020: 690 – 706.

10. Hyland S, Faltys M, Hüser M et al. Early prediction of circulatory failure in the intensive care unit using machine learning. Nature Medicine volume 26, pages 364–373 (2020)

11. Bradley J, Schelbert E, Bonnett L et al. Predicting hospitalisation for heart failure and death in patients with, or at risk of, heart failure before first hospitalisation: a retrospective model development and external validation study. Lancet Digit Health 2022; 4: e445–54

12. Arcadu F, Benmansour F, Maunz A et al. Deep learning algorithm predicts diabetic retinopathy progression in individual patients. npj Digital Medicine (2019) 2:92

13. Ting D, Cheung C, Nguyen Q et al. Deep learning in estimating prevalence and systemic risk factors for diabetic retinopathy: a multi-ethnic study. npj Digital Medicine (2019) 2:24

14. Kong J, Ha D, Lee J et al. Network-based machine learning approach to predict immunotherapy response in cancer patients. Nature Communications, 2022, 13:3703.

15. Seymour C; Kennedy J, Wang S et al. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. JAMA. 2019;321(20):2003-2017.

16. Gallardo M, Munk M, Kurmann T et al. Machine Learning Can Predict Anti-VEGF Treatment Demand in a Treat-and-Extend Regimen for Patients with Neovascular AMD, DME, and RVO Aassociated Macular Edema. Ophthalmology Retina 2021; 5:604-624.

17. *https://www.ema.europa.eu/en/documents/scientific-guideline/qualification-opinion-stride-velocity-95th-centile-secondary-endpoint-duchenne-muscular-dystrophy_en.pdf*. Assessed 25.02.2023

18. *https://www.fda.gov/regulatory-information/search-fda-guidance-documents/biomarker-qualification-evidentiary-framework*. Assessed 25.02.2023

19. Amur S , LaVange L, Zineh I et al. Biomarker Qualification: Toward a Multiple Stakeholder Framework for Biomarker Development, Regulatory Acceptance, and Utilization. Clin Pharmacol Ther . 2015 Jul;98(1):34-46.

20. *https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device*. Assessed 25.02.2023

21. FDA, Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) - Discussion Paper and Request for Feedback, 2019 *https://www.fda.gov/media/122535/download* Assessed 25.02.2023

22. *https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13684-Extension-of-the-transition-period-for-medical-devices_en*. Assessed 25.02.2023

23. *https://www.raps.org/news-and-articles/news-articles/2023/2/european-parliament-votes-to-extend-mdr-transition*. Assessed 25.02.2023 ■

A FRAMEWORK TO ACHIEVING QUALITY IN PHARMACEUTICAL THERAPEUTICS AND VACCINE DEVELOPMENT

Stan Altan (Janssen R&D, LLC), Dave LeBlond (Robert Singer Consulting), Tim Schofield (CMC Sciences, LLC)

Abstract

Recent regulatory and industry discussions on the concept of "patient-centric" specifications (PCS) have accelerated as companies pursue more efficient and harmonized "patient focused" drug development . The term "patient-centric" is understood as having many features in common with the term "clinically relevant". We build on recent industry discussions tying together the disparate domains of drug development. Each of these domains plays an important role in achieving the end goal of establishing PCS, with the goal of ensuring patient safety and efficacy. We also discuss the role of important stakeholders, the value of designs for clinical, process, formulation and analytical development, the technical roles of scientists and statisticians across the various development domains. We envision benefits to establishing greater synergy and communication among these domains. We propose that product development be viewed as a single unified framework, acknowledging the principle that PCS assures that product quality is built into the process and the product.

I. Introduction

The quality of pharmaceutical therapeutics and vaccines has been commonly associated with consistency of commercial manufacturing. Some efforts have been made to link quality to clinical trials, but these have been at best inadequate and at worst impossible due to disparate goals among domains within a company as well as regulatory expectations on clinical trial materials. The lack of common principles and standards of practice has led to inconsistencies in quality of licensed products across global markets, mostly due to differences in the definition of quality.

More recently, however, efforts are being made to address global harmonization through adherence to a common principle for quality. That principle is focused on decision rules such as specification limits which are used to release materials into the market. This principle has been referred to by various names such as "clinically relevant specifications" and "patient-centric specifications," all of which have refocused the basis of quality from consistency of manufacturing to product performance in patients.

Any one of us is exposed to this concept when being prescribed or purchasing a pharmaceutical or vaccine when we are told that the product "contains" a certain potency (or the amount of the active primary ingredient). What we don't commonly know is the range around this claim of quality consistent with clinical benefit. As a consumer of that drug or vaccine, we should hope that the true amount or potency is in a window that ensures satisfactory safety and efficacy. This patient expectation should be the driver for rules which lend assurance to patient centric quality and should be a driver for product development.

This article will articulate a framework for developing products which meet patient expectations related to quality. This will be restricted to measures of quality which are related to product safety and efficacy. Attributes related to characteristics such as taste and convenience – or characteristics currently mandated by compendial or other standards - will not be considered. A case will be made that this would best be accomplished by expanding the province of specifications-setting from the analytical domain of a company to all of development, including clinical, biomarkers, process, formulation, and analytical.

Under such a framework, developing PCS is the joint responsibility of scientists and statisticians in all these development domains. With the common mission of developing and licensing safe and effective pharmaceutical and vaccine products, these domains work together to perform studies which link desired outcomes

in patients to the analytics and process controls which ensure those results.

There are three important caveats in implementing a PCS framework. First, some specifications are dictated by compendia or from common experiences that are documented in the literature. This article will address specifications which are not dictated by such standards, or which are used to control drug product to levels which are otherwise justified. Second, the process of setting specifications is commonly reserved until the end of drug or vaccine development. This is due to the viewpoint that specifications should be linked to the lots manufactured using the final drug substance and drug product processes. This article advocates for using development experiences or novel studies to set commercial specifications. Third, PCS need not be justified in clinical studies but might be informed in preclinical or in vitro studies whose designs include input from both the clinical (e.g., biomarkers) and chemistry, manufacturing, and control (CMC) domains.

We will proceed with some terminology and concepts which will be used to communicate the elements needed to progress to PCS, followed by a viewpoint on the roles and responsibilities of development domains in that undertaking. From there we give a few examples of CMC and clinical designs which might be used to support PCS. We end with a summary in which we describe some conflicts and hurdles along with a proposal for a path forward to achieving PCS.

II. Definitions and concepts

Numerous terms and concepts are related to PCS. A list of these will help in both framing the topic and revealing components for consideration.

Terms related to the clinical domain

Clinical domain – the collection of clinical and laboratory functions that are related to evaluating patient outcomes.

Patient centric specifications (PCS) - limits on product critical quality attributes (CQAs) that are properly set and scientifically justified to assure the fitness for use (i.e., safety, efficacy, and availability) to the patient of the final pharmaceutical or vaccine product. Note that currently limits on CQAs are historically considered the sole responsibility of the CMC domain.

Clinical outcomes – patient safety or efficacy measures which are the subject of product claims and resolution in clinical studies..

Target Product Profile (TPP) - the target claims (including efficacy and safety) for a drug or vaccine leading to a clinical development plan. The TPP could become a place to include trials to support PCS.

Biomarkers – laboratory measurements which are predictive of a clinical outcome; sometimes called surrogate markers (e.g., in vitro dissolution, PK or immunogenic blood levels, etc.).

Terms related to the CMC domain

Chemistry, manufacturing, and control (CMC) domain – the collection of functions responsible for developing the manufacturing and control processes for a new drug or vaccine.

Quality Target Product Profile (QTPP) – the target commercial requirements such as shelf life and specifications which drive CMC development (note: there has been much discussion about whether specifications belong in the QTPP; it will be the position of this article that they should be in order to have a technical basis for development).

Critical quality attributes (CQAs) - measurable properties of a drug or vaccine (e.g., potency) that are thought (or demonstrated) to be linked to patient outcomes. Note that the study of patient outcomes is historically considered the sole responsibility of the clinical domain.

Critical process parameters (CPPs) – process parameters which when varied impact one or more CQAs and must therefore be controlled to ensure appropriate product quality. CPPs also include critical material attributes (CMAs).

Quality by design (QbD) – principles and practices that have been summarized in a series of ICH quality guidelines (Q8, Q9, Q10, Q12), which outline the pathway towards "building quality into a product," and maintaining quality throughout the product lifecycle. At the core of QbD is the definition of quality, the concept of fitness-for-use, and the steps necessary to deliver a com-

mercial control strategy (the engineering and analytical formulae and post-licensure protocols for maintaining product quality).

Design space – A multivariate region of CPP space within which CQAs are assured to be within their respective specification limits with high probability.

Commercial – the combined manufacturing and quality control functions which produce and control (release and stability) a commercial product; other activities related to quality are collected into a process called lifecycle management.

Reportable result – a summary value of one or more analytical determinations of a quality attribute intended for a regulatory purpose, for example, biological potency on a certificate of analysis.

Jointly owned terms

Value chain - the process or system of interrelated activities by which a company adds value to an article.

III. Fitness for use

A special term associated with QbD that is commonly used in CMC applications is "fitness for use." This serves as the goal of a QbD pathway in development. Thus, for example, an analytical procedure is fit for use if it meets its requirements in making an accurate decision (e.g., with respect to release of a drug product lot, shelf life estimation, or reagent bridging). Those requirements are codified in an analytical target profile (ATP) where limits on bias and variability (or their combination, total error) are determined to control the risk of an incorrect decision when using the procedure. The reliability of the reportable result is assured by limits placed on key suitability metrics (e.g., model parameter estimates) and through design considerations. At the execution level, the ATP requirements are in turn assured by establishing an analytical design space (also called the method operable design region) for procedural variables (e.g., incubation time, temperature). The key QbD concept is a strong chain made of functional linkages that maps the analytical method execution to decision maker's risk. In this paradigm, the ATP requirements are "customer centric" in the sense that they address the decision maker's risk of making the wrong decision.

The probability that the analytical reportable result will satisfy its required decision accuracy is analogous to the probability that the drug product will satisfy the safety and efficacy patient centric expectations. Analogous to the use of bias and variability as analytical surrogate metrics, biomarker laboratories develop procedures (biomarker assays such as cholesterol level, immunological response, and genomic readouts) which are used to make decisions about patient health (as a diagnostic) or as a surrogate endpoint in a clinical trial. Biomarker assays should meet their requirements for making accurate predictions regarding patient disposition or for clinical trial assessments. CQA limits for a drug product are analogous to the suitability limits for an analytical procedure. CPP limits for a drug product manufacturing process keep CQAs within their limits in the same way that analytical method procedure ranges assure that method performance will be at an acceptable level. In some cases, this may be an iterative process as additional knowledge and understanding are gained during the development process as implied by the double headed arrows in Figure 1.

QbD is typically associated with CMC practices. However, the science and risk-based approaches of QbD should resonate with all participants in all development domains. QbD with an appropriate definition of quality, ideally PCSs, brings all domains together to achieve a common goal of delivering safe and effective drugs and vaccines.

It's recommended, therefore, to apply the QbD concept of fitness for use to PCS based product requirements, or PCSs. Fitness for use of the product then becomes the "driver" for setting CPP ranges that predict efficacy and safety of a drug or vaccine, and through these for developing manufacturing and quality control to this standard.

IV. Development domains and their roles in assuring product quality

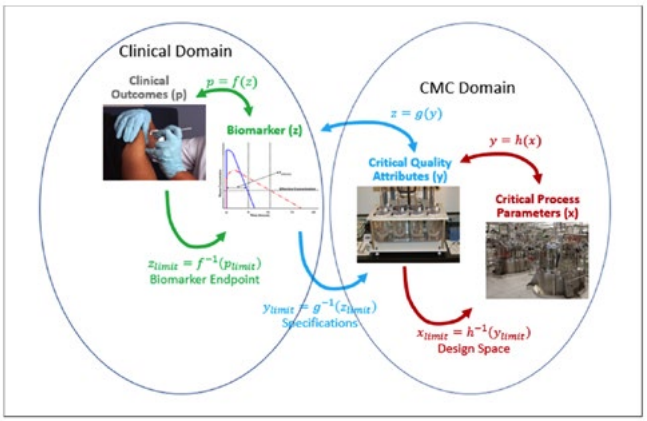
As illustrated in Figure 1, there are two major domains in product development, clinical and CMC. These are categorized by the types of information they contribute to the development process. For this article the clinical domain will be comprised of clinical, preclinical, translational medicine, and biomarker development. The CMC domain is comprised of process (synthesis,

upstream, and downstream), formulation, and analytical development. Regulatory affairs and quality assurance play a role in communicating and enforcing practices related to clinical and CMC development and lifecycle management. While devices and combination therapies (i.e., a drug supported by a companion diagnostics) might be included, they are not specifically considered here. This article focusses on those dimensions related primarily to drugs or vaccines and their respective clinical populations.

While there may be additional functions and different names for these functions, many companies operate along separate clinical and CMC development pathways (except as it relates to supplying materials for clinical trials), and with limited interactions between development and commercial divisions (except for technology transfers). The mission of clinical development is to demonstrate the safety and efficacy of a candidate drug or vaccine while the mission of CMC is to formulate a marketable drug product, necessary analytical measurement tools, a commercial manufacturing process and associated production strategy primarily based on process capability. These missions are often treated as independent and are sometimes in conflict. These will be discussed in the Summary section including Conflicts and Hurdles.

Figure 1 characterizes the roles of and relationships among domains and functions within each domain. A key aspect in Figure 1 is the nature of the information sharing between four key functions. Each function therefore has a role for both a scientist (call them a subject matter expert) and an information scientist (a data scientist and/or statistician). The double headed arrows represent the collaborations between the domains and the single headed arrows represent the development of decision limits for each domain.

Figure 1: Clinical and CMC domains covering the product development value chain



Two functions and the information they contribute are depicted in the *clinical domain*: *clinical development*, whose members design and implement clinical studies to support product efficacy and safety claims; and *biomarker development* whose members develop clinical markers and associated assays which are intended to be used as surrogate clinical responses. We use the term biomarkers synonymously with surrogate outcomes or measurements. Similarly, two functions comprise the *CMC domain: analytical development* whose members design and implement procedures used to characterize and control product; and manufacturing development whose members design the commercial synthetic, biotechnology, purification, and formulation processes. It is noteworthy that analytical procedures are used not only for characterization and control, but also to guide process and formulation development. In this sense it is important that these procedures either be carried out in a centralized laboratory or be bridged between functions to assure similar scales of information exchanged between the two functions.

The interfaces between functions offer the opportunities for collaboration on translation of information across the value stream. These are briefly described as follows:

- Clinical studies can be performed, or data collected from routine studies to evaluate the relationship between clinical outcomes (p) and a biomarkers (z). For older biomarkers (e.g., sugar level related to diabetes) this relationship may already be known. For newer diseases such as COVID these relationships must be built during an accelerated clinical program. These relationships may likewise be developed in preclinical models when a validated animal model is used to predict human outcomes. The relationship between clinical outcome and biomarker result is represented as $p=f(z)$.
- Likewise, studies can be performed that describe the relationship between a biomarker response (z) and a critical quality attribute (y). This is common in vaccine development where the mechanism of action of the vaccine is well understood (e.g., sterilizing or cell mediated immunity). In these cases the biomarker can be used initially to bridge clinical populations (e.g., by sex, age, and race) or when there is an

important process or formulation change. This relationship is depicted as $z=g(y)$. In solid dose products, IVIVC studies can serve a similar purpose. In a later example, we will show how this can be done in practice, combining a process study linked to pK outcomes.

- Within the CMC domain, models can be built which relate CPPs (x) to CQAs (y). These are sometimes developed during process characterization and apply to formulation as well as API or drug substance development. With either, the link can be denoted as $y=h(x)$.

The functions f, g, and h correspond to predictive models (which may be multivariate) that map variable values from one sub-domain to another. These form the links in the value stream. Having established one or several of these relationships provides the context for the value stream and for defining important decision criteria, including PCS. Working from clinical outcomes towards CPPs, a biomarker endpoint can be determined from the relationship $p=f(z)$ using traditional inverse function notation: $z_{Limit}=f^{-1}(p_{Limit})$, where p_{Limit} is a clinical product claim (e.g., 95% efficacy against severe disease) and z_{Limit} denotes a biomarker limit within which the p_{Limit} claim can be realized with high probability (see The role of statistics in ensuring quality).

Similarly, patient centric CQA specifications can be established from the inverse relationship between the biomarker limit z_{Limit} and the CQA specification y_{Limit}. Finally, the h model can provide a design space that describes the CPP ranges, x_{Limit}, that assure that the CQA specifications, y_{Limit}, can be met with high probability.

It is important to note that this is a simplification of a complex modeling process. But Figure 1 serves to illustrate that careful scientific communication between organizational functions can create links in a value chain that coordinates and synergizes clinical study and CMC testing designs.

Widely used models relating CPPs to CQAs can be used for two purposes. As discussed below in section V1, one use is to justify the selection of CPPs and the other is to use the relationship to define the process or formulation design space. A CPP is defined as a process parameter from which variation has *meaningful impact* on one or more CQAs. Here meaningful impact can be taken to mean that lack of control of a given CPP could cause a substantial enough shift in a CQA to increase the risk of exceeding its specification limit. The design space is the multivariate combination of levels of CPPs which predict that associated CQAs will remain within their specifications with high probability.

Finally, PCSs need not be addressed solely in clinical trials. Validated animal models and advanced *in vitro* systems (e.g., organ on a chip) can be employed to explore a range of levels of CQAs which are either directly translatable or which include adjustments to define levels of impurities and doses which are predicted to be safe and effective in patients. This approach is accepted for known toxicological impurities with well-established maximum exposure levels (ICHQ3B, 2006) and formulae that address important factors in a patient population (e.g., weight, sex, etc.).

V. The role of CMC statistics in ensuring quality

CMC statisticians are familiar with a process variability (distribution) approach to setting specifications. Specifications derived from process variability considerations have at least two important purposes: (1) to evaluate process capability against appropriate PCS; and (2) to serve as the independent basis of monitoring manufacturing consistency (Stage 3 Process Validation - continued process verification). Both purposes can have positive business impacts, but do not directly support determining a PCS.

It is acknowledged that a shift from process variability based specifications to PCS could be disorienting to CMC statisticians and their analytical, process, and formulation colleagues. Within the CMC domain, process variability limits are often derived from statistical intervals obtained from relevant batch data within their immediate purview. Which statistical interval to use (3-sigma, β-tolerance intervals β-γ-tolerance intervals, Bayesian) and which confidence/probability coefficient (i.e, β-γ) best reflects an appropriate risk level, are often the issues of debate, rather than the impact of CPP limits and CQA specifications on patient centric considerations. The issue is that the resulting calculations should not be used as CQA product specifications or CPP process limits because they do not assure patient safety or efficacy.

There is a more relevant role for statisticians (from all domains) in supporting a PCS approach. This is

to apply their statistical modeling expertise to help develop and characterize the predictive uncertainties in the *f*, *g*, and *h* (or other) models that form the links in the value chain. But these models are not merely an end in themselves. These models should, in principle, be leveraged jointly to obtain predictions and associated uncertainties to help manage the risks associated with propagating patient outcomes through to manufacturing controls. Statisticians can work with subject matter experts to design studies which minimize these uncertainties to some target risk level. The resulting integrated knowledge base can help identify and reduce the major sources of overall process variability to minimize loss from product failures or identify manufacturing formulae changes (such as overfill) which compensate for the variability. If designed properly, this knowledge can also point to other sources of variability, helping to refine the manufacturing control strategy.

CMC statisticians can contribute further to the goal of establishing PCS by using early modeling to forecast manufacturing variability and stability change of a drug or vaccine which is under development. Early development data from the Process Design phase for a candidate product together with data from similar commercial products can use Bayesian methods to predict the distribution of CQA values in the commercial product. This can be used to inform dose ranging experiments or in special studies to target clinical doses which represent the breadth of exposure to a critical attribute.

VI. An example of a patient centric QbD study for a solid dose compound in two Stages

An example is provided here of a design of studies used to develop some of the functional relationships illustrated in Figure 1, first linking CQAs and CPPs, and then CPPs to pK parameters. A conventional tool used in CMC development, design of experiments or DoE is used to show its power in resolving complex relationships across pharmaceutical domains.

Briefly, DOE is tool which relates controlled input factors to output variables. A familiar case is the function *h* in Figure 1 where one or more critical quality attributes (*y*'s) are expressed as a function of critical process parameters (*x*'s). DOE is a formal way to design a study to resolve this function using 2 or more levels of each factor (e.g., pH equal to 6.8 and 7.2) crossed in combinations with the same for other factors. In the simplest case this could be a study with 2k combinations of k-factors each at 2-levels. But as k increases, the laboratory might use fractions (e.g., ½ of the full design called a fractional factorial design) to manage resources. The function resulting from the analysis of DOE data is a linear function with coefficients on main effects (i.e., the effect of each factor alone) and interactions (i.e., the combined effects of 2 or more factors). With this, the inclusion of center points (i.e., points in the middle of the factor levels, usually representing an initial target), and the use of more than 2 levels of factors, the resulting function becomes a polynomial approximation to a more complex nonlinear relationship between the *x*'s and the *y*'s. Complications arise when the full design is over-fractionated leading to the inability to uniquely estimate some important coefficients. This is not because the coefficient isn't estimable but instead because it is shared with other factors (called confounded or aliased with other factors or their interactions). This condition can be improved using a technique called augmentation where additional combinations are added to the factorial design to provide data that can be used to uniquely estimate important main effects and interactions. A more complete description of experimental designs can be found in Montgomery (2019).

a. Step 1 – DoE 1

Consider an experiment studying the impact of CPPs on a single CQA (in vitro dissolution) as described in Table 1, showing an augmented ½*2^4 factorial design with center point and 5 replicates in a total of 16 runs. We use standard notation where -1, 0 and 1 represent the low, mid and high levels of the factors. The factors are comprised of 2 manufacturing process factors, Compression Force (CF), Blend Time (BT), and 2 material factors, Particle Size (PS), Binder Concentration (BC). For the purposes of this example, we will assume that the single CQA, dissolution has a working acceptance limit which can be utilized to identify the active factors, i.e. those factors whose "effect" is substantial. We hasten to add that if a PCS is available, then those can be used to establish a "safe space" (Zhao and Suarez, 2019) for the acceptance limit. But there will be cases when a PCS is not yet known, so in these cases, an iterative process would be followed to move on to DoE 2, and then return to Step 1 as additional information from DoE 2 becomes available, allowing a formal Design

Space construction. As the development process continues, and clinical data is collected and related to DoE 2 outcomes, we envision that this in turn would inform further refinement of the *g* and *h* models.

Table 1 – ½ 24 augmented factorial design

Run	CF	BT	PS	BC	Response (dissolution)
1	-1	1	-1	1	y_1
2	0	0	0	0	y_2
3	1	-1	-1	1	y_3
4	-1	-1	1	1	y_4
5	1	-1	1	-1	y_5
6	0	0	0	0	y_6
7	1	1	1	1	y_7
8	1	1	-1	-1	y_8
9	1	1	-1	1	y_9
10	-1	1	-1	-1	y_{10}
11	-1	-1	-1	-1	y_{11}
12	1	-1	-1	1	y_{12}
13	-1	1	1	-1	y_{13}
14	-1	-1	1	1	y_{14}
15	1	1	-1	-1	y_{15}
16	1	-1	1	-1	y_{16}

The linear statistical model includes a term for 4 main effects and the 2 interactions of interest plus an error term. A design space equation based on either working acceptance limits or PCSs can be derived from the intersection of the limits and the statistical model adjusting for uncertainty. The design space equation can be described by a hierarchical Bayesian model, whose posterior predictive distribution leads to a risk profile of acceptable product. The risk profile of a possibly multidimensional experimental space describes a region that produces the required product quality with a desirable level of probability (Peterson, 2008). It is noteworthy to point out that this approach could result in a highly restrictive design space and excessive predicted risk if both manufacturing and analytical measurement variability are incorporated into the acceptance or specification limits. In this example, there are 5 design combinations with replication, so a manufacturing process variance component with 5 degrees of freedom can be estimated. The main deliverable of the Step 1 experiment is to define the CPPs and the functional relationship described as *h* in Figure 1 .

b. Stage 2 – DoE 2

Table 1 provided an example of a simple manufacturing process and material study. The primary goal of that design is first to identify the active factors, and then later to establish the functional design space (*h*). We can build on this example to show how a statistical design can be used to broaden the relationships (including both *h* and *g*) between process parameters, CQAs and the surrogate outcomes, say one or more pK parameters, AUC, Cmax, and Tmax.

For purposes of demonstration of the statistical principles, let us assume that the DoE 1 model *g* for dissolution reduces to 3 main effects, and 2 interaction terms. An appropriate design was chosen as shown in Table 2 to investigate the effect of the 3 meaningful process/material parameters, linked to the CQA (*y*, in vitro dissolution), on the surrogate outcomes (*z*'s, the pK parameters), thus establishing the functional relationship *g*, or possibly a composite function (*g*+*h*). The y_i*'s represent least squares means calculated from DoE 1. Another option for DoE 2 is to simply carry into DoE 2 those specific design combinations from DoE 1 that appear sufficiently different from the center point based on a statistical comparison to a scientific criterion.

Table 2 Treatment Combinations for pK Study

Treatment Combination	Factor			Response pK parameters	Response Dissolution From DoE 1
	CF	BT	PS		
1	-1	-1	1	z_1	y_1*
2	-1	0	-1	z_2	y_2*
3	-1	1	1	z_3	y_3*
4	0	0	0	z_4	y_4*
5	1	-1	-1	z_5	y_5*
6	1	-1	1	z_6	y_6*
7	1	1	0	z_7	y_7*

Given the definitions of treatment combinations in Table 2, Table 3 shows a 4 -period crossover design with 7 subjects, where the impact of the process/material parameters on pK parameters, linked to one or more CQAs, can be studied with a balanced incomplete block design (BIBD). For a brief explanation of crossover designs, see Cochran and Cox (1957). The design was generated with the r function GEN_EYD, described in Hua et al (2021).

Table 3 Treatment Allocation by Period and Subject

Subject	Period			
	1	2	3	4
1	3	5	1	4
2	0	3	6	1
3	5	6	4	0
4	1	2	5	6
5	2	0	3	5
6	4	1	0	2
7	6	4	2	3

A between- and within-subjects model can be used to estimate treatment means, with an error term having 12 degrees of freedom. If this is insufficient precision, then we can consider expanding the size of the study to 14 subjects, and still preserve the orthogonal properties of the 7 subject design. It would therefore be prudent to consider this in the design of the study protocol, and describe it appropriately in any statistical analysis plan. Once the means are estimated with sufficient precision, one can model the pK parameter means plus dissolution rate as a multivariate response directly against the process parameters. This would be followed by a conditional regression of the pK parameters against the CPPs and dissolution rate.

The correspondence to Figure 1 across the two experiments is:

y^* (dissolution) = $h(x)$ where x = CPPs,

z (pK outcomes) = $g_c(y^*,x)$ where y^* = in vitro dissolution and x = CPPS, through a conditional regression (see Graybill, 1976) for additional description of conditional regression modeling.

A Bayesian approach can be used to classify the least squares means as a way to interpret the effect of the process parameters in relation to some meaningful clinically relevant criterion. Knowing the factor effects then can guide the construction of a Design Space that ensures a PCS.

VII. An example of a patient centric specification for vaccines

The concept of PCS has been acknowledged in vaccines for some time. Notably a combined measles, mumps, and rubella vaccine that was licensed in the mid-1900's used small clinical trials to establish doses which were adequately safe and effective, and then added a factor to the release limit which was adequate to compensate for the potency loss of each antigen over its shelf life.

This concept has been formalized in more recent vaccine development programs and built into vaccine clinical studies. This is best represented by studies with an "end of expiry" dose (usually potency) of a vaccine. Like the ad hoc approach used for historical vaccines early stability data for a candidate product can be used to forecast the change over a commercially viable shelf life (as specified in the QTPP). The end-of-shelf-life (EOSL) requirement can be supported by CMC statisticians first using modeling and simulation of early estimates of product variability and stability change, including the predicted uncertainty of the change from a planned design of the supporting stability study. This can be further formalized as a release limit, using a patient-centric specification representing the EOSL requirement on quality and adjusting this by the combination of the change in quality which has been estimated from product stability studies, the standard error of the change, and the variability of the release assay [Allen & Dukes; WHO Guidelines on Stability Evaluation of Vaccines; Schofield]. As with design space determination, Bayesian methods can be used together with prior knowledge of the release modeling parameters to improve the prediction.

An experimental lot can be prepared at the target dose or a clinical lot can be strategically degraded using mild elevated temperatures to obtain the expected end of expiry potency. A routine noninferiority clinical trial using a clinical biomarker can be designed to demonstrate that the altered material elicits similar immunological response as the unaltered material. The trial can be designed as a pseudo-efficacy study when there is a correlate of protection (i.e., a level of a vaccine biomarker that is predictive of efficacy). The predicted level of efficacy can be held to the target level specified in the product TPP.

VIII. Summary

The goal of the joint activities described in Figure 1 is to define manufacturing controls ensuring quality product to patients. Lack of manufacturing control or poor process capability (low probability of success in releasing product to the market) can result in product supply shortages which may put patients at risk. The FDA process validation guidance[1] includes stages directed towards manufacturing control. While Stage 1 (process design and development) has been illustrated with the factorially designed study above, Stage 2 (process qualification) and Stage 3 (continued process verification) are meant to demonstrate and continuously monitor process capability. The scientific and statistical basis of process qualification is questionable due to the short period of time (i.e., experience) and small number of lots, while continuous performance verification will be hampered by frequent investigations and diminished supply if specifications are based on manufacturing variability. This is further complicated by the expectation that product be of adequate quality throughout its shelf life and the need to evaluate product quality after a routine (or unexpected) change in the process or analytical method. It may be at this time that a company, and regulatory authorities who require that specifications be established from manufacturing variability realize the value of having established broader patient-centric specifications. The vision of a workable specification at time of regulatory filing is now being actively discussed in the context of utilizing platforms and of integrating "prior knowledge" into the specification setting process. Statisticians will be challenged to bring a Bayesian perspective on this effort.

As noted another important quality consideration in developing a commercial drug product is its stability. There are many different aspects of stability studies, but essentially the goal is to study the effects of environmental stress factors, such as temperature, humidity, and light, on the physical and chemical properties of a drug product's final market configuration. One objective of the formulation studies is to produce a "shelf stable" formulation. Such a formulation permits the ability to store at acceptable conditions (say 5C. for biologics, 30C/75% Relative Humidity for small molecule solid dose products) for a sufficient period of time to support a cost effective marketing strategy. Typically, this would be 24 months, although it can be shorter or longer in some cases. Considerations such as regulatory testing and excursions of temperature labile products should also be considered.

Other aspects of the stability studies that could have clinical implications are potency changes, formation of impurities on storage, and physical changes that could impact rates of systemic absorption and bioavailability. These are relevant to both clinical and preclinical safety studies. The statistical modeling and design of stability studies of the active molecule are discussed in ICH Q1E, where a confidence interval approach applied to a fixed effects linear model is suggested for a regulatory filing. Some products like vaccines use stability data to determine a release specification (see WHO Guideline on Stability Evaluation of Vaccines). In this case the objective of both long term and accelerated stability studies is to obtain a reliable estimate of the rate of degradation of the product.

Recent work suggests a Bayesian mixed model, acknowledging process variability and analytical uncertainty, has attractive statistical features that recommend its use in assigning a shelf life to a drug product (Sontag, et al., 2023). In addition, accelerated stability studies modelled using nonlinear approaches based on an extended Arrhenius relationship, are gaining in common practice (see Porter et al., 2018). This is an area that is evolving as the science advances and technology drive newer approaches that bring greater efficiency to the value chain, enhanced by QbD.

The expectation that measurements must fall within specification limits throughout shelf life has served as a disincentive for robust stability study design (see FDA

OOS Guideline). Revisions to guidelines like ICH Q1E should dispel this misconception and pave the way to better design and analysis of stability studies.

QbD and associated process validation stages should be understood as a logical and seamless series of activities, involving the science, engineering knowledge and statistical modeling and risk computations. The knowledge gained in each stage is built on to design the next stage. Consequently, validation protocols are developed specific to the level of understanding of the product and the process at the given stage, acknowledging the risks associated with the state of knowledge. These can be thought of as an extension of the basic principles of QbD, where process understanding and science drive the nature of the studies and quality control strategies. Quality risk management as described in ICH Q9 plays an important role throughout the stages of process validation.

In summary a framework to achieving quality in pharmaceutical therapeutics and vaccine development involves CMC, preclinical, and clinical statisticians participating in science and risk based (QbD) development through designs and analyses of studies which help define quality and build this into the product control strategy. Greater emphasis on "building quality into the product" takes precedence over regulatory validation, where effective use of information includes formal assessments of risks. Risks to patients are evaluated against a patient-centric definition of quality rather than against consistent performance of the production process, while the latter should be managed by the manufacturer to improve product knowledge and to reveal opportunities for manufacturing and analytical improvements. Much of the discussion in this paper has focused on a coherent organization and flow of information across subject matter domains, but this remains in many respects a vision of a future state. The isolation of disciplines, and misaligned missions between clinical and CMC require in some cases a cultural change in business practices intersecting scientific and technical activities. It is exacerbated by ambiguous guidelines and the lack of standards of practice, including the use of Bayesian methods in the CMC domain. CMC should follow the clinical and devices domains in supporting regulatory guidance for Bayesian methods in designing clinical trials (Ionan, 2020; FDA Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials).

References

1. Apostol, I., Schofield, T., Koeller, G., Powers, S., Stawicki, M., Wolfe, R.A. (2008). A Rational Approach for Setting and Maintaining Specifications for Biological and Biotechnology–Derived Products—Part 1. BioPharm International. 21(8). p. 40–45

2. Bercu, J., Berlam, S. C., Berridge, J., Cherney, B., Cowley, D., Laughton, H. W., McLoughlin, D., McMahon, M., Moore, C. M. V., Murti, C., O'Neill, J., Parsons, R., Peng, D. Y., Quan, R. W., Subashi, A.K., Teasdale, A., Tyler, S. M., Watson T. J. (2019). Establishing Patient Centric Specifications for Drug Substance and Drug Product Impurities. Journal of Pharmaceutical Innovation. 14. p.76–89 *https://doi.org/10.1007/s12247-018-9366-5*

3. Cochran, W. G. and Cox, G.M. (1957). Experimental Designs (Second Edition). John Wiley & Sons, New York

4. Committee for Human Medicinal Products (2022). Toolbox guidance on scientific elements and regulatory tools to support quality data packages for PRIME and certain marketing authorisation applications targeting an unmet medical need. Euopean Medicines Agency, Amsterdam, The Netherlands.

5. Dickens, Jason E. (2010). "Overview of Process Analysis and PAT", Chapter 1 in Bakeev, Katherine A., Editor. Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries, Second Edition, John Wiley & Sons, Ltd

6. FDA Guidance for Industry (2004). PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance. *https://www.fda.gov/media/71012/download*

7. FDA Guidance for Industry (2011). Process Validation : General Principles and Practices. *https://www.fda.gov/files/drugs/published/Process-Validation--General-Principles-and-Practices.pdf*

8. FDA Guidance (2018). Biomarker Qualification: Evidentiary Framework Guidance for Industry and FDA Staff DRAFT GUIDANCE

9. Graybill, F. A. (1976). Theory and Application of the Linear Model. Prindle Weber & Schmidt, Boston. ISBN 10: 0878721088ISBN 13: 9780878721085

10. Hua, y., Hedayat, A.S., Yang, M. and Altan, S. (2021). An Extended Youden Design in Biological Assays. Statistics in Biopharmaceutical Research. 13 (3).

11. International Conference on Harmonization (2003). Guidance Q1E Evaluation of Stability Data. *https://www.ema.europa.eu/en/documents/scientific-guideline/ich-q-1-e-evaluation-stability-data-step-5_en.pdf*

12. International Conference on Harmonization (2006). Guideline Q3B (Revision 3) Impurities in New Drug Products. *https://www.fda.gov/media/71733/download*

13. International Conference on Harmonization (2009). Guideline Q8(R2) Pharmaceutical Development. *https://database.ich.org/sites/default/files/Q8_R2_Guideline.pdf*

14. International Council on Harmonization (2021). Guideline Q9(R1, Draft) Quality Risk Management. *https://www.ema.europa.eu/en/documents/scientific-guideline/draft-international-conference-harmonisation-technical-requirements-registration-pharmaceuticals_en-1.pdf*

15. International Conference on Harmonization (2008). Guideline Q10 Pharmaceutical quality system. *https://www.ema.europa.eu/en/documents/scientific-guideline/international-conference-harmonisation-technical-requirements-registration-pharmaceuticals-human_en.pdf*

16. International Council on Harmonization (2019). Guideline Q12 Technical and Regulatory Considerations for Pharmaceutical Product Lifecycle Management. *https://database.ich.org/sites/default/files/Q12_Guideline_Step4_2019_1119.pdf*

17. International Council on Harmonization (2022). Guideline Q14 (Step 2b) Analytical Procedure Development. *https://www.ema.europa.eu/en/documents/scientific-guideline/ich-guideline-q14-analytical-procedure-development-step-2b_en.pdf*

18. International Council on Harmonization (2020). E9 (R1) Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. *https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf*

19. Ionan, A. C. (2020). Communicating Bayesian Design and Analysis to Statisticians. 2020 ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop

20. Junker, B., Zablackis, E., Verch, T., Schofield, T. and Douette, P. (2015). Quality-by-Design: As Related to Analytical Concepts, Control and Qualification. Chapter 12 in Vaccine Analysis: Strategies, Principles, and Control. Editors Brian K. Nunnally, Vincent E. Turula, Robert D. Sitrin *https://doi.org/10.1007/978-3-662-45024-6*

21. McAllister M, Flanagan T, Boon K, Pepin X, Tistaert C, Jamei M, Abend A, Kotzagiorgis E, Mackie C. Developing Clinically Relevant Dissolution Specifications for Oral Drug Products-Industrial and Regulatory Perspectives. Pharmaceutics. 2019 Dec 23;12(1):19. doi: 10.3390/pharmaceutics12010019. PMID: 31878006; PMCID: PMC7022466.Montgomer D. (2019). Design and Analysis of Experiments, 10th Edition. John Wiley & Sons, New York ISBN: 978-1-119-49244-3 2019

22. Peterson, J.J. (2008). A Bayesian approach to the ICH Q8 definition of design space. Journal of Biopharmaceutical Statistics, 18, p. 958 – 974

23. Porter, W.R., Chen, J., Altan, S. (2018). "Data Evaluation and Statistical Methods". Chapter 6 in Accelerated Predictive Stability (APS) Fundamentals and Pharmaceutical Industry Practices. Editors Fenghe Qiu, Garry Scrivens, Academic Press, Cambridge, MA. p. 147-174

24. Ruesch, M. N., Benetti, L. , Berkay, E., Cirelli, D. J., Frantz, N., Gastens, M. H., Kelley, W. P., Kretsinger, J., Lewis, M., Novick, S., Rellahan, B., Pack, L., Stroop, C. J. M., Subashi, A., Yin, P., Zeng, M., Stults, J. (2020). Strategies for Setting Patient-Centric Commercial Specifications for Biotherapeutic Products. J Pharm Sci. 110(2):771-784.

25. Sondag, P., Kim, J.Y., Natalis, L., Scherder, T. (2023). Chapter 12 Bayesian Alternatives to Traditional Methods for Estimating Product Shelf Life and Internal Release Limits in Chapter 12 Application of Bayesian Methods for Specification Setting in Case Studies in Bayesian Methods for Biopharmaceutical CMC, Edited by Paul Faya and Tony Pourmohamad. Chapman & Hall/CRC Press. Boca Raton. ISBN 9781032185484

26. Specification Design and Lifecycle Management Working Group of the PQRI Manufacturing Technical Committee (2012). Quality by Design Specifications for Solid Oral Dosage Forms: Multivariate Product and Process Monitoring for Managing Drug Quality Attributes Concept paper. *https://pqri.org/wp-content/uploads/2016/02/PQRI-QbD-Specs-Concept-Paper-Jan-2012-2.pdf*

27. Suarez Sharp, S. (2012). Establishing Clinically Relevant Drug Product Specifications: FDA Perspective. 2012 AAPS Annual Meeting and Exposition. Chicago, IL, October 16, 2012

28. Thompson, C. and Miro-Quesada, G. (2023). Chapter 13 Application of Bayesian Methods for Specification Setting in Case Studies in Bayesian Methods for Biopharmaceutical CMC, Edited By Paul Faya, Tony Pourmohamad. *Chapman & Hall. ISBN 9781032185484*

29. Woodcock, J. (2004). The concept of pharmaceutical quality. American Pharmaceutical Review. November 2004

30. Zhao, Y., Suarez Sharp, S. (2019). FDA expectations in building a safe space to gain regulatory flexibility based on Physiologically Based Biopharmaceutics Modeling (PBBM). REdI and M-CERSI Workshop 2019 Current State and Future Expectations of Translational Modeling Strategies to Support Drug Product Development, Manufacturing Changes and Controls Workshop, Sep 25, 2019 //cersi.umd.edu/sites/cersi.umd.edu/files/Day%203-1%20Zhao%20Suarez%20LM.pdf

THE BIOPHARMACEUTICAL STATISTICS LEADERS CONSORTIUM'S (BSLC) POSITION ON EFFECTIVE COLLABORATION BETWEEN STATISTICIANS AND DATA SCIENTISTS IN BIOPHARMACEUTICAL RESEARCH & DEVELOPMENT

Erik Pulkstenis, Xun Chen, Simon Davies, Pameljit Kalra, Becky Maksimovic, Christopher Miller, Jonathan Rowe, and Amy Xia
on behalf of the BSLC

1. Introduction

In recent years, the role of advanced data science capability has expanded within the biopharmaceutical industry. This shift brings forth an opportunity to improve insights and data-driven decisions, while also providing new opportunities for collaboration. However, there is also the potential for confusion between the role of statisticians, who have long been involved in data modeling, experimental design and interpretation of clinical data, and the role of data scientists. While there are some common skills across the two domains, each role brings a distinct set of skills to the table. This intersection and complementation of skills requires careful consideration of where and how to utilize these roles most effectively within an organizational model. While there are different choices in terms of organizational structures, there are some key success factors - high degree of collaboration, shared learnings, awareness of when to leverage complementary skills, and availability of well-designed software or systems. In addition to structure, leadership behaviors and culture are always important drivers of outcome. The purpose of this article is to discuss some of these factors and highlight solutions that support both statisticians and data scientists in biopharma to reach their full potential, allowing synergistic collaboration to help solve innovative business problems and improve patient outcomes.

2. Roles of Statisticians and Data Scientists within R&D

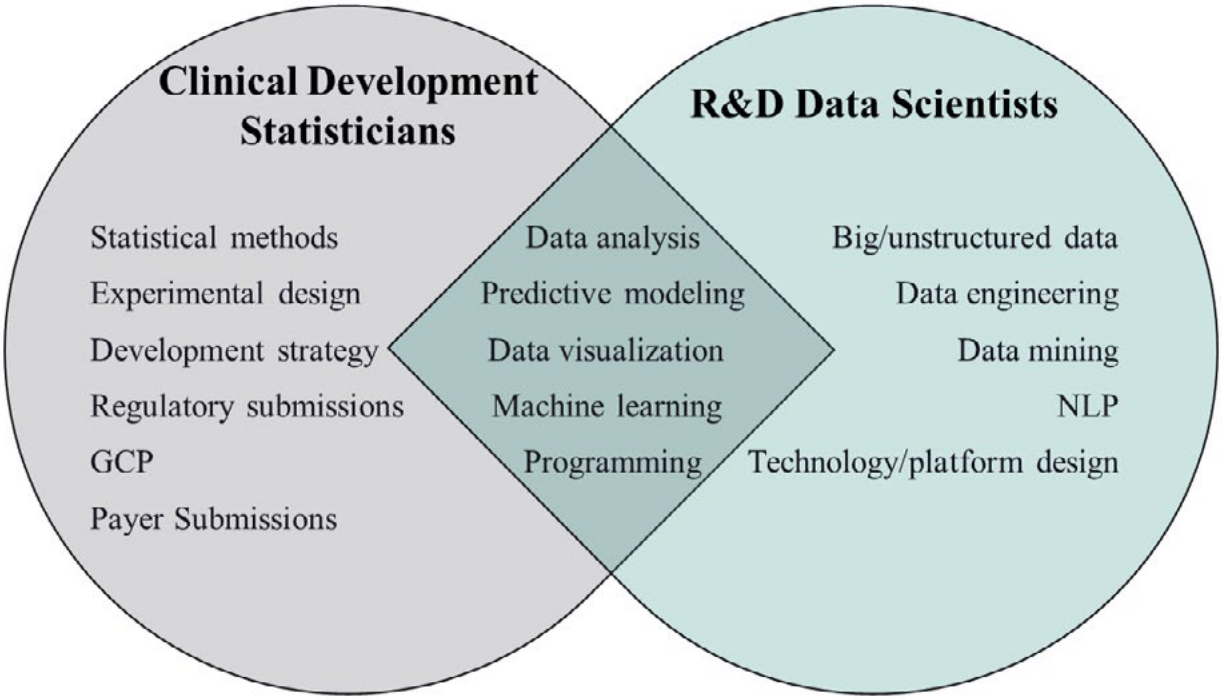
To effectively discuss the collaboration between the roles, we must first clarify the roles and the skillsets statisticians and data scientists bring. Statisticians have a diverse set of roles throughout the drug development life cycle spanning from non-clinical to clinical stages, and to post-marketing evidence generation and beyond. Similarly, data science encompasses a variety of skills such as Natural Language Processing (NLP), Artificial Intelligence (AI) and data engineering, with data scientists being able to specialize or serve as general practitioners.

While one could define a data scientist as someone who derives insights from data, this language obfuscates the various skills and talents that different quantitative specialists bring. For the purpose of this discussion, statisticians are characterized first by their formal training in statistical theory and methods, with statistical thinking as the basis. Crucially, this training includes developing a deep understanding of sources of bias, uncertainty, and techniques to address both. This methodological foundation allows statisticians to conduct quality non-clinical and clinical R&D work and meet the standard of Good Laboratory Practice (GLP) and Good Clinical Practice (GCP) and other relevant regulations in drug development. The role includes the

framing of research questions, experimental design and simulation, data analysis methods, and the interpretation of results – all of which require a high level of discipline and expertise to navigate properly. Statisticians, upon this foundation, seek to help their companies make the best strategic decisions possible at any point across development lifecycle, submissions, post-marketing evidence generation and beyond.

Data scientists in general are recognized for their training in aspects of data engineering, data visualization, data assembly, predictive modeling, and associated technologies that make complex data integrated and accessible to end users. While statisticians often manage well-structured data, data scientists are often trained to handle unstructured data and/or vast quantities of data. Areas of strength include technological acumen and strategic ability to organize and interrogate disparate data sources. Data scientists are often employed on predictive modeling and identification / classification problems.

At present, there is no clear line in the entry-level skills between statisticians and data scientists. Rather their skillsets can be more effectively described as somewhat overlapping and somewhat complementary. Statistical methodology and experimental design fall clearly into the statistician skillset, while skills such as data mining and data engineering fall more clearly under the purview of data scientists. Selected analysis techniques such as machine learning, data visualizations, etc. are leveraged by both groups, forming the overlap of this Venn diagram. These broad generalizations will, of course, fail to capture the unique skills of individual quantitative experts who may possess a broader range of skills, as well as the nuances that may differ across organizations. Instead, they describe the broad foundational aspects of the roles. Although cross-skilling is possible in the workplace, leaders have found that the educational backgrounds of the two groups position them to lead with different problem-solving approaches. These disparate approaches sometimes



Soft Skills that Inform Successful Collaboration

- *Process-Focused Approach*
- *Emotional Intelligence*
- *Entrepreneurial Mindset*
- *Non-Technical Communication*
- *Collaborative Mindset*
- *Continuous Exploration*
- *Problem Framing*
- *Adaptability*

make it challenging for data scientists to fully appreciate the standards and rigor required by the regulatory, payer, and commercial environments governing the pharmaceutical industry. In a similar vein, statisticians can struggle to assemble/structure large data sources in a manner that allows for meaningful analysis. Senior leaders have found that close collaboration is essential to mitigating both challenges. In the case of GCP, a careful focus on design, data stewardship, and process execution is key for bringing data scientists into clinical research. For exploratory analyses, collaboration that begins during solution design allows data scientists to build infrastructures that benefit both groups. While there are natural crossovers in the quantitative skills of data scientists and statisticians, it's equally salient to consider the soft skills that the two roles foster and leverage most frequently. Understanding the core areas of expertise and strengths of each discipline is essential for appropriate assignment of roles and responsibilities.

As this collaboration becomes more established, and academia shifts to support the industry, the Venn diagram between data science and statisticians begins to shift towards a landscape of technical skills along which quantitative experts fall and collaboration becomes more fluent. The similarities and differences between these roles present both challenges and opportunities. Challenges exist when stakeholders do not understand the differences between the roles and mis-match expertise to business problems. Opportunities exist for synergistic work between both disciplines. For example, by creating an NLP algorithm to detect signals in text-based data that might inform a trial, or by providing analytics tools to access real-time data. In fact, this may be where the greatest opportunity is. With the increasing complexity of clinical research, and diversity of data sources available, it is reasonable to conclude that impactful business problems will increasingly require the partnership and leveraging of unique talents of both roles.

3. Collaboration Requires Thoughtful Organizational Structures in R&D

When they collaborate effectively, data scientists and statisticians can amplify the impact of their work to increase the likelihood of clinical trial success and speed, improving patient outcomes, and meeting budgetary constraints. Although there are concrete examples of successful collaboration between statisticians and data scientists in biopharma, senior leaders highlight that the relationship between the two roles has not fully matured. Key headwinds include hype that may fail to deliver business value and the challenge of connecting quantitative experts to problems or inefficiencies that they can improve. These challenges may stem in part from the rapid build-out of new quantitative teams without clear goals or establishing a track record of success. In addition, functional leadership that leverages data scientists may not understand important issues related to quantitative work in R&D. This includes proper framing of research questions, design and analysis planning, executional validation, interpretive practices, and the distinction between hypothesis generation and confirmation. This is further challenged by the fact that the roles of statisticians and data scientists are often not well understood by management. Dedicated data scientists are relatively new within biopharma. Thus, their presence is likely to continue increasing as organizational trust continues to grow and more examples of data scientists' value are established.

To optimize the impact of both statisticians and data scientists, one must first consider the common organizational structures that inform their collaboration. While no biopharmaceutical companies position data scientists and statisticians in exactly the same way, a majority follow one of three organizational structures. The first is an integrated model where both types of quantitative experts work together as part of the same broader team inside the biometrics organization. The second is to have a separate data science group that sits outside of the biometrics group. The third is for data scientists to be deployed within a variety of functions throughout R&D alongside a dedicated biometrics function. The first model benefits from a unified quantitative landscape with sole leadership responsibility for deployment, collaboration and outcomes. At the same time, more effort may be required to connect with functional stakeholders. As the model moves to options two and three, the quantitative community becomes more dispersed. This may spread work, and in fact ownership, closer to functional stakeholders; but potentially at the expense of competitive dynamics, duplication of work

or organizational confusion.

In organizations which follow the integrated model, the burden of connecting problems to the right experts within the quantitative organization is borne by individual project leaders. In contrast, organizations with separate quantitative groups often require leadership to direct questions to the right stakeholders. There, the connectivity should occur at the leadership level rather than within a team. If not carefully managed, this structure may result in suboptimal deployment of work. In addition, as work often involves clinical trial data, important clinical trial/endpoint knowledge may be lost as work is transitioned external to biometrics. The fully decentralized model creates the most challenges. While functions initially enjoy the ownership of data science teams, they often face challenges of internal competition over projects (or staff) as well as scalability problems.

The collaboration between statisticians and data scientists within R&D requires a level of trust which can be supported by shared management objectives and a careful understanding of the roles that is not necessary in most businesses outside of R&D. Within R&D, statisticians can be hesitant to share data outside their organizations because of the risks associated with provision to analysts who are not knowledgeable in clinical trial design and the impact of design on the generated data. Sharing clinical data has the potential to lead to issues of secondary use, and the complexity of clinical trial data can mean that external stakeholders require time-consuming guidance to accurately interpret the data. Naturally, there is a risk associated with sending clinical data to other functions, thus breaking traceability and leaving the protection of a GCP procedural framework. In addition, when data from ongoing trials is shared without appropriate planning and firewall procedures, there is potential for operational bias that reduces the integrity of the trial. These risks and the associated costs can make statisticians hesitant to share clinical trial data with their data science colleagues.

One of the key points that senior statistics leaders highlight as successes within biopharma organizations is the ability to have collaborations occur without formal alignment or being connected by senior leadership. This is true even within a centralized model. This organic collaboration requires quantitative experts to recognize the skillsets of their colleagues and where they can add value. Having an organizational model that rewards partnerships between the functions makes it easier to bring the right experts onto a given project. When trust is established and quantitative experts have a good working relationship, both data scientists and statisticians can be deployed on the same problems to work together leveraging the significant strengths of both disciplines

Regardless of organizational structure, biopharma leaders have identified a few attributes of successful collaborations. Key among them are team-oriented thinking and a high level of emotional intelligence. This can be observed as humility and a willingness to bring in colleagues and share successes. From a management perspective, this model can be incentivized when success for both data scientists and statisticians is measured collaboratively. Achieving this goal requires tactful leadership to mitigate the fear that can arise among statisticians as data scientists encroach upon their role as the primary quantitative experts in R&D. This tension can be exacerbated by the broader organization celebrating the successes of their new data science capabilities rather than more accurately crediting them the achievements to all the quantitative experts whose work was represented. This can occur quite naturally as leaders seek to validate their expanded capabilities but has the potential to disincentivize organic collaborations.

Having leadership that rewards partnerships between the functions makes it easier to bring the right stakeholders onto a given project. Successful data science teams are characterized by their ability to leverage network effects and provide scalable solutions that become the new elevated standard, without linear scaling in headcount. This frees up data scientists to innovate and bring additional value to their organizations, while becoming a go-to resource for their respective statistics organizations.

4. Academia Can Continue to Support Collaboration Within Biopharma

Given the changing landscape of quantitative expertise within biopharma, it is worth exploring the academic programs that feed into the industry and how the two can be brought into better alignment. At present, academia is parallel to biopharma in recognizing the overlap in statistics and data science, as demonstrated by a growing number of combined departments at universi-

ties such as Cornell and Carnegie Mellon1,2. Academia can continue to support industry in creating successful collaborations by placing focus on encouraging partnerships between students studying data science and statistics, so they have a better understanding of each other's unique skillsets and fluency in collaborating even before entering industry.

At present, a key challenge of the academic programs that train quantitative experts who go on to work in biopharma R&D is the lack of training they provide on clinical trials and the unique challenges of working with biopharmaceutical data. This could be effectively addressed either through coursework that focuses specifically on clinical development and the unique challenges therein, or by forming closer relationships with biopharmaceutical companies to help inform curricula, or in some cases having a consulting lab in the universities as part of their program that includes statisticians and data scientists. By building a model with more clarity regarding the roles in industry and having this training available to both statisticians and data scientists in academia, industry could effectively mitigate the risk of detrimental results driven from misunderstanding of the data and entering the workforce without understanding the ways in which collaboration is mutually beneficial.

5. Conclusion

Both data scientists and statisticians should foster respect for each discipline and become familiar with the areas of unique contribution from the other. This awareness allows the right people to work on the right projects and helps shape collaboration opportunities. Secondly, having leadership understand the distinctions, and actively support a collaborative environment is key to reducing friction and driving a cooperative rather than competitive culture. Lastly, statisticians and data scientists should seek to build a mutual understanding which allows them to collaborate and innovate together. This facilitates the early and frequent interaction necessary to build trust, gain business acumen, and ultimately drive significant contributions. The quality of quantitative impact is likely to be directly related to the strength of the collaboration. The opportunity is there though challenges exist, and if we work together the future is truly bright for what we can accomplish.

References:

1. Cornell University Department of Statistics and Data Science. (n.d.). Retrieved August 23, 2022, from *https://stat.cornell.edu/*

2. University, C. M. (n.d.). *Statistics & Data Science*. Statistics & Data Science - Dietrich College of Humanities and Social Sciences - Carnegie Mellon University. Retrieved August 23, 2022, from *https://www.cmu.edu/dietrich/statistics-datascience/index.html*

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

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

On October 14, 2021, the American Statistical Association (ASA) Biopharmaceutical Section (BIOP) and LUNGevity Foundation hosted a virtual open forum to discuss statistical considerations in the evaluation and interpretation of data from cancer clinical trials conducted in a single foreign country as 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 (SWG), the FDA Oncology Center of Excellence (OCE), and LUNGevity Foundation.

International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH) Guideline E17 describes general principles for planning and designing multi-regional clinical trials (MRCTs) with the assumption that MRCTs provide more robust evidence than single region trials for the purpose of extrapolating study results for global registration. Extrapolation, however, depends on the completeness of the data package—with respect to regulatory requirements—that is submitted to the new region, as outlined in ICH Guideline E5: Ethnic Factors in the Acceptability of Foreign Clinical Data.

For many years, countries with relatively homogeneous racial and ethnic populations had to decide whether results from clinical trials conducted in the U.S. and other countries were applicable to them. Now, however, we are seeing the opposite: data from trials conducted in one country/region with a homogeneous population are being submitted to the U.S. and other countries who must then decide whether the data are generalizable to their more racially/ethnically heterogeneous populations. This open forum discussion among multi-disciplinary experts explored when and how foreign data from a single country are applicable to other countries/regions. Specific points of consideration were: how to determine whether extrapolated single-country data are valid for heterogeneous populations; what data need to be collected; when bridging studies are required; and what clinical trial designs will help assess the acceptability of foreign data.

The speakers/panelists* for the discussion included members of the BIOP Statistical Methods in Oncology SWG representing pharmaceutical companies, representatives from international regulatory agencies (FDA, Health Canada (HC), Medicines and Healthcare products Regulatory Agency (MHRA), Pharmaceuticals and Medical Devices Agency (PMDA), and Swissmedic (SMC)), 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 Medicinal Agency (EMA), Singapore Health Sciences Authority (HAS), Brazilian Health Regulatory Agency (ANVISA)). The discussion was moderated by the BIOP Statistical Methods in Oncology SWG 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.

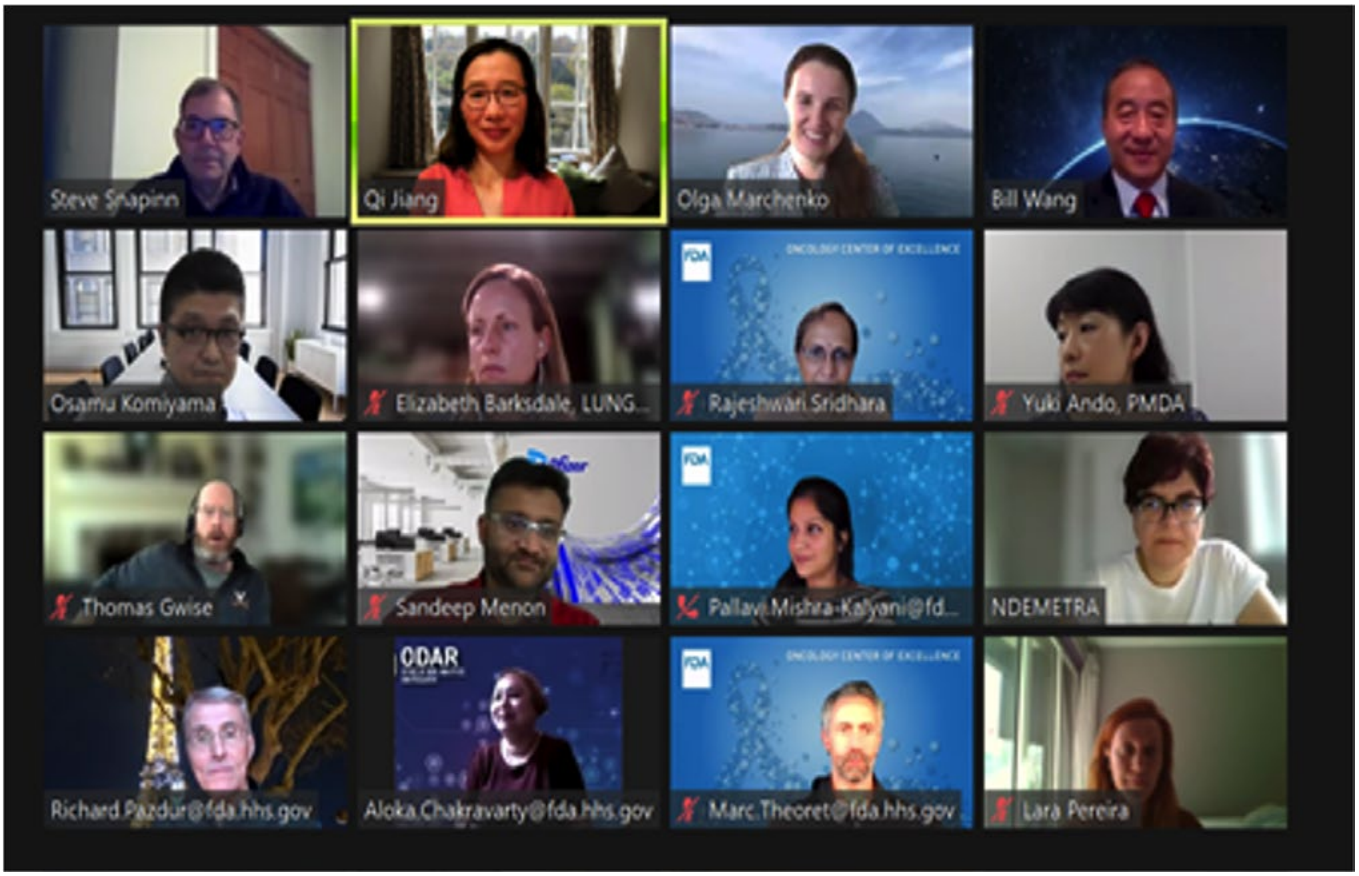
The forum opened with an introductory presentation and remarks from OCE leadership, who reviewed the tenets of ICH guidelines E17 and E5, including the intrinsic and extrinsic factors that need to be included in regulatory submissions for the purpose of extrapolation. It was noted that MRCTs are preferred to single-country/region clinical trials in order to see clinical differences related to the investigational medical product across geographic locations. This was followed by two presentations from speakers representing industry and regulatory viewpoints.

The industry speaker presented on the evolution of ICH E5 and E17 in oncology trials. The collective mindset has progressed from local to global in the nearly 20 years between the release of the guidances, with sequential bridging of local data giving way to evaluating regional consistency with global data from MRCTs. There are seven principles of good MRCT designs outlined in ICH E17 which should be applied from the earliest stages of drug development in order to fully explore and understand regional differences. Two case studies were provided which exemplify such a global development strategy: gefitinib for advanced non-small cell lung cancer, and zanubrutinib for mantle cell lymphoma. In closing, the speaker noted that global MRCTs should be the preferred option, and that ICH E17 promotes global drug development when used together with E5 and other ICH guidance, such as E9 R1. Additionally, the design and conduct of regional/country-level trials should align with the global development program to understand intrinsic and extrinsic factors and ensure trial quality.

The second speaker, from the Pharmaceuticals and Medical Devices Agency (PMDA) in Japan, talked about their specific experience using foreign data. Prior to 1998 there was no guideline for using foreign data in drug approvals, and PMDA relied solely on Japanese-only trial data. Bridging studies were allowed following the implementation of ICH E5 in 1998, and then foreign data from MRCTs was used upon release of Japanese guidance documents on global clinical trials (2007-2014) and finalization of ICH E17 in 2017. Under the bridging strategy, pharmacokinetics (PK) studies were conducted in Japan and compared with foreign Phase I data to assess intrinsic similarities; a Phase II dose-finding bridging study was conducted; and then, if data were similar, Phase III efficacy results were extrapolated to the Japanese population. As a result, drug development in Japan was delayed compared to other countries. When Japan started participating in MRCTs to support regulatory approvals, the number of bridging studies decreased, leading to faster approvals and access. Lessons learned from the Japanese experience include: prior consideration of possible ethnic factors is critical, and these factors may be important to explain observed differences in trial (sub)populations.

Key points raised during the subsequent multi-stakeholder panel discussion include:

1. The main takeaway from regulators was that MRCTs should be the goal, but this will require preplanning on the sponsors' part and having a solid understanding of the different populations involved. Sponsors need to demonstrate that their trials meet all regulatory requirements to support approval in countries to which they will be submitted, including having data that meet good clinical practice standards and adequately inform PK/PD, safety, and efficacy.
2. Differences in comparators used and standard of care (SOC) across regions can make extrapolation of trial results very difficult, as the effect of the treatment relative to different SOC regimens can differ. Relatedly, practice of medicine (POM) reflects cultural and regional differences and often differs significantly by region. It was proposed that sponsors could require uniform SOC/post-progression therapy in their trials, but differences in regional POM could affect the interpretability of results. Other suggestions were to use intent-to-treat analyses, which can assess the direct treatment effect by treatment assignment, or comparability of treatment effect across regions. Another proposal was to update the E17 guidance to address the complexity of evolving SOC.
3. Identifying appropriate intrinsic and extrinsic factors and determining their likelihood of affecting treatment safety and efficacy , respectively, can be difficult from both the biological and statistical perspectives. Bridging studies may help elucidate whether certain intrinsic factors are problematic. There were differing opinions on whether extrinsic factors are likely to impact



treatment effect when overall survival is the endpoint. Tumor intrinsic and extrinsic factors also need to be considered, and the effect different drug classes will have on them.

4. Bridging studies were raised as an important issue by multiple stakeholders. It is important to identify appropriate endpoints, study size, and when formal or informal comparisons are needed. It is important to clarify when a bridging study is needed vs. a safety study. Additionally, if sponsors bring in regulators early in the development process, it may be possible to include a sufficient number of patients in the MRCT to meet their requirements so that a bridging study would not be needed.
5. The issue of generalizability isn't specific to single-country clinical trial data and can be encountered when estimating treatment effect from any sub-population that isn't representative of the overall population. Other examples include multi-center phase III trials failing to replicate single-center phase II results. Shrinkage estimation, where the estimate for any subgroup is weighted average of the subgroup-specific estimate and the overall study result, may be useful in certain cases.

This forum provided an opportunity to have open scientific discussion among a diverse multidisciplinary stakeholder group, including clinicians, 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 multi-disciplinary open forum discussions on a variety of important topics.

Acknowledgement:

Authors thank Rick Peterson (ASA) and Joan Todd (FDA) for supporting the forum, and Dr. Nicole Li (BeiGene) for taking the meeting minutes.

*Speakers/ Panelists:

Dr. Yuki Ando (PMDA), Dr. Elizabeth Barksdale (LUNGevity Foundation), Dr. Angelo de Claro (FDA), Dr. Thomas Gwise (FDA), Lorenzo Hess (SMC), Dr. Qi Jiang (Seagen), Dr. Osamu Komiyama (Pfizer), Dr. Brent Logan (Medical College of Wisconsin), Dr. Olga Marchenko (Bayer), Dr. Sandeep Menon (Pfizer), Dr. Pallavi Mishra-Kalyani (FDA), Dr. Richard Pazdur (FDA OCE), Dr. Khadija Rantell (MHRA, UK), Mr. Andrew Raven (HC), Dr. Yuan-Li Shen (FDA), Dr. Steven Snapinn (Statistical Consultant), Dr. Rajeshwari Sridhara (FDA OCE), Dr. Marc Theoret (FDA OCE), Dr. William Wang (Merck & Co., Inc.). ■

SUMMARY OF ASA BIOP SECTION'S VIRTUAL DISCUSSION WITH REGULATORS ON CANCER CLINICAL TRIAL DESIGN CONSIDERATIONS IN EVALUATING TREATMENT EFFECT IN MARKER NEGATIVE POPULATION

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

On December 9th of 2021, the American Statistical Association (ASA) Biopharmaceutical Section (BIOP) and LUNGevity Foundation hosted a virtual open forum to discuss cancer clinical trial design considerations in evaluating treatment effect in a marker negative population as part of a series conducted under the guidance of the U.S. FDA Oncology Center of Excellence's (OCE) 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 OCE, and LUNGevity Foundation.

With advancements in precision medicine and development of molecularly targeted therapies, it is not uncommon to conduct cancer clinical trials in enriched populations, such as a marker positive population only. However, when there is no available therapy for a particular disease, the prevalence of the marker negative population is low, a reproducible assay has not been developed, or the classification threshold for marker positive/negative is not clear, especially for continuous markers, randomized clinical trials are often conducted in the overall population which includes both marker positive and marker negative patients. Given the potential for differential treatment effects among these subgroups of the overall intent-to-treat (ITT) population, trialists are faced with the conundrum of how to design the hypothesis testing hierarchy. One option is to test the hypothesis in the marker positive subgroup first; if the treatment effect in the marker positive subgroup is statistically significant, then test the hypothesis in the ITT population. In this scenario, the concern of concluding a significant treatment effect in the ITT population is that the results may be driven by a marker positive subgroup with a high prevalence rate, and the contribution of the marker negative treatment effect may be minimal and uncertain. Understanding the contribution of the marker negative treatment effect to the overall ITT results is important. This open forum discussion among multi-disciplinary experts focused on cancer clinical trial design considerations when a conclusion is reached on the overall ITT population after evaluation of treatment effect in the marker positive subgroup.

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, European Medicinal Agency (EMA), Health Canada (HC), Therapeutic Goods Administration (TGA)), academicians, and expert statistical consultants. In addition, over 100 participants attended the virtual meeting, including representatives from other international regulatory agencies (Medicines and Healthcare products Regulatory Agency (MHRA), Swissmedic (SMC), Singapore Health Sciences Authority (HAS), Brazilian Health Regulatory Agency (ANVISA), Pharmaceutical Division Israel

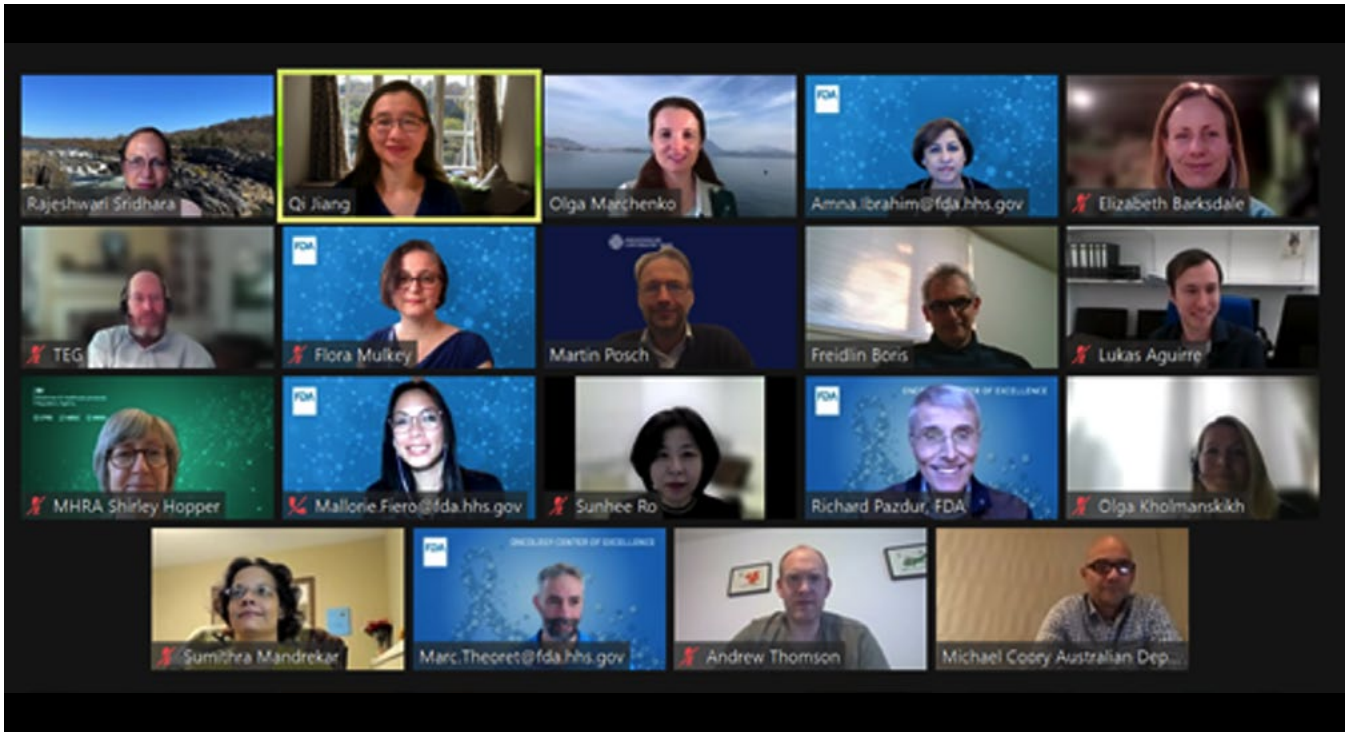
Ministry of Health), Japan (PMDA). The discussions were moderated by the BIOP Statistical Methods in Oncology Scientific Working Group co-chairs, Dr. Qi Jiang from Seagen and Dr. Olga Marchenko from Bayer; Dr. Elizabeth Barksdale from LUNGevity Foundation; and Dr. Rajeshwari Sridhara, consultant from OCE, FDA.

An introductory presentation and remarks by the OCE leadership highlighted the need for a multi-disciplinary approach to understand the contribution of the marker negative population to the overall ITT positive results. Based on emerging scientific knowledge of the mechanism of action of the investigational product, and understanding predictive or prognostic biomarkers associated with it, ideally enrichment on biomarker specific patient population or ITT population can be designed. The focus of this open forum discussion was whether or not a large treatment effect in the marker positive subgroup contributing to a significant effect in the ITT population may lead to an implicit conclusion of treatment effect in the marker negative subgroup. Points to consider in the discussions were (1) How to interpret treatment effect in ITT population? (2) What if ITT results are driven by the marker positive subgroup? (3) What if a marker assay is not standardized? In other words, how to interpret whether the marker negative population contributes to the overall positive results in the ITT population. This presentation was followed by presentations from statisticians representing the FDA and industry.

The FDA speaker pointed out general statistical issues with subgroup analyses, such as lack of adequate power and sample size to detect a clinically meaningful treatment effect and the lack of pre-specified alpha for multiplicity adjustment [1]. Two case studies, along with their review considerations, were presented: Keynote 590, an approval for the treatment in ITT population in esophageal cancer, and monarchE, an approval for the treatment in a subgroup of adjuvant breast cancer patients. In Keynote 590, results in both the pre-specified biomarker positive and ITT populations were compelling in terms of treatment benefit and were statistically significant. In addition, the biomarker negative subgroup had sufficient number of events, the treatment effect did not indicate a detrimental effect. On the other hand, in the monarchE example, the approval was restricted to a subgroup after taking into consideration limitations of the observed data in the remaining subgroup, e.g., data maturity, and small sample size and the limited number of outcome events.

The speaker from industry recommended a 2-stage, biomarker-based, adaptive design starting with ITT population, when considering situations in which there is uncertain marker prevalence and predictiveness. The promising marker-based subpopulation will be selected at an interim analysis, and the second stage of the study will focus on the selected subpopulation. The data from both stages will be used for the treatment effect estimates in the selected population with the use of a combination test to control type I error. Case studies that could have benefitted from the proposed approach were presented. Operational and regulatory considerations were also summarized.

The key points raised in the panel discussion following these presentations were the following. Currently, the majority of the phase III RCT designs that evaluate biomarker-defined subpopulations as well as the overall (ITT) population to infer on treatment recommendation for biomarker-negative patients based on the overall (ITT) population treatment effect (without formal evaluation of the biomarker-negative population). This design could be problematic as it can result in an unnecessary and sub-optimal treatment in the biomarker-negative patients; this is because treatment effect in ITT population could be driven by biomarker-positive patients with unfavorable risk-benefit profile in the biomarker-negative patients. When there is a hierarchical testing in the marker positive subgroup followed by ITT population, substantial uncertainties can arise on whether the benefit outweighs risk in the marker negative subgroup particularly if the sample size is small in this group. Ideally, this should be addressed at the design stage based on the scientific knowledge that is available regarding the disease, the mechanism of action, the expected magnitude of effect and safety profile of the investigational drug product. Key concepts for subgroup analyses can be applied at the design stage [2]. One possibility is to have a threshold for the treatment effect set up for the marker-negative group at the planning stage that would help in objective decision



making. Use of a Bayesian approach may also support in characterizing the uncertainty. Another possibility suggested was a use of modelling approaches.

This forum has provided opportunities to have open scientific discussion among a diverse multidisciplinary stakeholder group, including clinicians, 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 Mr. Syed Shah (FDA) for supporting the forum, and Dr. Yuan-Li Shen (FDA) for taking the meeting minutes.

References:

[1] Amatya et al. (2021) Subgroup analyses in oncology trials: regulatory considerations and case examples [DOI: 10.1158/1078-0432.CCR-20-4912]

[2] Guideline on the investigation of subgroups in confirmatory clinical trials (EMA/CHMP/539146/2013)

* Speakers/ Panelists:

Dr. Elizabeth Barksdale (LUNGevity Foundation), Dr. Michael Coory (TGA), Dr. Lukas Aguirre Davila (Paul Ehrlich Institute, EMA), Dr. Mallorie Fierro (FDA), Dr. Boris Freidlin (NCI), Dr. Thomas Gwise (FDA), Dr. Amna Ibrahim (FDA), Dr. Mette Linnert Jensen (Danish Medicines Agency, EMA), Dr. Qi Jiang (Seagen), Dr. Olga Kholmanskikh (Federal Agency for Medicines and Health Products, EMA), Dr. Sumithra Mandrekar (Mayo Clinic), Dr. Olga Marchenko (Bayer), Ms. Flora Mulkey (FDA), Dr. Richard Pazdur (FDA), Dr. Reena Philip (FDA), Prof. Martin Posch (Center for Medical Statistics, Informatics, and Intelligent Systems at the Medical University of Vienna), Mr. Andrew Raven (Health Canada), Dr. Sunhee Ro (Sierra Oncology), Dr. Richard Simon (Simon Consulting), Dr. Rajeshwari Sridhara (OCE), Dr. Marc Theoret (FDA), Mr. Andrew Thomson (EMA).

SUMMARY OF ASA BIOP SECTION'S VIRTUAL DISCUSSION WITH REGULATORS ON IMPACT ON TYPE I ERROR WITH UNPLANNED ANALYSES IN CANCER CLINICAL TRIALS

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

On April 7th of 2022, American Statistical Association (ASA) Biopharmaceutical Section (BIOP) and LUNGevity Foundation hosted an open forum discussion among industry, academia, and regulators on Impact on Type I Error with Unplanned Analyses in Cancer Clinical Trials as part of a series of discussions conducted for the United States Food and Drug Administration (US FDA) Oncology Center of Excellence (OCE) initiative, Project SignifiCanT (Statistics in Cancer Trials). The goal of Project SignifiCanT is to advance cancer drug development through collaboration and engagement among stakeholders in the design and analysis of cancer clinical trials. Organized jointly by the ASA BIOP Statistical Methods in Oncology Scientific Working Group (SWG), LUNGevity Foundation, and the FDA Oncology Center of Excellence, the purpose of this forum was to explore the impact of unplanned analyses on Type I error in cancer trials which are intended to support regulatory approval.

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 (US FDA, Health Canada, European Medicine Agency (EMA), Medicines and Healthcare products Regulatory Agency (MHRA) from the United Kingdom, and Department of Health from Australia), academicians and expert statistical consultants. In addition, over 100 members attended the virtual meeting including representatives from other international regulatory agencies (e.g., from Brazil, Israel, Singapore, Switzerland). The discussions were moderated by the BIOP Statistical Methods in Oncology SWG 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 the FDA OCE.

Randomized cancer clinical trials intended to support regulatory approval are generally designed with a set of primary and secondary endpoints and corresponding statistical analyses accounting for multiple analyses and control of Type I error. However, unplanned analyses occur in some trials for different reasons including: (1) observed safety concerns within the study, particularly when overall survival is being evaluated, or efficacy analysis for benefit-risk consideration; (2) prompted by new data from an external source (e.g., to change the study population); (3) to inform regulatory decision making; and (4) sponsor-initiated administrative analyses. In all these scenarios it is important to understand whether Type I error maybe be inflated and how to control it.

The meeting started with an introductory presentation and remarks from the FDA OCE leadership. Questions specific to panelists from academia, industry, and regulatory agencies were reviewed, with general themes being: (1) reasons for unplanned analyses; (2) when and how Type I error adjustment should be implemented; (3) how to assess integrity of ongoing trial(s) when unplanned analyses occur; and (4) how to interpret and

communicate results from unplanned analyses.

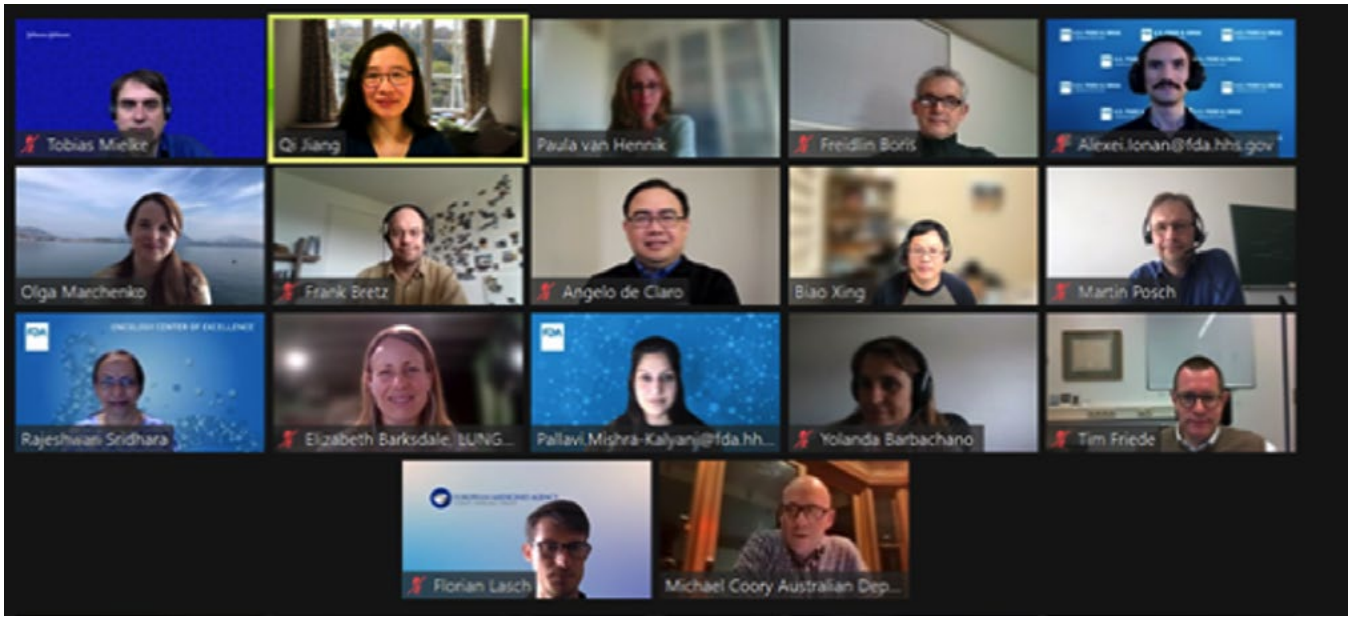
The first presenter was a clinical expert from EMA who provided a regulatory view on the topic. The presenter quoted ICH E9 Statistical Principles for Clinical Trials stating that unplanned interim analyses should be avoided and "if unplanned interim analysis is conducted, the clinical study report should explain why it was necessary, the degree to which blindness had to be broken, provide an assessment of the potential magnitude of bias introduced, and the impact on the interpretation of the results." Transparency around the rationale for and impact of unplanned analyses is essential. Sponsors should explain, and regulators evaluate, impacts on Type I error, study design features (e.g., blinding, treatment adherence), and trial integrity.

The second presenter, representing industry, also quoted ICH E9, specifically Section 5.1 Prespecification of the Analysis. The presenter acknowledged that unplanned interim analyses may create difficulties in controlling Type I error rate, maintaining trial integrity, and interpreting trial results. To minimize the impact, sponsors should maintain the appropriate firewalls, document reasons for unplanned analyses, and make appropriate adjustments in the statistical analysis plan. Types and examples of unplanned analyses were reviewed and discussed. The presentation concluded with questions about and challenges to current practices.

The third presenter, from academia, discussed potential impacts based on the information and types of decision rules used to trigger an unplanned analysis. Depending on data used to trigger the analysis and the decision made, the impact on Type I error can range from none to a substantial inflation. Practical issues, such as confirming what information was available when the analysis was triggered and what decision rule was used to trigger the unplanned analysis, might also be not trivial. The presenter concluded that strict Type I error control is not possible without strong assumptions if an analysis of hypothesis testing may be triggered by trial data correlated with the primary endpoint. He also noted that Bayesian analyses are valid even after unplanned interim looks, regardless of how they have been triggered. Additionally, unplanned analyses can have an indirect impact on the remainder of the trial, and therefore, sensitivity analyses comparing outcomes and baseline characteristics before and after the interim analysis are recommended (Dimairo, M. et al. 2020).

The presentations were followed by a panel discussion covering opinions from diverse stakeholders. The discussion focused on the questions raised at the beginning of the meeting and on points brought by the presenters. Presenters and panelists agreed that unplanned analyses should be avoided, if possible, but when they are conducted, better transparency on why and how is needed. It was noted that many unplanned analyses can be planned and formalized through protocol amendments and/or SAP amendments (the rationale and timing for such amendments should be clearly specified). Regulators mentioned that unplanned analyses are done more often for single arm trials than for randomized trials, and that most unplanned analyses are conducted after a trial completion. The main reason regulators ask for unplanned analyses is to evaluate overall survival information for regulatory decision making when a submission is based on an ongoing single arm trial or subgroup analysis for labeling considerations. Adjustments of Type I error have been evaluated on a case-by-case basis, for example, adjustments are not required for unplanned analyses done for safety or futility. If the regulatory agency requested an unplanned analysis after a study completion, the regulators usually do not request Type I error adjustment. It was also mentioned that additional analyses are often requested by payers in one or multiple countries. Sponsors might have different analysis plans for different regions, at times even using different primary endpoints. Integration of results in such cases might be challenging. In the situations when Type I error needs to be adjusted, it also might be challenging to understand what portion of alpha should be spent at unplanned analyses.

This forum, similar to previous ones, provided an opportunity to have open scientific discussions among a diverse stakeholder group. We plan to continue with similar multi-disciplinary open forum discussions in the future on a variety of important statistical aspects in cancer drug development.



* Speakers/ Panelists:

Dr. Yolanda Barbachano (MHRA, UK), Dr. Elizabeth Barksdale (LUNGevity Foundation), Dr. Frank Bretz (Novartis), Dr. Angelo de Claro (FDA), Dr. Michael Coory (Department of Health, AU), Dr. Boris Freidlin (NCI), Prof. Tim Friede (University Medical Center Göttingen), Dr. Paula van Hennik (Medicines Evaluation Board, The Netherlands, Dutch alternate CHMP member, European Committee for Human Medical Products), Dr. Alexei Ionan (FDA), Dr. Qi Jiang (Seagen), Dr. Florian Lasch (EMA), Dr. Olga Marchenko (Bayer), Dr. Tobias Mielke (J&J), Dr. Pallavi Mishra-Kalyani (FDA), Dr. Richard Pazdur (FDA), Dr. Inna Perevozskaya (GSK), Prof. Martin Posch (Center for Medical Statistics, Informatics, and Intelligent Systems at the Medical University of Vienna, Austria), Dr. Khadija Rantell (MHRA, UK), Andrew Raven (Health Canada), Dr. Lisa Rodriguez (FDA), Dr. Rajeshwari Sridhara (Contractor, Oncology Center of Excellence, FDA), Dr. Marc Theoret (FDA), Dr. Biao Xing (Seagen).

Acknowledgement:

Authors thank Ms. Joan Todd (FDA) and Mr. Syed Shah (FDA) for supporting the forum and Dr. Weidong Zhang (Sana Biotechnology) for taking the meeting minutes.

References:

ICH E9 Statistical Principles for Clinical Trials: *https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials*

Dimairo, M. et al. "The Adaptive designs CONSORT Extension (ACE) statement: a checklist with explanation and elaboration guideline for reporting randomised trials that use an adaptive design", BMJ 2020; 369:m115: *https://www.bmj.com/content/369/bmj.m115*

SUMMARY OF ASA BIOP SECTION'S VIRTUAL DISCUSSION WITH REGULATORS ON CONSIDERATION IN THE EVALUATION AND INTERPRETATION OF INTERIM OVERALL SURVIVAL RESULTS FROM RANDOMIZED CANCER CLINICAL TRIALS IN CHRONIC DISEASES

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

The American Statistical Association (ASA) Biopharmaceutical Section (BIOP) and LUNGevity Foundation hosted a virtual open forum on May 12, 2022, to discuss statistical considerations in the evaluation and interpretation of interim overall survival (OS) results in patients with chronic diseases from randomized cancer clinical trials as 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 cancer clinical trials in chronic diseases, such as chronic lymphocytic leukemia (CLL), which are intended to support regulatory approval often have progression-free survival (PFS) as the primary outcome of interest. However, OS is an important endpoint in all randomized cancer clinical trials as it is considered a measure of both efficacy and safety of treatments. Given the long course of such chronic diseases, the number of events for the OS analysis can be limited at the time of the final analysis of PFS. Additionally, in some of these cases, OS analysis may not be hypothesis driven. There are challenges in interpreting interim OS analysis results in such cancer trials, particularly if the early analysis suggests the possibility of a detrimental effect of the investigational drug, and in assessing the benefit–to-risk evaluation of the drug. This open forum discussion among multi-disciplinary experts delved in ways to quantify uncertainty in the evaluation and interpretation of early OS analysis in cancer trials that evaluate treatments for patients with chronic diseases and are intended to support regulatory approval.

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, Health Canada (HC), Therapeutic Goods Administration (TGA), Medicines and Healthcare products Regulatory Agency (MHRA), 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 Medicinal Agency (EMA), Swissmedic (SMC), 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.

An introductory presentation and remarks by the OCE leadership highlighted the need for a multi-disciplinary approach to understand the risks and benefits of treatments for chronic cancers where early analyses suggest a potential detriment in OS. For example, in randomized clinical trials evaluating six PI3K inhibitors for the treatment of CLL, each trial showed a statistically significant improvement in PFS along with a potential detriment in OS based on early OS analysis. This presentation was followed by three others, from statisticians representing industry, academia, and the FDA.

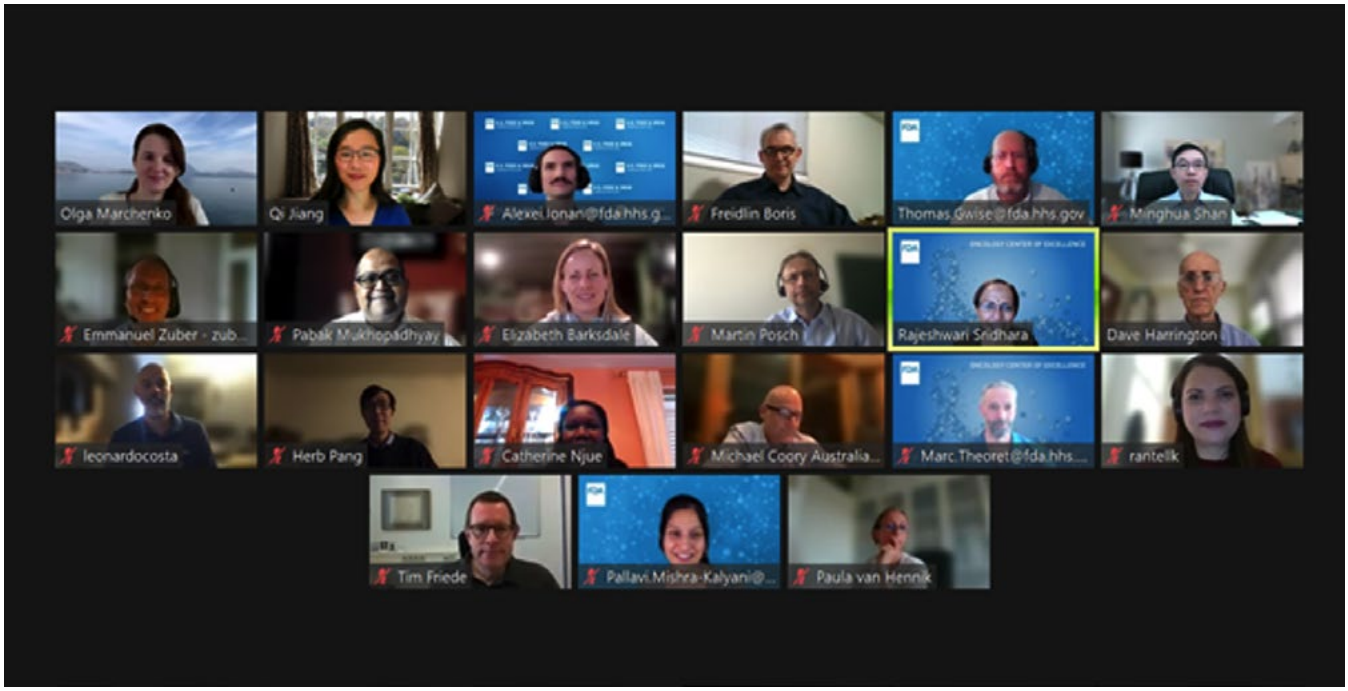
The speaker from industry laid out a scenario where PFS was the primary efficacy endpoint and OS was a safety endpoint, and a meaningful benefit in PFS was demonstrated. In this scenario, ruling out any detrimental effect (HR > 1) in OS would require a large sample size and may not be practical. The speaker proposed a potential approach utilizing pre-specified decision criteria of ruling out a 'substantial OS detriment,' mentioning that what is substantial needs to be prespecified in the statistical analysis plan. A suggestion was also made by the speaker to use Bayesian posterior probability to evaluate OS.

The academic statistician discussed how, unlike commonly used nonparametric and semiparametric approaches, parametric models allow for extrapolation of the survival curve when using data from randomized controlled trials and external data. Different data sources may be combined to increase precision. Dynamic borrowing based on similarity of data could be considered to obtain robust shrinkage parameter estimation in Bayesian hierarchical models. Furthermore, the speaker suggested that one could use joint models of longitudinal and survival data to get more efficient estimates.

The FDA speaker emphasized the importance of an investigational treatment benefit-risk assessment and the clinical context of indolent diseases. OS is an important factor, but only one of many factors in the benefit-risk assessment. Identification of an optimal dose and including patient-reported outcomes, among others, may help with benefit-risk assessment. Prespecifying OS analyses at the design stage, regardless of plans for OS hypothesis testing, helps to make optimal decisions under uncertainty. Defining OS "harm" or "detriment" and reaching alignment among stakeholders at the design stage may improve clarity and planning. The cause of a potential detrimental OS effect needs to be examined carefully as it could be due to the specific conduct of the study, adverse events, additive toxicity, dosage, or might be explained by intercurrent events such as treatment discontinuations, subsequent anti-cancer therapies, etc. Implementation of the ICH E9(R1) estimand framework at the design stage and performing simulation can inform study conduct, analyses, interpretation of the results, and benefit-risk assessment.

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

- Examining cause of death is always important as early patterns consistent with a detrimental OS effect may disappear as the trial continues. It is also possible that the treatment is detrimental to a specific subgroup of patients.
- PFS may not be an appropriate primary endpoint in certain diseases/treatments. Patient reported outcomes may be more appropriate endpoints for the evaluation of clinical benefit.
- A Bayesian framework might be useful to interpret the available data.
- In chronic diseases, metrics other than hazard ratio may need to be considered: for example, survival rates at prespecified landmark times (e.g., at 1, 3 or 5 years).
- Longer follow-up and continued data collection can help in reducing uncertainty of the findings.
- Pre-specification of OS analysis and documenting cause of death is always important even when a specific hypothesis is not being tested.



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. We plan to continue with similar multi-disciplinary open forum discussions on a variety of important topics.

Acknowledgement:

Authors thank Joan Todd (FDA) for supporting the forum, and Dr. Nicole Li (BeiGene) for taking the meeting minutes.

* Speakers/ Panelists:

Dr. Elizabeth Barksdale (LUNGevity Foundation), Dr. Michael Coory (TGA, AU), Dr. Leonardo Filho (ANVISA, BR), Dr. Boris Freidlin (NCI), Prof. Tim Friede (University Medical Center, Göttingen), Dr. Nicole Gormley (FDA), Dr. Thomas Gwise (FDA), Prof. David Harrington (Harvard University), Dr. Alexei Ionan (FDA), Dr. Qi Jiang, (Seagen), Dr. Olga Marchenko (Bayer), Dr. Pallavi Mishra-Kalyani (FDA), Dr. Pabak Mukhopadhyay (AstraZeneca), Dr. Catherine Njue (Health Canada), Dr. Herbert Pang (Genentech), Prof. Martin Posch (Center for Medical Statistics, Informatics, and Intelligent Systems at the Medical University of Vienna), Dr. Khadija Rantell (MHRA, UK), Dr. Minghua Shan (Bayer), Dr. Rajeshwari Sridhara (OCE), Dr. Marc Theoret (FDA), Dr. Emmanuel Zuber (Novartis)

SUMMARY OF ASA BIOP SECTION'S VIRTUAL DISCUSSION WITH REGULATORS ON CONSIDERATIONS IN THE EVALUATION OF PROGRESSION-FREE SURVIVAL WITH INFORMATIVE CENSORING IN CANCER CLINICAL TRIALS

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

On September 8, 2022, the American Statistical Association (ASA) Biopharmaceutical Section (BIOP) and LUNGevity Foundation hosted a virtual open forum to discuss considerations in the evaluation of progression-free survival with informative censoring in cancer trials as 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.

Progression-free survival (PFS), defined as time from randomization to disease progression or death, is commonly used as the primary endpoint to demonstrate efficacy in randomized cancer clinical trials. Global regulatory agencies have accepted PFS as the primary efficacy outcome measure for regulatory decision-making in specific cancers and treatment settings. However, there is disagreement in the community as to whether PFS indicates clinical benefit, and there are challenges in measuring PFS without bias. Assessments of disease progression are conducted at discrete time points and observations in the time-to-event analysis are censored for various reasons. A key assumption when interpreting this type of analysis is that the outcomes of censored patients are similar to those of patients who remain on study. If these assumptions are not met, censoring is informative and disturbs the principle of randomization. Informative censoring introduces bias, and the magnitude of bias depends on the proportion of patients where informative censoring is identified (1,2). In this open forum, experts representing multiple disciplines discussed how to assess the impact of informative censoring on treatment effect estimate based on PFS.

The speakers/panelists* for the discussion included members of the BIOP Statistical Methods in Oncology Scientific Working Group including pharmaceutical companies; representatives from international regulatory agencies (FDA, European Medicinal Agency (EMA), Health Canada (HC), Medicines and Healthcare products Regulatory Agency (MHRA), Therapeutic Goods Administration (TGA), Brazilian Health Regulatory Agency (ANVISA)); clinical investigators; academicians; patient advocacy groups; and expert statisticians from industry. In addition, over 100 participants attended the virtual meeting, including representatives from other international regulatory agencies (Pharmaceuticals and Medical Devices Agency (PMDA), Singapore Health Sciences Authority (HAS) and Swissmedic (SMC)). 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 forum opened with an introductory presentation and remarks from OCE leadership for context. While overall survival (OS) remains the gold standard and preferred endpoint for clinical trials in oncology, PFS is used as the primary endpoint in evaluating treatment effects in certain diseases. However, a disconnect between OS and PFS findings has been observed in some clinical trials. Because of intercurrent events such as toxicity, discontinuation of treatment, or discrepancy in measurements between local and blinded independent review of radiological scans, informative censoring of observations occurs in assessing PFS.

The introduction was followed by presentations from an academic clinician and a statistician from industry. The clinician presented examples to illustrate that PFS assessments cannot be separated from biases, and that PFS is not an appropriate endpoint to evaluate treatments. In the presenter's experience and opinion, PFS does not measure clinical benefit in late-stage diseases; OS is the best measure to evaluate treatment; and time to treatment failure (TTF) is a better outcome measure with less bias than PFS.

The industry statistician focused on how informative censoring can arise from blinded independent central review (BICR) as well as local site review of radiologic scans, and the implications for evaluating treatment effect. Based on simulations, informative censoring leads to bias in BICR Kaplan-Meier curves and estimates of medians, but not in the estimate of relative treatment effect when compared with local site evaluations. Confirmation of progression can minimize bias in informative censoring. Using BICR only as an audit of local evaluation can reduce the burden of performing BICR regularly as local site evaluation provides reliable estimate of relative treatment effect. This approach is included in both FDA and EMA guidance documents.

Key points raised during the subsequent multi-stakeholder panel discussion include:

- Intent-to-treat analysis is best when there is no informative censoring.
- There is no practical interpretation of the PFS results for patients when intercurrent events lead to censoring. The relative treatment effect based on PFS is biased when the informative censoring affects the treatment arms differently. Increasing the number of patients in the study can reduce random error, and having near real-time BICR evaluations, and using double-blind studies (where possible) may reduce bias.
- TTF includes subjective components (toxicity, treatment discontinuation and starting non-protocol therapy) that are likely to be determined differently between the arms thus introducing bias, and also TTF is difficult to generalize beyond the given trial. Moreover, weighing all intercurrent events equally in the definition of TTF complicates its interpretation.
- Understanding the reasons for censoring and the censoring patterns in both the treatment and control arms, as well as potential subgroup-treatment interactions, is important for interpreting the treatment effect.
- Sensitivity analyses such as tipping point analyses can help assess the impact of informative censoring.

This forum provided an opportunity to have open scientific discussion among a diverse multidisciplinary stakeholder group, including clinicians, 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 multi-disciplinary open forum discussions on a variety of important topics.

Acknowledgement:

Authors thank Ms. Joan Todd (FDA) and Mr. Syed Shah (FDA) for their logistical and technical support, and Dr. Weidong Zhang (Sana Biotechnology) for taking the meeting minutes.

* Speakers/ Panelists:

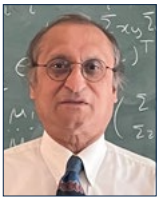
Dr. Ohad Amit (Teva Pharmaceuticals), Dr. Elizabeth Barksdale (LUNGevity Foundation), Dr. Alex Bliu (HC), Dr. Joyce Chen (FDA), Dr. Michael Coory (TGA, AU), Dr. Jaleh Fallah (FDA), Dr. Leonardo Filho (ANVISA, Brazil), Dr. Boris Freidlin (NCI), Dr. Christine Gause (Merck), Dr. Qi Jiang (Seagen), Dr. Sumithra Mandrekar (Mayo Clinic), Dr. Olga Marchenko (Bayer), Dr. Gregory Pond (McMaster University), Dr. Khadija Rantell (MHRA, UK), Mr. Andrew Raven (Health Canada), Dr. Steven Snapinn (Statistical Consultant), Dr. Rajeshwari Sridhara (OCE), Dr. Ian F. Tannock (Princess Margaret Cancer Centre, University of Toronto, Canada), Dr. Marc Theoret (FDA), Dr. Zachary Thomas (Eli Lilly), Dr. Andrew Thomson (EMA), Dr. Qing Xu (FDA)

References:

1. Fojo T, Simon RM. Inappropriate Censoring in Kaplan-Meir analysis. Lancet oncology 2021, 22:1358-1360

2. Tannock IF, Pond GR, Booth CM. Biased evaluation cancer drug trials – How use of progression-free survival as the primary endpoint can mislead. JAMA Oncology 2022, E1-E2 Published Online: March 10, 2022. doi:10.1001/jamaoncol.2021.8206

Big congratulations to the BIOP members who became **ASA FELLOWS** in 2022! Your hard work and influence are really valuable to the BIOP community and you continue to be our inspiration.



Lakshminarayan K. Choudur

Stevens Institute of Technology

For outstanding, sustained, and wide-ranging practical contributions involving the application of statistics in engineering, biology, and business, as well as dissemination of statistical knowledge through mentoring, workshops, and consulting.

Stephen W. Gulyas

Everest Clinical Research Services Corporation

For insightful applications of statistics in pharmaceutical research and health care; for longstanding commitment to the dissemination of statistical knowledge and mentoring; and for extraordinary leadership and innovation in industry and the ASA.

Mark S. Levenson

US Food and Drug Administration

For his statistical leadership and mentoring of statisticians at FDA, the public health impact of his research in drug safety evaluation and real-world evidence, and his service to the profession.

Steven J. Novick

AstraZeneca

For exceptional leadership, service, and commitment to the profession and for contributing influential and impactful methods in the biopharmaceutical space to improve discovery, development, and manufacturing quality.

Satrajit Roychoudhury

Pfizer

For major impact in developing and implementing novel statistical methods in high-priority medical settings, advancing statistical knowledge through numerous courses on cutting-edge topics, and substantial service to the ASA Biopharmaceutical Section.

Xian-Jin Xie

University of Iowa

For leadership in creating and building statistical groups in medical and dental schools; for exceptional collaboration in medical science; for outstanding teaching; and for the development of methods for design and analysis in cancer research.

Jose Zubizarreta

Harvard Medical School

For outstanding contributions to the design and analysis of observational studies, particularly on matching and weighting methodologies; for excellence in teaching of causal inference; and for dedicated mentoring of young scholars.

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 27to 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 WNAR

The WNAR 2023 will be held in Anchorage Alaska from June 15th 17th ,2023. To register, please visit: *https://wnar.org/wnar2023*

2023 Symposium on Data Science & Statistics (SDSS)

The American Statistical Association invites you to join us at the sixth annual Symposium on Data Science and Statistics in St. Louis, Missouri, May 23–26, 2023. SDSS provides a unique opportunity for data scientists, computer scientists, and statisticians to come together and exchange ideas. To find out more visit: *https://ww2.amstat.org/meetings/sdss/2023/*

- **Early Registration Closes:** April 20, 2023
- **Hotel Reservations Deadline:** May 1, 2023
- Registration via PDF files closes May 15, 2023. Online registration will remain open through the end of the conference.

JSM 2023

The Joint Statistical Meetings (JSM) is the largest gathering of statisticians and data scientists held in North America. JSM 2023 will be held in Toronto, Ontario, Canada from August 5–10, 2023. It is one of the broadest ranging conferences in statistics. To register, please visit: *https://ww2.amstat.org/meetings/jsm/2023/index.cfm*.

- **Early Registration:** May 1–31, 2023
- **Speaker Registration Deadline:** May 31, 2023
- **Regular Registration:** June 1–29, 2023
- **Late Registration:** June 30, 2023 onward
- **Housing:** May 1–June 30, 2023

2023 Nonclinical Biostatistics Conference

The ASA Biopharmaceutical Section Nonclinical Biostatistics Conference is hosted by the ASA Biopharmaceutical Section in cooperation with the Rutgers University Statistics Department. The biennial conference lasts four days with invited and contributed talks on nonclinical biostatistics topics with speakers from industry, regulatory, and academia. Two short courses are offered on the first conference day. Conference will be held at Rutgers University from June 19-21, 2023. Registration for the conference is open. Please visit the conference page here: *https://community.amstat.org/biop/events/ncb/index*