



Volume 30, No. 3 • Winter 2023

BIOPHARMACEUTICAL REPORT

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

CONTENTS

Overview of Artificial Intelligence (AI) in the Pharmaceutical Industry and Healthcare – A Paradigm-Shift

Mark Chang, PhD, Boston University..... 2

AI in Pharma for Personalized Sequential Decision-Making: Methods, Applications, and Opportunities

Yuhan Li (University of Illinois Urbana-Champaign), Hongtao Zhang (Merck & Co., Inc.), Keaven Anderson (Merck & Co., Inc.), Songzi Li (Agenus Inc.), and Ruoqing Zhu (University of Illinois Urbana-Champaign) 14

Big Data, Digital and Artificial Intelligence (AI) in a Patient-Centric Era: Perspectives of the Most Valuable Data & Insights Initiative Award Winner

Kelly H. Zou, Viatrix Inc..... 27

Using Bayesian Network to Identify Statistically Causal Clinical Biomarkers in High Dimensional Datasets

John Kang, I, Wei Wei | Merck & Co., Inc..... 33

Brief Overview of Dose Optimization in Oncology

Kentaro Takeda (Astellas), Yusuke Yamaguchi (Astellas)..... 39

2023 JSM session on "Clinical and Nonclinical Statistics Roles in the Pharmaceutical Industry: How Knowledge of One Helps the Other"

Stan Altan, W. Scott Clark, Imola Fodor, Cyrus Hoseyni, John Kolassa, Eve Pickering, and Shanthi Sethuramen..... 48

Report on Pharmaceutical Industry Survey of Nonclinical Statistics Resources

NCBLF Communications Committee: Stan Altan, W. Scott Clark (Chair), Katherine Giacoletti, Kristi Griffiths, John Kolassa, Katja Remlinger, Chi-Hse Teng, Donghui Zhang..... 55

Summary of ASA BIOP Section's Virtual Discussion with Regulators on Consideration of Criteria for Evaluation of Surrogate Endpoints

Rajeshwari Sridhara (FDA), Olga Marchenko (Bayer), Qi Jiang (Seagen), Elizabeth Barksdale (LUNGeVity Foundation), Yiyi Chen (Seagen), Marc Theoret (FDA)..... 58

Summary of ASA BIOP Section's Virtual Discussion with Regulators on Cancer Clinical Trial Design and Analysis Considerations in Evaluating Treatment Effect in Marker Negative Population – Part 2

Rajeshwari Sridhara (FDA), Olga Marchenko (Bayer), Qi Jiang (Seagen), Elizabeth Barksdale (LUNGeVity Foundation), Yiyi Chen (Seagen), Marc Theoret (FDA)..... 61

Report from the Cell and Gene Therapy Scientific Working Group

Alan Y Chiang (Lyell Immunopharma), Daniel Li (Bristol Myers Squibb), Zhenzhen Xu I (FDA)..... 64

Important Information for Prospective ASA Fellows

Paul Gallo, Brenda J Crowe, Bruce Binkowitz, Ilya Lipkovich, and Amy Xia, and updated by Paul Gallo, Bruce Binkowitz, Weili He, Inna Perevozskaya..... 70

Check Out the Latest Episodes of the ASA Biopharmaceutical Section Podcast

Christina Nurse (Takeda), Rebecca Wilson (J&J)..... 71

A Successful Conference Season

Brian Millen, Chair, ASA Biopharmaceutical Section (Biogen)..... 72

Regulatory Industry Statistics Workshop 2023 Recap

Erik Bloomquist, 2023 RISW Co-Chair (Merck)..... 75

Recap of 2023 Boston Pharmaceutical Symposium

Jianjun Hua (Dartmouth College), Zhaoyang Teng (Servier), Gautier Paux (Sanofi), Xihao Li (UNC-Chapel Hill), Kristin Baltrusaitis (Center for Biostatistics in AIDS Research), Wenting Cheng (Biogen)..... 77

ISBS to Hold Symposium in Baltimore

ISBS 2024 7th Symposium Organizing Committee..... 79

Congratulations to the New ASA Fellows..... 80

Upcoming Conferences..... 81

Notes from the editors

Holiday season is right around the corner! With the year 2023 wrapping up, hope you have all shared many productive and happy moments with colleagues, family and friends. The recently held JSM and RISW have successfully made many connections for statisticians, data scientists and quantitative researchers alike. In the meantime, discussions about new research in all fronts were happening there. For this fall issue of the biopharmaceutical report, our theme is "The Artificial Intelligence Era for Biopharma". We are joining the ongoing discussion on the role of artificial intelligence and machine learning (AI/ML) in drug and device development. So let the drum roll begin!

Our featured articles that highlight this issue's theme start with an overview of AI in the pharmaceutical industry and healthcare – a paradigm shift, contributed by **Mark Chang** (Boston University). In this article, he provided a very clear and unique way of distinguishing the classical statistics mindset vs the AI mindset when thinking about drug development. Next, the conversation continues with the review article by **Yuhan Li** (UIUC), **Hongtao Zhang** (Merck), **Keaven Anderson** (Merck), **Songzi Li** (Agenus) and **Ruoqing Zhu** (UIUC), in which they discussed AI in pharma for personalized sequential decision making. Our third featured article is by **Kelly Zou** (Viatrix). She gave perspectives on big data, digital and artificial intelligence in a patient centric era that many of us can learn from. Zooming in AI/ML's impact on a more specific area, we all know that biomarker development is oftentimes AI/ML centric. **John Kang** (Merck) and **Wei Wei** (Merck) shared with us how to use Bayesian networks to identify statistically causal clinical biomarkers in high dimensional datasets. To finish off our discussions on statistical innovation and dose optimization from last issue, **Kentaro Takeda** (Astellas) and **Yusuke Yamaguchi** (Astellas) brought us an overview of dose optimization in oncology and recent developments in this field.

For career development of clinical and non-clinical statisticians, **Stan Altan** (Janssen), **W. Scott Clark** (Eli Lilly), **Imola Fodor** (Genentech), **Cyrus Hoseyni** (Janssen), **John Kolassa** (Rutgers University), **Eve Pickering** (Pfizer) and **Shanthi Sethuramen** (Eli Lilly) summarized the 2023 JSM session on "Clinical and Nonclinical statistics roles in the pharmaceutical industry: how knowledge of one helps the other". There were a lot of interesting questions and answers in the session that could help many members navigating careers. In addition, the Nonclinical Biostatistics Leaders Forum communications committee, with members **Stan Altan** (Janssen), **W. Scott Clark** (Eli Lilly), **Katherine Giacoletti** (Merck), **Kristi Griffiths** (Eli Lilly), **John Kolassa** (Rutgers University), **Katja Remlinger** (GSK), **Chi-Hse Teng** (Novartis), **Donghui Zhang** (Sanofi), shared the report on pharmaceutical industry survey of nonclinical statistics resources.

Later in this issue, we have summaries of two 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 two discussions are: 1) Consideration of Criteria for Evaluation of Surrogate Endpoints, 2) Cancer Clinical Trial Design and Analysis Considerations in Evaluating Treatment Effect in Marker Negative Population – Part 2. The Cell and Gene Therapy Scientific Working Group, led by **Alan Y Chiang** (Lyell Immunopharma), **Daniel Li** (BMS) and **Zhenzhen Xu** (FDA), brought us the updates from their group, including the introduction and regulatory guidelines in the space which could benefit many of our BIOP members. Next we have information for perspective ASA fellows and nominators on the nomination process, shared by **Weili He** (Abbvie) on behalf of the ASA BIOP Fellows Nomination Committee. **Christina Nurse** (Takeda) and **Rebecca Wilson** (J&J) gave us the pointers to the ASA BIOP section's podcast, and a few upcoming podcasts with interesting discussion topics. As we mentioned at the beginning here, there were many great conferences in the last few months. Our ASA BIOP session chair **Brian Millen** (Biogen) brought us his recollection of JSM and RISW this year with shared photos, and **Erik Bloomquist** (Merck, 2023 RISW co-chair) shared with us his recap of the RISW conference, also with photos! In addition to JSM and RISW, there are many other conferences with invigorating discussions. Please take a look at the recap of 2023 Boston Pharmaceutical Symposium, shared by **Jianjun Hua** (Dartmouth), **Zhaoyang Teng** (Servier), **Gautier Paux** (Sanofi), **Xihao Li** (UNC Chapel Hill), **Kirstin Baltrusaitis** (Harvard) and **Wenting Cheng** (Biogen). Organizers of the 7th symposium on statistical innovation for medical product development, which is part of the international society for biopharmaceutical statistics, gave an overview of the conference, which will be held March 6-9 in Maryland, USA. We would like to send congratulations to the BIOP sections newest ASA fellows **Margaret Gamalo** (Pfizer), **Bo Huang** (Pfizer), **Xiaofeng Wang** (Cleveland Clinic) **Inna Perevozskaya** (GSK) and **Kalyan Ghosh** (Inference)! We close this issue with a list of upcoming conferences that may be of interest to you. Thank you to all of our contributing authors and ASA colleagues for your continued support!

OVERVIEW OF ARTIFICIAL INTELLIGENCE (AI) IN THE PHARMACEUTICAL INDUSTRY AND HEALTHCARE – A PARADIGM-SHIFT

Mark Chang, PhD, Boston University, Boston MA

Thank you for the invitation of the editorial board, I would like to humbly provide an overview of Artificial Intelligence (AI)/Machine Learning (ML) in the pharmaceutical Industry and Healthcare, covering the following topics: AI landscape in the Pharmaceutical Industry, AI and ML versus Classical Statistical Approaches, Challenges in Classical Statistics, AI solutions to the Challenges and Applications in Clinical Trials. Summary and Discussions. I have also included regulatory aspects and suggestions for future work. The contents are primarily based on two recently published books: (1) *Artificial Intelligence for Drug Development, Precision Medicine, and Healthcare*[1] and (2) *Foundation, Architecture, and Prototyping of Humanized AI*[2].

AI Landscape in the Pharmaceutical Industry

AI and ML for Drug Development have recently attracted great interests. According to NetBase Quid, 2022 AI Index Report[3], the focus areas with the top private-investments in AI from 2017 to 2021 are

medicine and healthcare, at nearly \$30 billion, followed closely by Data Management. Nearly 18,000 publications can be found in medical AI at PubMed.com; publications in the top 3 research areas, pathology, radiology, and surgery, have reached over 8,000 (Figure 1).

Today, AI research and applications cover the entire landscape of the pharmaceutical Industry, from Drug Discovery to Preclinical Research, Clinical Trials, Pharmacovigilance, and Drug Manufacturing (Figure 2).

AI applications include quantitative structure-activity relationships (QSARs) in drug discovery[4], cancer prediction using microarray data[5], deep learning for medical image analysis[6], healthcare[7], clinical trials[8], and drug safety monitoring[9]. AI methods for gene expression data analysis can be used for disease diagnosis and prognosis and have provided opportunities for early effective treatment and optimal treatment for individual patients based on their disease stage and other characteristics. Cancer is a worldwide genetic-related disease, which imposes significant mortality and cost. Cancer has been characterized as a heteroge-

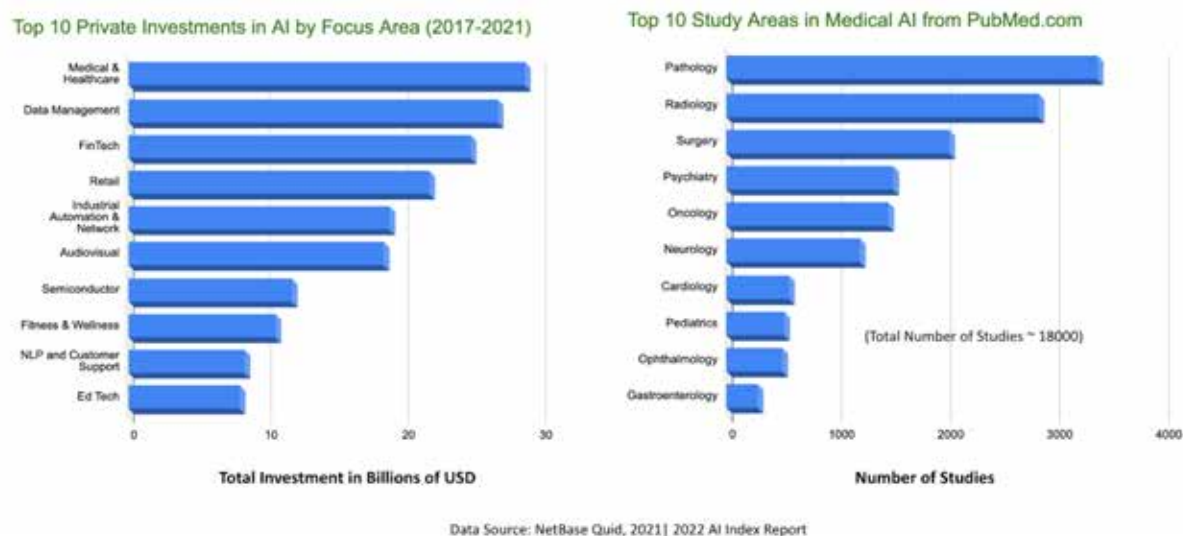


Figure 1: Medical AI Research and Investment

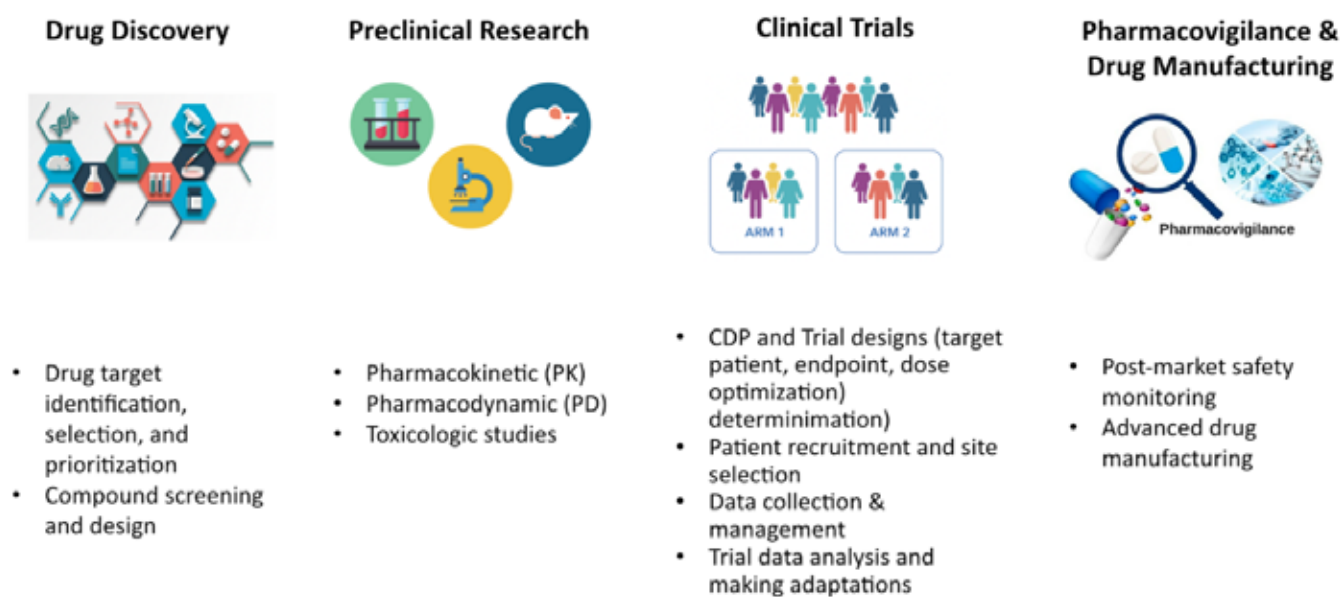


Figure 2: AI Research and Applications Across the Drug Development Landscape

neous disease consisting of many different subtypes. The early diagnosis and prognosis of a cancer type have become a necessity in cancer research, as these can facilitate the subsequent clinical management of patients. The importance of classifying cancer patients into high- or low-risk groups has led to much research in AI methods to model the progression and treatment of cancerous conditions.

AI research in healthcare can help physicians make better clinical decisions or even replace human judgment in certain functional areas of healthcare. The increasing availability of healthcare data and the rapid development of big data analytic methods have made possible the recent successful applications of AI in healthcare[7]. In addition to the popular deep learning neural network strategies for structured data, natural language processing for unstructured data, such as that in physicians' notes, deep learning such as particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have been used to improve the accuracy of disease diagnosis. They can analyze medical images such as X-rays, MRIs, and CT scans to detect conditions like cancer, pneumonia,

and diabetic retinopathy. AI systems can provide up-to-date medical information from publications and large patient populations to assist physicians making real-time inferences, for health risk alerts and health outcome prediction, and in reducing diagnostic and therapeutic errors.

The applications of AI in clinical trials include stochastic decision processes for clinical development programs and similarity-based machine learning, but these are limited today. To solve this problem effectively requires a paradigm shift: a focus on the prediction of drug effects (efficacy and safety) instead of type-I error control. Similarity-based machine learning (SBML) provides a powerful AI approach for small and big data in clinical trials and in scientific discoveries generally[8].

In drug safety and pharmacovigilance, Sparkes, among other researchers, discusses the role of artificial intelligence within pharmacovigilance and medical information[10]. Last but not least, (electronic) data collection, fusion, and sharing all play an integral and critical part in successful applications of AI technologies.

technologies.

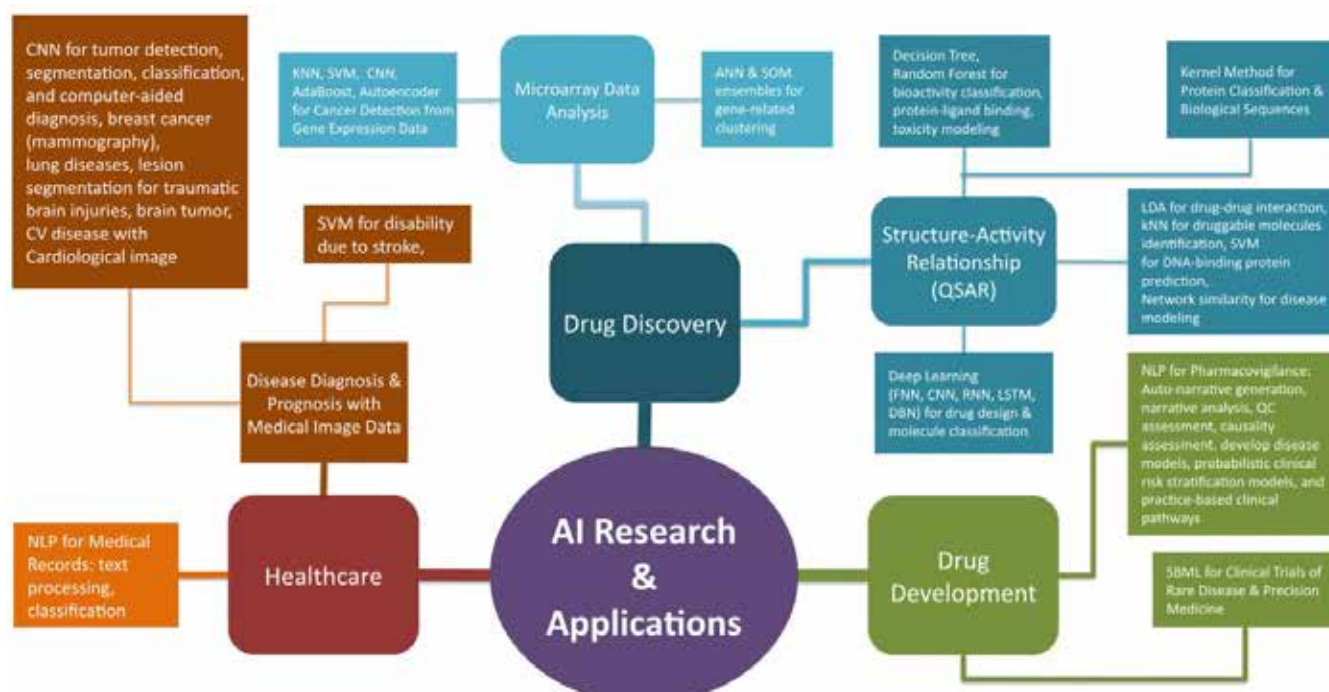


Figure 3: Summary of AI Applications in the Medical Field

AI Applications in the medical field are summarized via the tree diagram (Figure 3), including drug discovery, drug development, and healthcare. Details, including R-code and extensive references, can be found in Chang’s book[1].

AI/ML versus Classical Statistics

ML is the core AI, focusing on the data model part instead of user-interfaces. ML and Classical Statistics (CS) have many similar aspects, but AI/ML is clearly not equal to statistics. Here are some subtle differences: (1) AI/ML emphasizes learning and prediction, whereas classical statistics often focuses on type-I error rate; thus CS tends to use all data to get a minimal p-value, while ML splits available data into training and validation/evaluation sets, (2) CS and ML both deal with uncertainty, but the former focuses on the mathematical approach and probability distributions, while the latter is mainly algorithm-based, (3) ML takes its main aim at real world experiences with an unclearly defined target population, while CS often performs under an ideal assumed probability

distribution for the target population, (4) AI/ML can often deal with big and unstructured data, CS does not, (5) Data mining, NLP, technology and design platforms are also topics in ML, but not in CS, and (6) AI and ML encourage continuous exploration, whereas CS limits reusability of data due to multiplicity problems in hypothesis testing (Figure 4).

ML methods can be classified into five general categories: supervised, unsupervised, reinforcement, evolutionary, and swarm intelligence learning methods (Figure 5).

A typical task for supervised learning is classification, e.g., noting when there is disease or no disease. In supervised learning, the learner will give a response y based on an input x and will be able to compare his response y to the target (correct) response. In other words, the “learner” presents an answer y for each x in the training sample, and the supervisor provides either the correct answer or an error associated with the learner’s answer. The term learning here refers to the learner (a model) adjusting its parameters to reduce the error by using the training dataset. The trained AI model

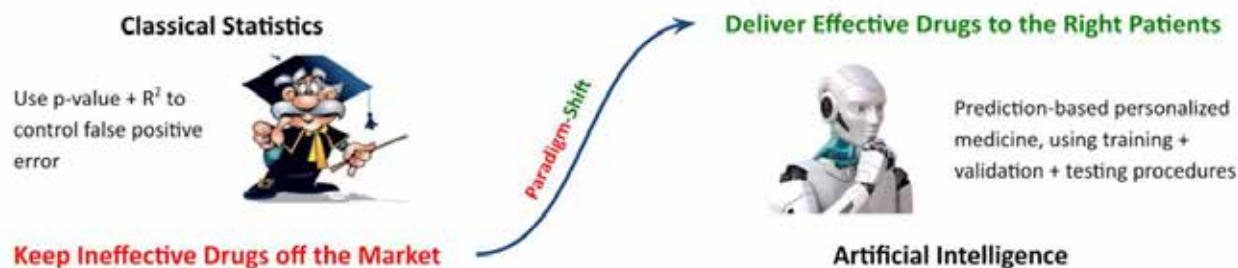


Figure 4: How AI/ML Approaches Differ from Classical Statistics

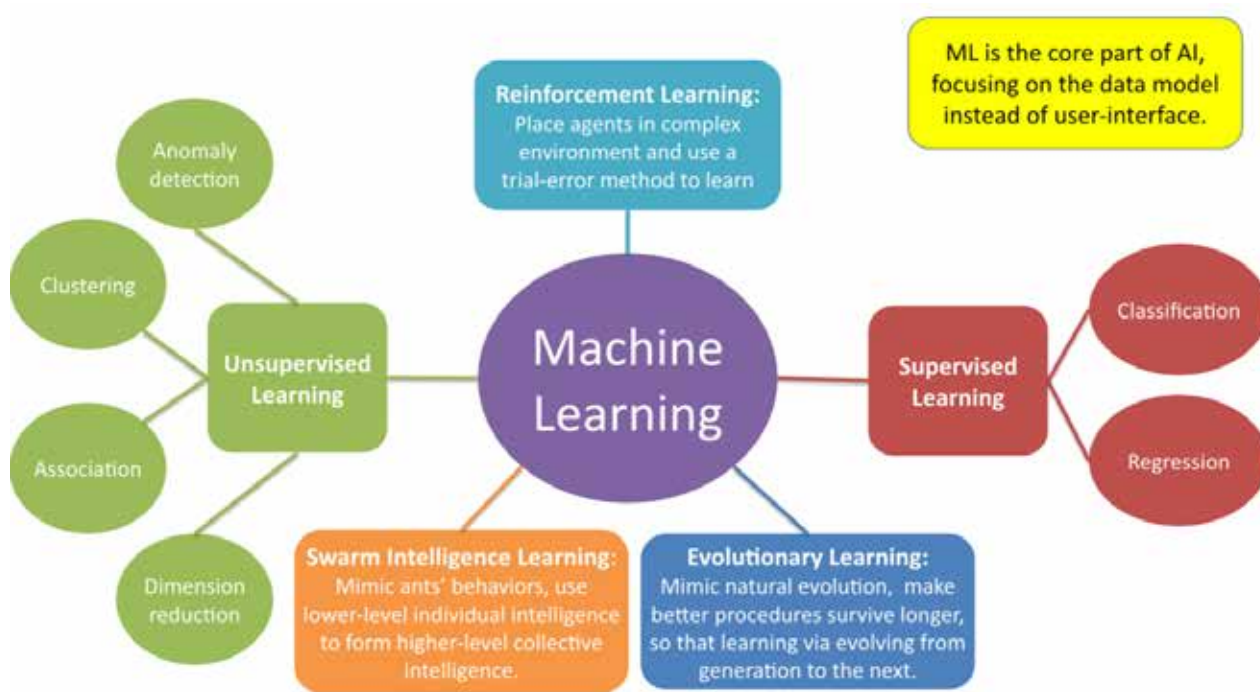


Figure 5: Types of Machine Learning Approaches

can be used for future predictions. Supervised learning has been used in disease diagnosis, drug safety signal detection, and other medical fields[11][12].

Two typical tasks of unsupervised learning are document clustering and information retrieval. In unsupervised learning, the learner receives no feedback from the supervisor at all. Instead, the learner's task is to re-represent the inputs in a more efficient way, for instance, as clusters or with a reduced set of dimensions. Unsupervised learning is based on the similarities and differences among input patterns. The goal is to find hidden structures in unlabeled data without the help of a supervisor

providing a correct answer. In drug development, unsupervised learning is often used for data preprocessing before adopting supervised learning[13][14].

Reinforcement learning (RL) concerns how a learner should take actions in an environment so as to maximize some notion of long-term reward. RL gets feedback from real-world experiences; its algorithms attempt to find a policy (or a set of action rules) that maps states of the world to the actions the learner should take in those states. Unlike supervised learning, in RL the correct input-output pairs are never presented. Furthermore, there is a focus on on-line performance, which involves

finding a balance between exploration of uncharted territory and exploitation of one's current knowledge. RL has been suggested for drug development programs, including design molecular structures with desired properties, autonomous clinical trials in oncology, and a clinical development planning[1][15][16].

Biological evolution can be viewed as a learning process: how biological organisms have offspring and adapt to their environment can improve the probability of the species' survival and success. Inspired by such biological evolutionary mechanisms, genetic programming (GP) was developed. In GP, computer programs are encoded as a set of genes that are then modified (evolved) using an evolutionary algorithm. The methods used to encode a computer program in an artificial chromosome and to evaluate its fitness with respect to the predefined task are central in the GP technique. Ghaheri et al. introduced the genetic algorithm and its applications in medicine and provided a comprehensive review of the applications of genetic algorithms in medicine in 15 different areas, including disease screening, diagnosis, treatment planning, pharmacovigilance, prognosis, and health care management[1][17].

Systems in which organized behavior arises without a centralized controller or leader are called self-organized systems. The intelligence possessed by a self-organized system is called Swarm Intelligence (SI) or Collective Intelligence. Artificial SI is an emerging field of biologically inspired artificial intelligence characterized by micro motives and macro behavior. A good example of Swarm Intelligence is that of ant colonies, which optimally and adaptively forage for food. Ants are able to determine the shortest path leading to a food source simply by following pheromones. This works only because the shortest path will have more ant traffic and stronger pheromone scents than other paths. A new docking algorithm called PLANTS (Protein-Ligand ANTSystem) is developed based on ant colony optimization, to facilitate structure-based drug design[18]. An artificial ant colony is employed to find a minimum energy conformation of the ligand in the protein's binding site[1][17].

Challenges in Classical Statistics Calling for AI/ML

We now discuss the challenges we are facing in CS and the benefits of switching to the AI/ML paradigm. In the first dilemma in decision-making, I will show the practical invalidity of virtually all one-sided rank tests and further illustrate that the so-called right choice in our daily life could be just an illusion. In the second

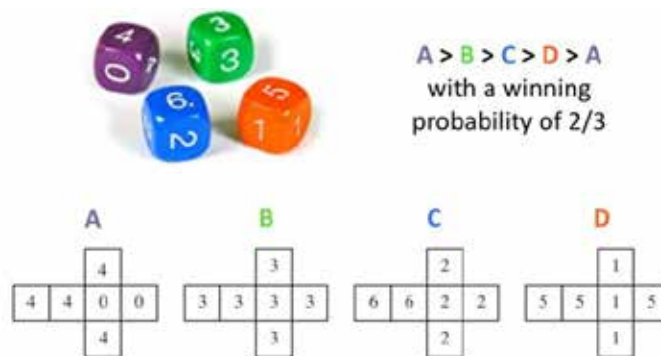


Figure 6: A set of Four Intransitive Dice

paradox, we will show how CS can arrive at completely opposite conclusions from the same data collected from the same patients when different clinical trial designs are used. The third paradox is to reveal the controversies in CS interpreting the effect of each patient characteristic. All these suggest the need for a paradigm shift from CS to AI/ML.

A Shocking Paradox in Decision-Making

Each of us has to make many choices in our lives, from the trivial to life-changing. Choices can be emotional or rational. Here we are interested in the latter, rational choices. We illustrate our point with Efron's intransitive dice (Figure 6).

Efron's dice are the four dice A, B, C and D with the following numerals on their six faces: A displaying {4, 4, 4, 0, 0}, B with {3, 3, 3, 3, 3}, C having {6, 6, 2, 2, 2}, and D, {5, 5, 5, 1, 1}. It can be easily proved that die A beats die B; B beats C; C beats D, and D beats A, all with the same probability of 2/3. Therefore, the four dice are equally good.

Now imagine if the numbers represent the evaluation scores of the four social systems (or products, medical interventions) at six different times or aspects. If we are provided with social system options A, B and C without knowing the existence of option D, we might think A is the right choice, but actually the four choices are equally good. The conclusion can be applied to our decision-making in other situations, such as medical treatments of a certain disease. In this case, different dice may



Figure 7: Simpson's Paradox in Running Clinical Trials

present different treatments A, B, C, and D, whereas the face values of a die may indicate the responses of different patients to that treatment. Without knowing the possible treatment D, we would conclude A is the best treatment after we run a clinical trial. However, in fact, all four treatments can be equally good. These examples seem to make us completely lose confidence in virtually any decision we have made or are going to make in our daily life. Therefore, the 'right' decision might be just an illusion in the eyes of the decision-maker[2][19]. Note that the effect of any medical treatment and comparison conclusions are independent of the availability of other treatments, as long as they theoretically exist.

There are many other sets of intransitive dice consisting of three or more dice. For instance, the set of three dice, Red {3, 3, 3, 3, 3, 6}, Blue {2, 2, 2, 5, 5, 5}, and Olive {1, 4, 4, 4, 4, 4} is intransitive. Intransitive dice do not have to be 6-faced and the numbers do not have to be integers. The set of dice of {1, 4, 4, 4}, {2, 2, 5, 5}, and {3, 3, 3, 6}, and the set of {1, 1, 4, 4, 4, 4, 4, 4}, {2, 2, 2, 5, 5, 5, 5}, and {3, 3, 3, 3, 3, 3, 6, 6} are two more sets of intransitive dice. We can replace 2 by a number such as 2.3 or 2.8, or even by a random variable distributed between 2 and 3, without any impact on the conclusion.

This paradox implies that virtually all one-sided rank methods that are commonly used in clinical trials practically fail miserably since a one-sided rank-test can at most tell treatment groups involved are different but cannot tell which one is better.

Unresolved Simpson's Paradox in Clinical Trials with Classical Statistics

We are going to use Simpson's paradox to show you that we could arrive at completely opposite (but both statistically significant) conclusions on medical treatment effects based on a very same set of data from the same patients.

Suppose we have options to run a randomized clinical trial with two different treatments on the same patients: Option 1, a single large trial with male and female patients and Option 2, the same patients but treated in two gender-specific trials (Figure 7). Given the data in Figure 6, if Option 1 is chosen, we will conclude that the Test drug is significantly worse than Control with response rate 79% versus 83% ($p < 0.005$), respectively. However, if Option 2 is chosen, with the same data as in Option 1 we will arrive at a completely opposite conclusion: Test drug has statistically significantly better effects in the both females (74% vs 67% with $p < 0.005$) and the males (93% vs 87% with $p < 0.005$).

Would you conclude that the Test drug fails when Option 1 is actually adopted, or will you do subgroup analyses? Remember, the population can be further divided into smaller groups based on other characteristics of the patients. When should such subgroup analyses be stopped? The same dilemma can arise in multiregional (global) trials, in which different regions will show different treatment effects[19][20].

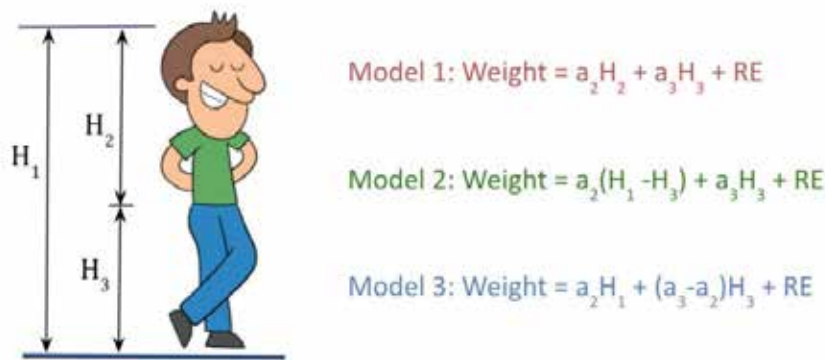


Figure 8: Controversies in Modeling Body Weight by Heights

Paradox in Interpreting Effects from Statistical Modelling

Controversies in the interpretation of Associative or Causal Effects arise when using a classical statistical model, since the effects will depend on attributes (independent variables) engaged. For instance, to study how the lower body height will affect the body weight, we use the initial model (Figure 8) that includes the attributes upper body height (H_2) and lower body height (H_3), and a random error (RE). This initial model can be mathematically rewritten as Model 2, and further reduced to Model 3 with attributes: the height H_1 and the lower body height H_3 . We have to point out that Model 1 and Model 3 are mathematically equivalent. However, the interpretations of how much H_3 affects the weight can be very different. Using Model 1, we would conclude that “every inch increase in the lower body will lead to an a_3 pound increase in body weight”, while using Model 3, we would conclude with: “every inch increase in the lower body will lead to an $a_3 - a_2$ pound increase in body weight”. Even when the models fit to the data perfectly ($RE = 0$), different models will lead to different conclusions regarding the effect of a given factor/attribute. The effect of an attribute depends generally on the mathematical form chosen!

As pointed out by Chang[2], the reason that different models give different interpretations of the effects of individual attributes is because the attributes can be associated (H_1 has already included H_3). Likewise, in life sciences we often study how phenotypes (human behaviors) depend on genotypes, while genotypes are associated. Such associations make the interpretation of

attribute-effect subjective, depending on what genes or other attributes one wants to be included in the model.

We can easily list more controversies in CS, for example in hypothesis-test-based adaptive trials, an interim smaller p-value can fail to reject the null hypothesis, while a larger p-value at the final analysis can reject the null hypothesis.

AI Solutions and Applications in Clinical Trials

The controversies we have discussed above can be overcome or reduced through AI/ML, such as similarity-based ML approaches (SBML). Before we introduce SBML techniques, we discuss the similarity-principle[1][16][19].

Similarity Principle Based ML

The similarity principle is a hidden principle that we most commonly use, whether consciously or subconsciously, in our lives every day. The principle can be stated as: similar things or individuals will likely behave similarly, and the more similar they are the more similarly they behave. Here are three simple examples from drug discovery, preclinical research, and clinical trials: (1) ligands with similar structures will behave similarly or have a similar mechanism of action, (2) the effectiveness of a drug on patients can be somewhat predicted based on its effects on animals, and (3) future patients who are similar in disease, gender, and age to the clinical trial patients will likely have similar responses to the medical intervention.

We all might agree that to qualify as a true scientific discovery, a finding must be verifiable. Otherwise, it cannot be called science. However, as history is unique,

- Informational contributions based on their similarities to Lisa.
- Similarities determined via training (attribute-scaling factors).
- SBML works even when data is small (rare diseases)

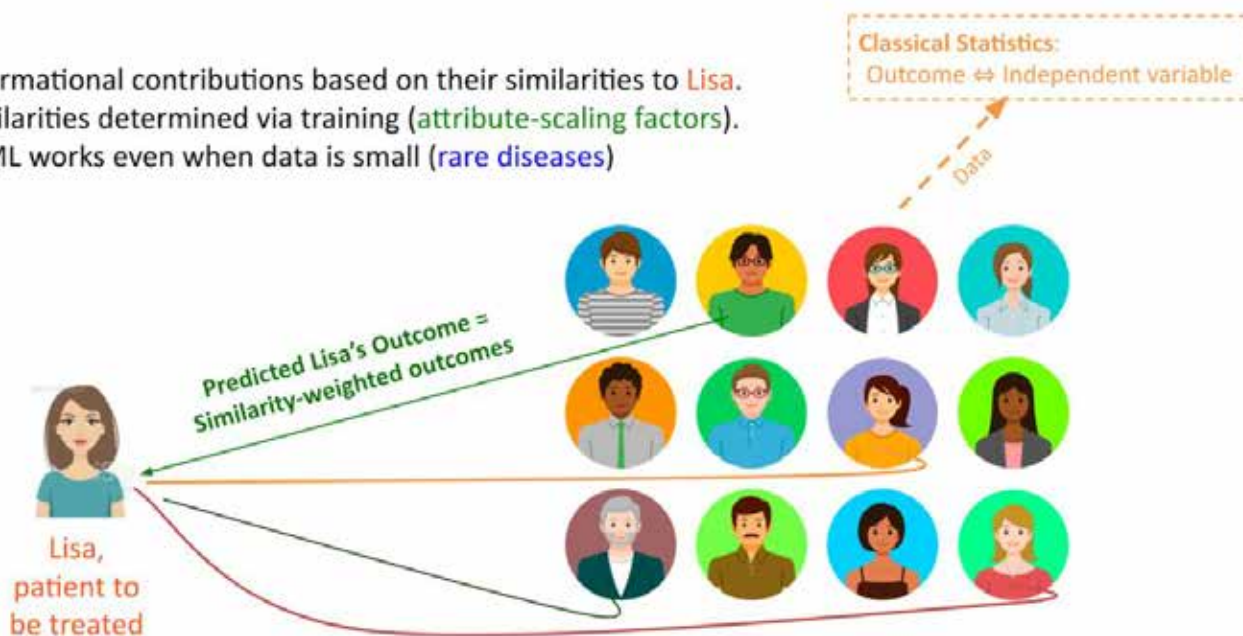


Figure 9: Similarity-Based Machine Learning in Predicting Drug Effects on a Future Patient

no two events are identical or repeat exactly; even the same individual (especially a living being) will change constantly. For this reason, we have to group similar things together and, considering them as approximately the same, study their common or overall behaviors. Pharmaceutical scientists treat people with the “same” disease to study the overall effect of a drug even though individual responses to the drug may be different. Indeed, similarity grouping is the basis for scientific discovery and prediction and the similarity principle is the backbone behind causality.

We are now ready to illustrate SBML. In a statistical approach, we usually link the outcome (dependent variable) to the attributes (independent variables) directly using a preselected model. The success of this approach very much depends on prior knowledge on the model selection. Even with good prior knowledge, it is still very difficult if not impossible to resolve the previously mentioned paradoxes. In contrast, SBML seeks to project the outcomes for patients of interest based on the outcomes of similar patients. Specifically, with SBML we are predicting a new patient’s outcome based on similarity-weighted observed outcomes from similar patients. In this approach, the key is to determine the similarities among patients. In daily life, such similarity

is semi-subjectively determined based on prior knowledge. In SBML the similarities are objectively determined (learned) through training and the importance of each attribute in similarity-determination is learned via so-called attribute scaling factors (Figure 9).

Mathematically,

$$\text{Future Patient Response } Y = c \sum_n S_n Y_n$$

$$\text{Similarity Score can be } S_n = \exp(-\sum_k R_k d_{nk})$$

Here d_n is the absolute difference between the future patients and the n -th patient in the k -th attribute and R_k is the scaling factor associated with the k -th attribute[1][8].

SBML shows itself to be more attributes-inclusive regardless of statistical significance, and more data-inclusive since less relevant information will automatically be weighted less, and being able to deliver the right drug with the right amount to the right patient at the right time, naturally a precision medicine approach. Figure 10 shows the SBML results in comparison with other statistical modeling approaches for a clinical trial with a rare disease[8]. SBML outperforms (smaller errors) all other approaches across a range of training sample sizes.

As we can see, SBML resolves the Simpson paradox by weighting data from different patients differently in predicting a new patient’s response based on simi-

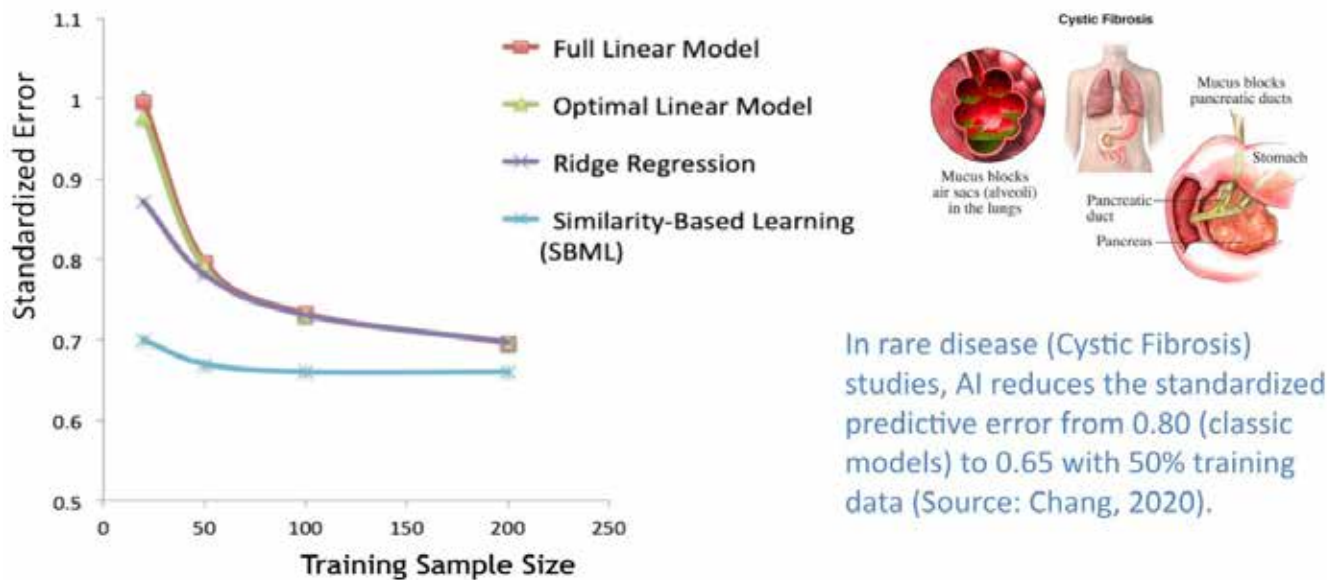


Figure 10: Similarity-Based Machine-Learning Performance

larities between new patients and clinical trial patients and by updating the conclusions as new data come in. SBML resolves the controversy in modeling the effect of an attribute by emphasizing prediction instead of interpreting the ill-defined attribute effects. SBML also reduces the chance of an intransitive-dice dilemma by focusing on predicting the response in each individual (precision medicine) instead of the effect on a heterogeneous patient group.

Concerns About AI Technologies

Despite the many positive sides of AI applications, there are concerns about the lack of transparency and accountability in AI decision-making: the potential for biased algorithms, privacy violations, and the use of AI in surveillance and decision-making processes that could harm individuals or society.

AI technology such as ChatGPT can easily generate fake news and fake data, which might lead to pseudoscience that can be very difficult to detect. Because fake stuff is much easier to produce than real data and real science, we could be buried with and eventually live in such a virtual reality. Where is the balance between authenticity and the right of free-speech? This is an imperative question we have to consider now!

Indeed, we expect AI (recommendation systems) to aid our choice and decision-making, but at the same time we feel such systems take away some of our freedoms. The two-way interactions between humans and AI have been affecting every aspect of our lives, even unnoticeably our view on AI itself. Recently, people are also concerned about job displacement by AI. All these concerns call for regulation and oversight on AI.

Some of us may be worried about our ability to change so many aspects of human hardware. We completely replace malfunctioning organs with healthy ones, so, for example, maybe we'll be able to use medical equipment to erase undesirable memories in the future. As these processes continue, are we making a human-machine mixed race? When does a person lose his or her identity in the process?

Concerns about the potential for AI systems to operate beyond human control are often referred to as the "control problem." Some of us start to fear: Can, and in what ways, might humanized AI (HAI) agents surpass human beings? Will we become unnecessary? I'd rather answer the question from a social instead of a technical perspective as most people have tried to do. During the long future course of HAI's development, we humans will develop emotions towards HAI agents as they live with us on a daily basis.

Regulatory submissions with key terms ML or AI by Therapeutic Area

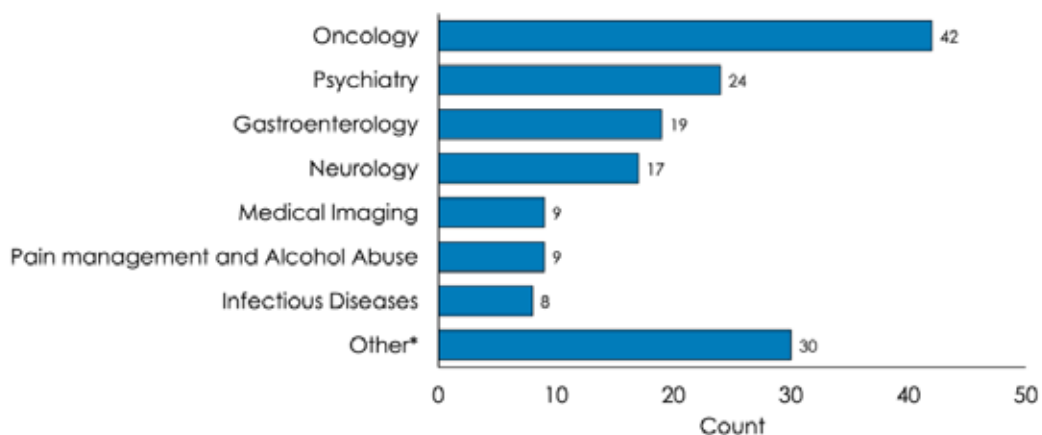


Figure 11: Regular Submissions with Key Term ML or AI by Therapeutic Area[25-28]

We will not discriminate against “anyone” because of race, color, gender, sexual orientation, or origin (machine-made or not); all that will matter are time and intellectual interactions, be they technical or emotional. The concept (connotation and denotation) of a human being, like all other concepts, is subject to the dynamics of evolution. Before we can develop the full capacity of HAI, our societal view - our definition - of mankind will have to experience dramatic modifications. HAI agents will be recognized as the machine-race of humankind. On one hand, HAI will move closer and closer to human intelligence. On the other hand, humankind becomes more and more accepting of machine-kind. The two parties will meet and unite in a middle way[2]. Technically, if super AI surpassing human beings is possible, there is no reason to believe that we all might already be super-AI made human beings.

Summary and Discussion

AI/ML has been used in pharmaceutical discovery and development since the 1990’s[14][21]. It can help drug discovery and development, health management, improving people’s health and QOL. Recently the FDA reported a summary of regulatory submissions that involve AI and ML terms in different therapeutic areas (Figure 11). However, the applications of AI in clinical trials are limited today, and one of main reasons is that

current statistical strategies implemented in the regulatory guidance are predominantly based on CS type-I error control. Adopting this approach, we are facing even bigger challenges than we’ve ever faced before. For, on the one hand, as the available drugs become ever more effective, the diminished efficacy margin for improvement requires an impractically larger sample size for clinical trials so as to control the type-I error and maintain a sufficient power for the hypothesis test; on the other hand, as awareness of disease heterogeneities and needs for precision medicine increase, the sample size available for each specific disease trial becomes even smaller. To solve this problem effectively requires a paradigm shift: a focus on prediction of drug effects (efficacy and safety) instead of type-I error control. For this, AI/ML provides a viable solution, as we have illustrated with SBML.

To accelerate the applications of AI/ML in the entire pharmaceutical industry landscape, there are additional barriers that need to be overcome: (1) the lack of basic understanding of AI technologies, (2) data privacy and security, (3) limited data availability and quality, (4) the need for regulatory clarity or guidance on AI applications. All these barriers generate concerns of business risk and abide a general resistance to change.

To overcome the barriers, the joint effects of industry, academics and government are needed. Among them, government guidance plays a crucial

role. Recently, the FDA moved toward a new, tailored review framework for artificial intelligence-based medical devices[22]. The CDER established an AI Steering Committee in 2020 to facilitate effective use and sustainment of AI in CDER's decision-making and operations. Two discussion papers to spur conversation about AI and ML in drug development & manufacturing with the goal of developing and adopting a flexible risk-based regulatory framework that promotes innovation and protects patient safety[23][24].

Strategically, to speed up AI application in clinical trials and reduce the cost and time to market we can start from existing studies and early clinical trials, and gradually move to new pivotal trials. In so doing we ought to proactively work with regulatory agencies, synergizing AI with other innovative approaches such as precision medicine, adaptive clinical trials, clinical trial simulation, real-world experiences, and stagewise market authorization with enhanced pharmacovigilance. At the same time, we should exercise AI caution to avoid pseudoscience.

No doubt, data scientists with hand-on AI/ML knowledge are very much desired. Chang[1] comprehensively discusses different methods in supervised and unsupervised learning, reinforcement learning, evolutionary learning and swarm intelligence. For most methods, each chapter of the book provides examples of applications using R. The chapter, Applications of AI in Medical Science and Drug Development, is a comprehensive review of applications of different AI methods in drug discovery, in cancer prediction using microarray data, as well as in medical image analysis, healthcare, clinical trials, and drug safety monitoring. Many different AI methods can be used for the same application problems and many different application problems can be solved using the same AI method. The R programs for the book are available on www.statisticians.org. Beyond these weak AI approaches, Humanized AI (HAI) as strong AI, in a narrow sense, embodies the development of humanized assistants, especially for seniors. In a broad sense HAI is seen as a new race of human beings, the machine-race human. To stay at the forefront of HAI, see my recent book: *Foundation, Architecture, and Prototyping of Humanized AI*[2].

References

1. Chang, M. (2020). *Artificial Intelligence for drug development, precision medicine, and healthcare*. CRC. Boca Raton, FL.
2. Chang, M (July, 2023). *Foundation, Architecture, and Prototyping of Humanized AI: A New Constructivist Approach*. CRC Press.
3. <https://aiindex.stanford.edu/report/>
4. Ghasemi, F. et al (2018). Neural network and deep-learning algorithms used in QSAR studies: merits and drawbacks. *Drug Discov Today*. 2018 Oct;23(10):1784-1790.
5. Daouda, M. and Mayo, M. (2019). A survey of neural network-based cancer prediction models from microarray data. *Artificial Intelligence In Medicine* 97 (2019) 204-214
6. Anwar, S.M., et al. (2018). Medical Image Analysis using Convolutional Neural Networks: A Review *Journal of Medical Systems* November 2018, 42:226
7. Dua, S. Acharya, U. R., Dua, P. (2014). *Machine Learning in Healthcare Informatics*, Springer; 2014 edition.
8. Hwang, S. and Chang, M (2022). Similarity-principle-based machine learning method for clinical trials and beyond. *Computer Science. Statistics in Biopharmaceutical Research*.
9. [9]Schmider, J., Kumar, K., et al (2019). Innovation in Pharmacovigilance: Use of Artificial Intelligence in Adverse Event Case Processing. *Clinical pharmacology & therapeutics* I volume 105 number 4 I April 2019
10. Sparkes, S. (2018). The role of artificial intelligence within pharmacovigilance and medical information. *PIPELINE* Issue 56; September 2018
11. Lowe, R. et al. (2011) Classifying molecules using a sparse probabilistic kernel binary classifier. *J. Chem. Inf. Model.* 51, 1539-1544

12. Lu, L. Zheng, Y., et al (2017). *Deep Learning and Convolutional Neural Networks for Medical Image Computing*. Springer, Switzerland, 2017.
13. Aliper, A. et al. (2016) Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Mol. Pharmaceut.* 13, 2524-2530
14. Rose, V.S. et al. (1991) An application of unsupervised neural network methodology Kohonen topology-preserving mapping to QSAR analysis. *Quant. Struct. Act. Relat.* 10, 6-15
15. Chang, M. (2010). Monte Carlo Simulation for the pharmaceutical industry. CRC
16. Chang M. et al. (2019). *Innovative Strategies, Statistical Solutions and Simulations for Modern Clinical Trials*. CRC
17. Ghaheri, A. et al. (2015). The application of genetic algorithms in medicine. *Oman Med J.* Nov; 30(6): 406-417.
18. Korb, O. et al. (2006). PLANTS: Application of ant colony optimization to structure-based drug design. *ANTS 2006: Ant Colony Optimization and Swarm Intelligence* pp 247{258.
19. Chang, M. (2014). *Principles of Scientific Methods*. CRC
20. Chang, M. (2012). *Paradoxes in Scientific Inference*. CRC
21. Chang, M. (2011). *Modern Issues and Methods in Biostatistics*, Springer. N.Y., N.Y.
22. Gottlieb, S. (2021). Statement from FDA Commissioner Scott Gottlieb, M.D. on steps toward a new, tailored review framework for artificial intelligence-based medical devices. <https://www.fda.gov/news-events/press-announcements/statement-fda-commissioner-scott-gottlieb-md-steps-toward-new-tailored-review-framework-artificial>
23. FDA guidance. *Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan* January 2021.
24. The U.S. Food and Drug Administration (FDA) issued the “Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device” – discussion paper. <https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf>
25. Zhu, H. Division of Pharmacometrics, FDA/CDER/OTS/OCP. Application of Artificial Intelligence and Machine Learning in Drug Development and Precision Medicine. FDA/UMD CERSI Workshop. February 2023.
26. Fakhouri, T. (FDA). CDER Office of Medical Policy. AI in Drug Development.. FDA/UMD CERSI Workshop. February 17th, 2023.
27. Q i Liu, R. H., Julie Hsieh, Hao Zhu, et al. (2022). Landscape Analysis of the Application of Artificial Intelligence and Machine Learning in Regulatory Submissions for Drug Development from 2016 to 2021. *Clinical Pharmacology & Therapeutics*.
28. Haug, C.J. and Drazen, J.M. (2023). Artificial intelligence and machine learning in clinical medicine, 2023. *N.Engl J Med* 388; 13 ■

AI in Pharma for Personalized Sequential Decision-Making: Methods, Applications and Opportunities

Yuhan Li^{1*}, Hongtao Zhang^{2*}, Keaven Anderson², Songzi Li³ and Ruoqing Zhu^{1†}

¹ Department of Statistics, University of Illinois Urbana-Champaign, Champaign, IL, USA

² Biostatistics and Decision Research Science, Merck & Co., Inc., North Wales, PA, USA

³ Biostatistics, Agenus Inc., Lexington, MA, USA

1 Introduction

In the pharmaceutical industry, the use of artificial intelligence (AI) has seen consistent growth over the past decade. This rise is attributed to major advancements in statistical machine learning methodologies, computational capabilities and the increased availability of large datasets. AI techniques are applied throughout different stages of drug development, ranging from drug discovery to post-marketing benefit-risk assessment. Kolluri et al. [1] provided a review of several case studies that span these stages, featuring key applications such as protein structure prediction, success probability estimation, subgroup identification, and AI-assisted clinical trial monitoring. From a regulatory standpoint [2], there was a notable uptick in submissions incorporating AI components in 2021. The most prevalent therapeutic areas leveraging AI were oncology (27%), psychiatry (15%), gastroenterology (12%), and neurology (11%).

The paradigm of personalized or precision medicine has gained significant traction in recent research, partly due to advancements in AI techniques [3]. This shift

has had a transformative impact on the pharmaceutical industry. Departing from the traditional “one-size-fits-all” model, personalized medicine incorporates various individual factors, such as environmental conditions, lifestyle choices, and health histories, to formulate customized treatment plans. By utilizing sophisticated machine learning algorithms, clinicians and researchers are better equipped to make informed decisions in areas such as disease prevention, diagnosis, and treatment selection, thereby optimizing health outcomes for each individual [4, 5].

In this article, we explore a range of methods and algorithms in the field of personalized medicine. While these techniques share the overarching aim of crafting personalized treatment plans, they differ in terms of problem formulations and practical applications. We delve into specific examples within the healthcare sector, categorizing them as either established in research and practice, or as aspirational approaches with potential for significant impact. The article concludes with a discussion of pertinent challenges and outlines avenues for future research.

*The first and second authors contribute equally.

†Corresponding author: rqzhu@illinois.edu

2 Methods and Applications

2.1 Optimal Treatment Sequence

Dynamic Treatment Regime Dynamic Treatment Regime (DTR) represents a cutting-edge paradigm in the realm of personalized medicine, aiming to tailor medical interventions to individual patients' evolving health status [6]. Within the context of clinical research, data concerning a DTR are usually collected from multi-stage clinical trials or longitudinal observational studies on the disease of interest [7]. These studies often involve a finite number of decision stages. An optimal DTR aims to find a sequence of decision rules that assign treatments at each stage based on a patient's baseline characteristics and historical information.

Suppose we have a pre-specified finite T decision points, indexed by $t = 1, 2, \dots, T$. Let $S^t \in \mathbb{R}^p$ represent all related patient characteristics at time t , such as age, gender, and lab results, which may vary across time points and reflect the patient's current condition. The treatment given at time t is denoted by $A^t \in \mathcal{A}$, which may include drug choice and/or dosage selected from a possible set of treatments \mathcal{A} , which could be either discrete or continuous. The potential treatment trajectory up to point t is denoted by $\bar{A}^t = (A^1, A^2, \dots, A^t)$, and $\bar{S}^t = (S^1, S^2, \dots, S^t)$ represents the cumulative information of the patient leading up to t . Realizations of such a treatment path and accumulated patient data are denoted as $\bar{a}^t = (a^1, a^2, \dots, a^t)$ and $\bar{s}^t = (s^1, s^2, \dots, s^t)$, respectively.

At each decision point, we further observe an immediate reward R^t that may depend on all the previous history leading up to t . The immediate reward serves

as an indicator of the individual's response to the selected treatment, where a larger reward signifies a more favorable response. Hence, the collected dataset is of the form $\{S_i^1, A_i^1, R_i^1, S_i^2, \dots, S_i^T, A_i^T, R_i^T, S_i^{T+1}\}_{i=1}^n$, which comprises n i.i.d. trajectories with T decision points. The objective is to identify the optimal DTR that maximizes the cumulative reward, i.e., $R = \sum_{t=1}^T R^t$, from $t = 1$ to T . Under certain scenarios, it is also possible to only observe the final reward R at the last stage, e.g., event-free survival or overall survival. In either case, the goal is to maximize R by choosing a sequence of decisions.

A DTR is defined as $\pi = (\pi_1, \dots, \pi_T)$, which forms a sequence of decision rules to treat a patient over time. The decision rule at each time point t , π_t , can be thought of as a mapping from a patient's history \bar{S}^t to the available treatment option set \mathcal{A} . The optimal DTR $\pi^* = (\pi_1^*, \pi_2^*, \dots, \pi_T^*)$ is defined as the DTR that achieves the maximum expected reward, i.e., $R_{\pi^*} = \mathbb{E}(\sum_{t=1}^T R_{\pi_t}^*) \leq \mathbb{E}(\sum_{t=1}^T R_{\pi_t}^t) = R_{\pi^*}$ for all π .

Q-learning To estimate the optimal DTR, Q-learning is widely used, particularly in the finite-horizon setting where decision stages are limited and predetermined [8, 9, 10, 11, 12]. Q-learning adopts a backward induction mechanism, starting its estimation at the last decision point and working its way back to the beginning.

We begin at the final stage T and posit models such as linear models or random forests to estimate the $Q_T(\bar{s}^T, \bar{a}^T)$ [10]. The observed cumulative reward $\sum_{t=1}^T R^t$ serves as the response variable, while \bar{s}^T and \bar{a}^{T-1} are used as covariates. Once the model is estimated, we identify the treatment $\hat{\pi}_T^*$ that maximizes the expected reward for a patient at decision point T , given their historical profile \bar{s}^T, \bar{a}^{T-1} . To find the

optimal treatment regime, we work backward, and treat the already-estimated Q-function value as the new response variable for previous decision points. Following similar steps, we estimate $\hat{\pi}^* = (\hat{\pi}_1^*, \dots, \hat{\pi}_T^*)$ as the overall optimal treatment regime. For a more comprehensive discussion, we refer readers to [7].

In a finite-horizon setting, Q-learning has gained tremendous popularity due to its ease of implementation and overall strong performance [5]. Nonetheless, its sensitivity to model misspecification presents a challenge. If the posited model for Q-functions is misspecified, performance may suffer significantly, as the bias would backpropagate to the very first stage. Various alternatives and variations have been proposed to address such limitations. For instance, Advantage Learning (A-learning, [13]) estimates the optimal DTR by modeling the difference in outcomes between two treatment options, making it more robust to model misspecification [14]. Robust Q-learning [15] introduces data-adaptive techniques for nuisance parameter estimation, tackling both residual confounding and efficiency loss.

There are also other notable advancements and extensions in Q-learning, such as statistical inference for Q-learning based on the asymptotic normality of estimators [16] and bootstrap methods [17]. [18] developed a Bayesian framework for finding the optimal DTR to accommodate prior knowledge and measure the uncertainty of the estimated DTR. [19] extended original Q-learning methods to survival outcomes, and [20] considers high-dimensional settings and variable selection in Q-learning. For an exhaustive overview, we direct readers to [5, 7].

Application: Treatment Regime in Perioperative Setting Numerous studies have explored the

application of Q-learning and its variants in clinical trial settings, aiming to find the optimal DTR from clinical trial data [10, 21, 22]. To illustrate this, we consider the treatment of early-stage malignant tumors, which could be surgically removed at this stage. The perioperative process typically begins with neoadjuvant therapy, designed to shrink the tumor and thereby enhance the chances of a successful surgery. Following the operation, adjuvant therapy is administered to prevent cancer recurrence.

We can model this as a Q-learning problem with two decision points. The state variables might include factors such as tumor stage, resection margin (R0/R1/R2), pathology, tumor imaging data, and patient health status. The first decision action, A^1 , is the choice of neoadjuvant treatment. The second decision action, A^2 , could be two-dimensional, incorporating both the choice of adjuvant treatment and its duration (number of cycles). Event-free survival (EFS) can serve as the final reward. By integrating the corresponding covariates into the Q-learning framework, we can estimate the optimal treatment sequence for both the neoadjuvant and adjuvant periods, tailored to the characteristics of individual patients.

Application: Lines of Therapies for Metastatic Cancers When treating metastatic cancer, the typical medical practice is to treat patients with the same drug until either the disease progresses or the patient becomes intolerant to the drug. The next-in-line treatment is then initiated. Identifying a personalized optimal treatment regime or sequence with the aim of maximizing a certain metric, such as overall survival, is of tremendous significance in healthcare. Given the multiple decision points associated with prescribing next-in-line treatments, such as drug choice and the

time point to switch to the next-line treatment, Q-learning becomes a natural fit for tackling this issue. For example, Zhao et al. [21] illustrated their reinforcement learning (RL) model in the context of lines of chemotherapy for metastatic non-small-cell lung cancer (NSCLC).

Fast-forward to the era of immuno-oncology, the narrative has evolved to focus on the personalized optimal sequence involving the PD-1/PD-L1 checkpoint inhibitors: Should they be given as monotherapy or in combination with other drugs, in what order, and to which patients? Most major checkpoint inhibitors, such as pembrolizumab and nivolumab, have been evaluated either as monotherapy or in combinations in different lines to treat patients with metastatic NSCLC. Therefore, existing data may already hold the answers to these questions. Applying Q-learning and other appropriate RL methods to the aggregated data could provide extremely valuable insights for improving the treatment of these patients.

2.2 Adaptive Clinical Trial Design

Adaptive Clinical Trials Q-learning is primarily concerned with estimating optimal DTR using pre-collected datasets. However, adaptive clinical trials require real-time, data-dependent decision making, such as selecting treatment arms based on historical data up to a certain cutoff point [23]. This real-time utilization of cumulative data is known as the “online setting”, which stands in contrast to the “offline setting” in which pre-collected datasets are used [24].

To formalize this problem in the context of adaptive clinical trial design, we consider a trial with N treatment arms. Each arm i is associated with an unknown probability distribution D_i , which describes the treatment outcomes (efficacy or toxicity) when

assigning that particular treatment to a patient. At each decision point t , a reward R^t is obtained from the corresponding distribution D_i when treatment arm i is selected. The objective is to determine the recommendation rule at each decision point based on the accumulated data. This rule aims to maximize the expected cumulative reward $\mathbb{E}[\sum_{t=1}^T R^t]$.

Such formulation transforms the adaptive design into a multi-armed bandit (MAB) problem [25, 26]. The major challenge in solving such a problem lies in balancing the trade-off between “exploration”, where less-understood arms are chosen to collect more data about their distributions, and “exploitation”, where arms with higher observed cumulative rewards are chosen to maximize the expected outcome [27]. Therefore, effective solutions to the MAB problem in the context of adaptive clinical trials must address this exploration-exploitation dilemma to achieve optimal patient outcomes.

Multiarmed Bandit Various methods have been developed to tackle the MAB, such as the ϵ -greedy algorithm [28], Thompson sampling [29], and Upper Confidence Bound [30], among others. The ϵ -greedy algorithm takes a straightforward approach to the exploration-exploitation dilemma. With probability $1 - \epsilon$, the algorithm selects the arm with the highest empirical mean reward observed so far, known as the “greedy” action. With probability ϵ , it randomly selects an arm, thereby exploring the action space. The parameter ϵ controls the trade-off between exploration and exploitation. A higher ϵ promotes more exploration at the cost of immediate reward, while a lower ϵ focuses more on exploitation. Meanwhile, Thompson sampling takes a Bayesian approach to the MAB problem. It maintains a probability distribution over the

expected reward for each arm, updating these distributions as more data are collected. At each round t , a sample is drawn from each arm's posterior distribution, and the arm with the highest sample is selected. The Upper Confidence Bound (UCB) algorithm selects the arm with the highest upper confidence bound on its expected reward. At each time step, it calculates the upper bound for each arm using both the estimated mean reward and its uncertainty. The arm with the highest calculated upper bound is then selected, aiming to minimize long-term regret.

ϵ -greedy is straightforward and computationally efficient but suffers from constant, often unnecessary, exploration due to its ϵ parameter [31]. Thompson sampling provides a more nuanced balance between exploration and exploitation by incorporating uncertainty through probabilistic models [29]. While this leads to better performance in complex environments, it may require greater computational resources, particularly for complex posterior distributions [32]. UCB has strong theoretical bounds on regret and is deterministic. However, it makes strong assumptions about the reward and can be less effective in non-stationary environments [28].

Several extensions to the original MAB algorithms have also been proposed to address real-world challenges, such as the analysis on sample complexity of MAB [33], MAB under dependent arms [34], MAB with safety constraints [35, 36], and MAB with multiple objectives [37]. To further incorporate the patient-specific information to the decision-making process, contextual bandit framework has been introduced with additional state variables [38, 39]. Such extension enables personalized treatment recommendations in adaptive clinical trials.

In the pharmaceutical setting, the MAB framework has been employed to study oncology dose-finding and response-adaptive randomization designs. We elaborate the first application and refer the readers to [26] for the latter.

Application: Oncology Dose-Finding One primary objective of phase I oncology dose-finding trials is to identify the maximum tolerated dose (MTD) of the drug candidate to inform the dose level(s) to be investigated in subsequent phases of development. They start treating one cohort of patients, usually of size 3, at the lowest provisional dose level. Upon observing the data of the cohort, a recommendation (escalation/stay/de-escalation) is rendered regarding the dose level at which the next cohort of patients should be treated according to a certain statistical design. This process is repeated until the total sample size is exhausted or certain pre-specified early-stopping rules are met.

Dose-finding has been an active area of statistical innovation. One important class of designs is the model-based designs [40, 41, 42]. These designs postulate a parametric form of the dose-toxicity relationship and utilize the cumulative data to make a dose recommendation. The endpoint in most cases is a binary indicator of the presence of dose-limiting toxicity (DLT) within a certain period (e.g., 28 days). Patient-level covariate information can be intuitively incorporated in the model-based designs [43].

Dose-finding trials are great candidates for applying the MAB framework due to their sequential and adaptive nature [44, 45]. Specifically, patients in the t^{th} cohort are assigned to dose D_t from the set of provisional doses $\{1, \dots, K\}$. The objective is to identify the dose level that is closest to the pre-specified target

toxicity rate θ . Mathematically, this can be expressed as $k^* = \arg \min_k |\theta - p_k|$, where p_k is the observed toxicity rate at dose k . We define the reward function R^t as $R^t = -|\theta - \hat{p}_{D_t}^t|$, where $\hat{p}_{D_t}^t$ is the estimated toxicity rate of the selected dose for cohort t . By employing suitable MAB algorithms, the optimal dose level can be effectively identified.

In recent years, the need for precision medicine is emerging more and more frequently with the development of new cancer treatments like T-cell engagers and cell therapies. Dosing of such therapies might need to be more personalized to avoid adverse events known to the mechanism of action, such as the cytokine release syndrome. The contextual bandit framework can be useful to incorporate patient-level information in this case [39].

2.3 Mobile Health for Enhanced Patient Management

Mobile Health (mHealth) Section 2.1 details statistical methods for estimating optimal DTRs with a finite number of decision points. However, with the recent advancement of sensor technologies and wearable devices, it has become possible to record personal health information over an extremely long period with the help of mHealth technologies [46]. Consequently, leveraging such data to formulate personalized treatment plans, addressing chronic diseases and various health issues across an infinite horizon with numerous decision points, has emerged as a prominent research area in recent years.

To date, mHealth has been used extensively in managing various health-related conditions including stress, depression, and other chronic diseases such as diabetes and cardiovascular diseases. It enhances patient monitoring and treatment for healthcare providers

[47]. In mHealth settings, the data follows a similar pattern as Section 2.1, which also consists of n i.i.d trajectories with T decision points, in the form of $\{S_i^1, A_i^1, R_i^1, S_i^2, \dots, S_i^T, A_i^T, R_i^T, S_i^{T+1}\}_{i=1}^n$. Compared with the finite horizon, several key differences should be noted.

First, the Markov property is assumed under the infinite horizon, meaning the next state and reward depend only on the current state and action, i.e., $P(S^{t+1} = s^{t+1} | \bar{S}^t = \bar{s}^t, \bar{A}^t = \bar{a}^t) = P(S^{t+1} = s^{t+1} | S^t = s^t, A^t = a^t)$. Following the Markov property, the policy π is a function of the current state only, mapping it to a distribution on the action space where $\pi(s) = P(A^t = a | S^t = s)$. Finally, a discount factor $\gamma \in [0, 1)$ is introduced to ensure that the sum of rewards $\sum_{k=0}^{\infty} \gamma^k R^{t+k}$ remains finite. A larger γ would place more weight on future rewards.

We generally model the whole process as a Markov decision process (MDP). An MDP is defined as a tuple $\langle \mathcal{S}, \mathcal{A}, \mathbf{P}, R, \gamma \rangle$, where \mathcal{S} is the state space, \mathcal{A} is the action space, $\mathbf{P} : \mathcal{S} \times \mathcal{A} \rightarrow \Delta(\mathcal{S})$ is the unknown transitional kernel, $R : \mathcal{S} \times \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ is a bounded reward function, and $\gamma \in [0, 1)$ is the discount factor. A policy π is a mapping from the state space to the action space $\pi : \mathcal{S} \rightarrow \mathcal{A}$. The goal is to find an optimal policy π^* that maximizes the expected discounted sum of rewards $\mathbb{E}_{\pi}[\sum_{k=1}^{\infty} \gamma^{k-1} R^{t+k} | S^t = s]$.

Reinforcement Learning (RL) When the number of decision points approaches infinity, the task of determining the optimal policy transforms into a reinforcement learning (RL) problem [48]. In RL literature, we define the value function and state-value function for a given policy π as $V_t^{\pi}(s) = \mathbb{E}[\sum_{k=0}^{\infty} \gamma^k R^{t+k} | S^t = s]$, and $Q_t^{\pi}(s, a) = \mathbb{E}[\sum_{k=0}^{\infty} \gamma^k R^{t+k} | S^t = s, A^t = a]$. The only difference between the V -function and Q -function

is whether we specify the action at time t .

Based on these definitions, we can treat both the V -function and Q -function as measures of how good a policy is for a patient in any given state. By finding a policy that maximizes these quantities, we essentially achieve the goal of constructing a personalized treatment plan. However, this is not trivial given the dynamics over a long period, which can be difficult to model. Hence, the *Bellman optimality equation* becomes an important tool.

We first define the optimal value function as $V^*(s) = \max_{\pi} V^{\pi}(s)$, and the optimal Q -function is similarly defined as $Q^*(s, a) = \max_{\pi} Q^{\pi}(s, a)$. These functions are interrelated through the equation $V^*(s) = \max_a Q^*(s, a)$. The policy π^* that maximizes these functions is referred to as the optimal policy, denoted by $V^{\pi^*}(s) = V^*(s)$ and $Q^{\pi^*}(s, a) = Q^*(s, a)$. Both $V^*(s)$ and $Q^*(s, a)$ are unique and must satisfy the corresponding *Bellman optimality equation* [49]:

$$V^*(s) = \max_a \mathbb{E}_{S^{t+1}|s,a} [R^t + \gamma V^*(S^{t+1}) | S^t = s, A^t = a],$$
$$Q^*(s, a) = \mathbb{E}_{S^{t+1}|s,a} [R^t + \gamma \max_{a'} Q^*(S^{t+1}, a') | S^t = s, A^t = a].$$

Thus, $V^*(s)$ and $Q^*(s, a)$ serve as the fixed points of their respective Bellman optimality equations, and π^* can be solved accordingly.

One major challenge in solving the Bellman optimality equation arises when the dataset is collected under a policy that diverges from the optimal policy π^* , while the Bellman optimality equation requires that actions be generated based on π^* to be valid [50]. Such distribution mismatch is the dominant case in mHealth setting and introduces both theoretical and computational challenges in finding the optimal policy.

To tackle these challenges, Greedy Gradient Q-learning (GGQ) [51] and V-learning [52] have been developed, formulating estimation equations based on the Q -function and V -function, respectively. GGQ has the

advantage of enabling the construction of confidence intervals for the mean outcome difference between the optimal policy and any alternative policies. However, its estimation equation contains a non-smooth max operator, making estimation difficult without large amounts of data [6]. Furthermore, GGQ consistently selects the best arm at each decision stage, often resulting in sub-optimal outcomes in complex dynamic environments [53]. In contrast, V-learning adopts a stochastic policy distribution and avoids the non-smooth max operator, leading to more stable optimization. The stochastic policy class also makes V-learning more robust in the face of unexpected situations [54].

While V-learning's stochastic policy class offers flexibility in action selection, it can degenerate into a uniform distribution in a large action space. To mitigate this, pT-learning was introduced, confining the support set to near-optimal actions at each decision point and allowing sparsity control through a tuning parameter [54]. Extending this, the Quasi-optimal Learning framework adapts the method to continuous action spaces, making it applicable to challenges such as optimal dose-finding over an infinite horizon [55].

Application: Glucose Management for Diabetes

Glucose management in diabetes is a key mHealth application. By continuously monitoring the glucose level, food intake, and physiological information, a series of just-in-time interventions, such as insulin injection, can be delivered to patients to improve long-term health outcomes [56]. An application example is the OhioT1DM study [57], featuring 12 Type 1 Diabetes patients with continuous glucose monitoring (CGM) data, self-reported activity logs such as meal intakes and sleep status, and insulin injection dosages and timing over eight weeks. Figure 1 provides a

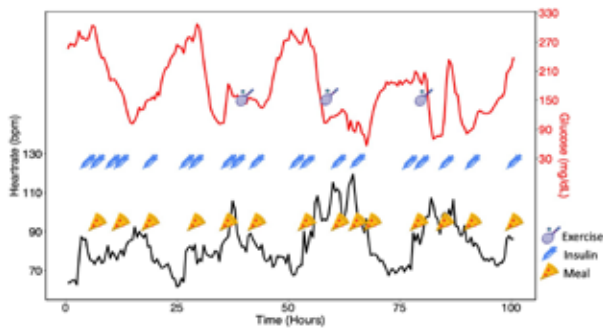


Figure 1: *OhioT1DM Data: A longitudinal observation of a patient*

snapshot of the fluctuation of glucose level, insulin injections, meals, exercise, and heart rate of a patient during a 100 hour time interval.

As glucose dynamics can vary significantly between individuals, clinicians aim to personalize insulin injection doses based on each patient’s health status [58]. Our objective is to develop a personalized treatment policy that optimally controls glucose levels for each individual.

We define the state variables as health status measurements for individual patients, and the action space refers to the insulin injection dose levels at each decision point. The glycemic index serves as the reward function, measuring the proximity of glucose levels to the normal range [59]. By applying methods like V-learning, pT-learning, and Quasi-optimal learning, we can determine an optimal policy for controlling each patient’s glucose levels. Implementation details are available in [52, 54, 55].

3 Discussion

We have introduced a wide range of methods and algorithms in personalized medicine. Under a finite horizon, methods like Q-learning and its variants, as well as MAB algorithms, have matured considerably in finding optimal DTRs and guiding the design of clinical trials. Nevertheless, these finite-horizon models have underlying assumptions that could be further relaxed

to enhance their applicability.

Confounding and causality are critical issues in policy learning. Current methods often assume a fully observable environment; however, the true policy may be influenced by unmeasured confounders such as genetic factors [24]. Incorporating recent advances in causal inference to address these unmeasured confounders [60, 61] has emerged as a promising research direction [62, 63, 64].

In offline settings like Q-learning, where data is pre-collected and no online interaction with the environment occurs, algorithms may suffer from inadequate coverage of state-action pairs. This can lead to imprecise estimations of value functions [50, 65]. Hence, the pessimism principle is advised to limit the learned policy from visiting poorly-covered states, ensuring safety and avoiding undesired behaviors [66, 67]. Balancing pessimism and policy optimality represents another interesting research avenue [68].

Furthermore, the performance of an estimated DTR is assessed by its value function. Thus, it’s essential to quantify uncertainties and conduct statistical inferences related to the value function. This challenge is closely tied to an emerging field of research known as off-policy evaluation (OPE), which aims to evaluate the value of a certain policy based on data generated from a different policy [69]. Notably, constructing confidence intervals for these value functions [70, 71] and evaluating the value disparity between a particular policy and the optimal one are also pivotal research questions [72].

Under infinite horizons, in addition to the challenges present in finite horizons, further issues emerge that require extensive investigation. For example, the Markov property is a fundamental assumption under an infi-

nite horizon. In mHealth settings, however, outcomes may be influenced by decisions made before the immediately preceding time point. Developing methods to test the validity of the Markov property [73], and to address violations in the data-generating process, is an important extension to existing frameworks.

Lastly, survival data is common in mHealth applications. Such data often includes treatment and covariate information that may be censored in follow-up stages, complicating policy learning. Although recent advancements in optimal policy estimation have been made within the survival data framework [19, 74, 75], adapting these approaches to an infinite horizon remains a challenge.

References

- [1] Sheela Kolluri et al. “Machine learning and artificial intelligence in pharmaceutical research and development: a review”. In: *The AAPS Journal* 24 (2022), pp. 1–10.
- [2] Qi Liu et al. “Landscape analysis of the application of artificial intelligence and machine learning in regulatory submissions for drug development from 2016 to 2021”. In: *Clinical pharmacology and therapeutics* 113.4 (2023), pp. 771–774.
- [3] Margaret A Hamburg and Francis S Collins. “The path to personalized medicine”. In: *New England Journal of Medicine* 363.4 (2010), pp. 301–304.
- [4] Bibhas Chakraborty and Erica E Moodie. “Statistical methods for dynamic treatment regimes”. In: *Springer-Verlag. doi 10.978-1* (2013), pp. 4–1.
- [5] Michael R Kosorok and Eric B Laber. “Precision medicine”. In: *Annual review of statistics and its application* 6 (2019), pp. 263–286.
- [6] Eric B Laber et al. “Dynamic treatment regimes: Technical challenges and applications”. In: *Electronic journal of statistics* 8.1 (2014), p. 1225.
- [7] Jesse Clifton and Eric Laber. “Q-learning: Theory and applications”. In: *Annual Review of Statistics and Its Application* 7 (2020), pp. 279–301.
- [8] SA Murphy. “A Generalization Error for Q-Learning.” In: *Journal of Machine Learning Research: JMLR* 6 (2005), pp. 1073–1097.
- [9] James M Robins. “Optimal structural nested models for optimal sequential decisions”. In: *Proceedings of the second seattle Symposium in Biostatistics*. Springer. 2004, pp. 189–326.
- [10] Yufan Zhao, Michael R Kosorok, and Donglin Zeng. “Reinforcement learning design for cancer clinical trials”. In: *Statistics in medicine* 28.26 (2009), pp. 3294–3315.
- [11] Inbal Nahum-Shani et al. “Q-learning: A data analysis method for constructing adaptive interventions.” In: *Psychological methods* 17.4 (2012), p. 478.
- [12] Michael R Kosorok and Erica EM Moodie. *Adaptive treatment strategies in practice: planning trials and analyzing data for personalized medicine*. SIAM, 2015.
- [13] Susan A Murphy. “Optimal dynamic treatment regimes”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 65.2 (2003), pp. 331–355.

- [14] Phillip J Schulte et al. “Q-and A-learning methods for estimating optimal dynamic treatment regimes”. In: *Statistical science: a review journal of the Institute of Mathematical Statistics* 29.4 (2014), p. 640.
- [15] Ashkan Ertefaie et al. “Robust Q-learning”. In: *Journal of the American Statistical Association* 116.533 (2021), pp. 368–381.
- [16] Rui Song et al. “Penalized q-learning for dynamic treatment regimens”. In: *Statistica Sinica* 25.3 (2015), p. 901.
- [17] Bibhas Chakraborty, Eric B Laber, and Ying-Qi Zhao. “Inference about the expected performance of a data-driven dynamic treatment regime”. In: *Clinical Trials* 11.4 (2014), pp. 408–417.
- [18] Thomas A Murray, Ying Yuan, and Peter F Thall. “A Bayesian machine learning approach for optimizing dynamic treatment regimes”. In: *Journal of the American Statistical Association* 113.523 (2018), pp. 1255–1267.
- [19] Hunyong Cho et al. “Multi-stage optimal dynamic treatment regimes for survival outcomes with dependent censoring”. In: *Biometrika* 110.2 (2023), pp. 395–410.
- [20] Wensheng Zhu, Donglin Zeng, and Rui Song. “Proper inference for value function in high-dimensional Q-learning for dynamic treatment regimes”. In: *Journal of the American Statistical Association* 114.527 (2019), pp. 1404–1417.
- [21] Yufan Zhao et al. “Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer”. In: *Biometrics* 67.4 (2011), pp. 1422–1433.
- [22] Gregory Yauney and Pratik Shah. “Reinforcement learning with action-derived rewards for chemotherapy and clinical trial dosing regimen selection”. In: *Machine Learning for Healthcare Conference*. PMLR. 2018, pp. 161–226.
- [23] William R Zame et al. “Machine learning for clinical trials in the era of COVID-19”. In: *Statistics in biopharmaceutical research* 12.4 (2020), pp. 506–517.
- [24] Antonio Coronato et al. “Reinforcement learning for intelligent healthcare applications: A survey”. In: *Artificial Intelligence in Medicine* 109 (2020), p. 101964.
- [25] William H Press. “Bandit solutions provide unified ethical models for randomized clinical trials and comparative effectiveness research”. In: *Proceedings of the National Academy of Sciences* 106.52 (2009), pp. 22387–22392.
- [26] Sofia S Villar, Jack Bowden, and James Wason. “Multi-armed bandit models for the optimal design of clinical trials: benefits and challenges”. In: *Statistical science: a review journal of the Institute of Mathematical Statistics* 30.2 (2015), p. 199.
- [27] Jean-Yves Audibert, Rémi Munos, and Csaba Szepesvári. “Exploration–exploitation tradeoff using variance estimates in multi-armed bandits”. In: *Theoretical Computer Science* 410.19 (2009), pp. 1876–1902.
- [28] Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.
- [29] Shipra Agrawal and Navin Goyal. “Analysis of thompson sampling for the multi-armed bandit problem”. In: *Conference on learning theory*.

- JMLR Workshop and Conference Proceedings. 2012, pp. 39–1.
- [30] Aurélien Garivier and Eric Moulines. “On upper-confidence bound policies for switching bandit problems”. In: *International Conference on Algorithmic Learning Theory*. Springer. 2011, pp. 174–188.
- [31] John White. *Bandit algorithms for website optimization*. " O'Reilly Media, Inc.", 2013.
- [32] Izzatul Umami and Lailia Rahmawati. “Comparing Epsilon greedy and Thompson sampling model for multi-armed bandit algorithm on marketing dataset”. In: *Journal of Applied Data Sciences 2.2* (2021).
- [33] Shie Mannor and John N Tsitsiklis. “The sample complexity of exploration in the multi-armed bandit problem”. In: *Journal of Machine Learning Research* 5.Jun (2004), pp. 623–648.
- [34] Sandeep Pandey, Deepayan Chakrabarti, and Deepak Agarwal. “Multi-armed bandit problems with dependent arms”. In: *Proceedings of the 24th international conference on Machine learning*. 2007, pp. 721–728.
- [35] Samuel Daulton et al. “Thompson sampling for contextual bandit problems with auxiliary safety constraints”. In: *arXiv preprint arXiv:1911.00638* (2019).
- [36] Cong Shen et al. “Learning for dose allocation in adaptive clinical trials with safety constraints”. In: *International Conference on Machine Learning*. PMLR. 2020, pp. 8730–8740.
- [37] Saba Q Yahyaa and Bernard Manderick. “Thompson Sampling for Multi-Objective Multi-Armed Bandits Problem.” In: *ESANN*. 2015.
- [38] Ambuj Tewari and Susan A Murphy. “From ads to interventions: Contextual bandits in mobile health”. In: *Mobile health: sensors, analytic methods, and applications* (2017), pp. 495–517.
- [39] Yogatheesan Varatharajah and Brent Berry. “A contextual-bandit-based approach for informed decision-making in clinical trials”. In: *Life* 12.8 (2022), p. 1277.
- [40] John O’Quigley, Margaret Pepe, and Lloyd Fisher. “Continual reassessment method: a practical design for phase 1 clinical trials in cancer”. In: *Biometrics* (1990), pp. 33–48.
- [41] Beat Neuenschwander, Michael Branson, and Thomas Gsponer. “Critical aspects of the Bayesian approach to phase I cancer trials”. In: *Statistics in medicine* 27.13 (2008), pp. 2420–2439.
- [42] Hongtao Zhang, Alan Y Chiang, and Jixian Wang. “Improving the performance of Bayesian logistic regression model with overdose control in oncology dose-finding studies”. In: *Statistics in Medicine* 41.27 (2022), pp. 5463–5483.
- [43] Beat Neuenschwander et al. “A Bayesian industry approach to phase I combination trials in oncology”. In: *Statistical methods in drug combination studies 2015* (2015), pp. 95–135.
- [44] Maryam Aziz, Emilie Kaufmann, and Marie-Karelle Riviere. “On multi-armed bandit designs for dose-finding clinical trials”. In: *The Journal of Machine Learning Research* 22.1 (2021), pp. 686–723.
- [45] Lan Jin, Guodong Pang, and Demissie Alemayehu. “Multiarmed Bandit Designs for Phase I Dose-Finding Clinical Trials With Multiple Tox-

- icity Types". In: *Statistics in Biopharmaceutical Research* 15.1 (2023), pp. 164–177.
- [46] Bruno MC Silva et al. "Mobile-health: A review of current state in 2015". In: *Journal of biomedical informatics* 56 (2015), pp. 265–272.
- [47] James M Rehg, Susan A Murphy, and Santosh Kumar. "Mobile health". In: *Cham: Springer International Publishing* (2017).
- [48] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.
- [49] Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.
- [50] Scott Fujimoto, David Meger, and Doina Precup. "Off-policy deep reinforcement learning without exploration". In: *International conference on machine learning*. PMLR. 2019, pp. 2052–2062.
- [51] Ashkan Ertefaie and Robert L Strawderman. "Constructing dynamic treatment regimes over indefinite time horizons". In: *Biometrika* 105.4 (2018), pp. 963–977.
- [52] Daniel J Lockett et al. "Estimating dynamic treatment regimes in mobile health using v-learning". In: *Journal of the American Statistical Association* (2019).
- [53] Christoph Dann, Gerhard Neumann, Jan Peters, et al. "Policy evaluation with temporal differences: A survey and comparison". In: *Journal of Machine Learning Research* 15 (2014), pp. 809–883.
- [54] Wenzhuo Zhou, Ruoqing Zhu, and Annie Qu. "Estimating optimal infinite horizon dynamic treatment regimes via pt-learning". In: *Journal of the American Statistical Association* (2022), pp. 1–14.
- [55] Yuhan Li, Wenzhuo Zhou, and Ruoqing Zhu. "Quasi-optimal Reinforcement Learning with Continuous Actions". In: *The Eleventh International Conference on Learning Representations*. 2022.
- [56] Shruti Muralidharan et al. "Mobile health technology in the prevention and management of type 2 diabetes". In: *Indian journal of endocrinology and metabolism* 21.2 (2017), p. 334.
- [57] Cindy Marling and Razvan Bunescu. "The OhioT1DM dataset for blood glucose level prediction: Update 2020". In: *CEUR workshop proceedings*. Vol. 2675. NIH Public Access. 2020, p. 71.
- [58] Jiansong Bao et al. "Improving the estimation of mealtime insulin dose in adults with type 1 diabetes: the Normal Insulin Demand for Dose Adjustment (NIDDA) study". In: *Diabetes Care* 34.10 (2011), pp. 2146–2151.
- [59] David Rodbard. "Interpretation of continuous glucose monitoring data: glycemic variability and quality of glycemic control". In: *Diabetes technology & therapeutics* 11.S1 (2009), S–55.
- [60] Wang Miao, Xu Shi, and Eric Tchetgen Tchetgen. "A confounding bridge approach for double negative control inference on causal effects". In: *arXiv preprint arXiv:1808.04945* (2018).
- [61] Yifan Cui et al. "Semiparametric proximal causal inference". In: *Journal of the American Statistical Association* (2023), pp. 1–12.

- [62] Nathan Kallus and Angela Zhou. “Confounding-robust policy improvement”. In: *Advances in neural information processing systems* 31 (2018).
- [63] Jiayi Wang, Zhengling Qi, and Chengchun Shi. “Blessing from experts: Super reinforcement learning in confounded environments”. In: *arXiv preprint arXiv:2209.15448* (2022).
- [64] Chengchun Shi et al. “A minimax learning approach to off-policy evaluation in confounded partially observable markov decision processes”. In: *International Conference on Machine Learning*. PMLR, 2022, pp. 20057–20094.
- [65] Tengyang Xie et al. “Bellman-consistent pessimism for offline reinforcement learning”. In: *Advances in neural information processing systems* 34 (2021), pp. 6683–6694.
- [66] Ying Jin, Zhuoran Yang, and Zhaoran Wang. “Is pessimism provably efficient for offline rl?” In: *International Conference on Machine Learning*. PMLR, 2021, pp. 5084–5096.
- [67] Masatoshi Uehara and Wen Sun. “Pessimistic model-based offline reinforcement learning under partial coverage”. In: *arXiv preprint arXiv:2107.06226* (2021).
- [68] Kamyar Ghasemipour, Shixiang Shane Gu, and Ofir Nachum. “Why so pessimistic? estimating uncertainties for offline rl through ensembles, and why their independence matters”. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 18267–18281.
- [69] Masatoshi Uehara, Chengchun Shi, and Nathan Kallus. “A review of off-policy evaluation in reinforcement learning”. In: *arXiv preprint*
- [70] Philip Thomas, Georgios Theocharous, and Mohammad Ghavamzadeh. “High-confidence off-policy evaluation”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 29. 1. 2015.
- [71] Yihao Feng et al. “Non-asymptotic confidence intervals of off-policy evaluation: Primal and dual bounds”. In: *arXiv preprint arXiv:2103.05741* (2021).
- [72] Chengchun Shi et al. “Statistical inference of the value function for reinforcement learning in infinite-horizon settings”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84.3 (2022), pp. 765–793.
- [73] Chengchun Shi et al. “Does the markov decision process fit the data: Testing for the markov property in sequential decision making”. In: *International Conference on Machine Learning*. PMLR, 2020, pp. 8807–8817.
- [74] Ying-Qi Zhao et al. “Constructing dynamic treatment regimes with shared parameters for censored data”. In: *Statistics in medicine* 39.9 (2020), pp. 1250–1263.
- [75] Fei Xue et al. “Multicategory angle-based learning for estimating optimal dynamic treatment regimes with censored data”. In: *Journal of the American Statistical Association* 117.539 (2022), pp. 1438–1451. ■

BIG DATA, DIGITAL AND ARTIFICIAL INTELLIGENCE (AI) IN A PATIENT-CENTRIC ERA: PERSPECTIVES OF THE MOST VALUABLE DATA & INSIGHTS INITIATIVE AWARD WINNER

Kelly H. Zou, PhD, PStat®, ASA Fellow
Head, Global Medical Analytics and Real World Evidence,
Viatris Inc.
Email: Kelly.Zou@viatris.com

1. A New Era of Statistics and Data Science in Healthcare

Nowadays, the healthcare industry has seen a rapidly emerging fields through statistical and technological advancements. The key ingredient is data science, where big data, real-world data (RWD), digital innovation, and artificial intelligence (AI) are key elements.

“Artificial intelligence (AI) and machine learning are critical to the U.S. Department of Health and Human Services (HHS) in accomplishing our mission to enhance the health and well-being of all Americans.” [1]

Specifically, AI “enables computer systems to perform tasks normally requiring human intelligence.” Within the scope of AI, machine learning (ML) is “a type of artificial intelligence, gives computers the ability to learn without being programmed by humans.” Within the scope of ML, “deep learning (DL) systems learn from large amounts of data to subsequently recognize and classify related, but previously unobserved, data.” [2]

On the other hand, the European Commission proposed the first EU regulatory framework for AI. The proposed regulatory framework on AI included a number of specific objectives: “ensure that AI systems placed on the Union market and used are safe and respect existing law on fundamental rights and Union values; ensure legal certainty to facilitate investment and innovation in AI; enhance governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems; facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation.” [3]

Recently, in the latest amendments to the Commission’s proposal, the European Parliament would like to see that “AI systems are overseen by people, are safe, transparent, traceable, non-discriminatory, and environmentally friendly.” Given the broad spectrums of AI applications, the definition of AI can be “designed to be technology-neutral, so that it can apply to the AI

systems of today and tomorrow.” [4] “The EU wants to regulate AI to ensure better conditions for the development and use of this innovative technology. AI can create many benefits in improved healthcare and other aspects in terms of social determinants of health [5, 6]

2. Big Data and Real-World Data in Healthcare

To conduct sophisticated analyses or apply complex algorithms, the input data, ideally of high quality, are critical. As scientific discoveries and methodologies continue to advance, RWD and their companion technologies, such as digital and AI, which often employ statistical methods and data science tools, offer powerful ways for pharmaceutical industry to generate evidence.

Since the 21st Century Cures Act in 2016, the U.S. Food and Drug Administration has defined real-world evidence (RWE) as “evidence generated from real-world data (RWD) outside randomized controlled trials (RCTs).” Subsequently, “real-world evidence (RWE) is the clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of RWD.” [7]

The European Commission, on the other hand, is establishing an overarching and comprehensive European Health Data Space (EHDS) for the purpose of regulation. The EHDS “is a health specific ecosystem comprised of rules, common standards and practices, infrastructures and a governance framework...” Its aims include: “empowering individuals through increased digital access to and control of their electronic personal health data, at national level and EU-wide, and support to their free movement, as well as fostering a genuine single market for electronic health record systems, relevant medical devices and high risk AI systems; providing a consistent, trustworthy and efficient set-up for the use of health data for research, innovation, policy-making and regulatory activities.” [8]

Central to and critical for generating and harnessing big data and RWD, as well as for adopting and deploying AI is digital medicine, which “describes a field, concerned with the use of technologies as tools for measurement, and intervention in the service of human health.” Below can be why digital medicines can be ripe

Table 1. List of topics in Zou et al. (2022) . [10]

Chapter	Topic
Preface	Real-world evidence and digital innovation to combat noncommunicable diseases
1	Real World Evidence Generation
2	Applications of RWE for Regulatory Uses
3	Ethics & Bioethics
3	Real- World Data, Big Data and Artificial Intelligence: Recent Development and Emerging Trends in the European Union
4	Patient centricity and Precision Medicine
5	Health Information Technology
6	Digital Health Technologies and Innovations
7	Economic Analysis and Outcome Assessment
8	Partnerships and Collaborations
9	Global Perspective: China Big Data Collaboration to Improve Patient Care
10	The Future of Patient-Centric Data-Driven Healthcare

for the adoption of AI, since “digital medicine products are driven by high-quality hardware and software that support the practice of medicine broadly, including treatment, recovery, disease prevention, and health promotion for individuals and across populations.” [9]

3. An Award-Winning Book

We have recently published a book "Real-World Evidence in a Patient-Centric Digital Era", that provided methods, perspectives, examples, and insights on the innovative application of RWE to meet patient needs and improve healthcare, with a focus on the pharmaceutical industry. [10, 11] The authors have presented an overview of key analytical issues and best practices. Special attention is paid to the development, methodologies, and other salient features of the statistical and data science techniques that are customarily used to generate real-world evidence. It provides a review of key topics and emerging trends in cutting-edge data science and health innovation. Several key highlights and special features include the following: (1) Provided an overview of statistical and analytic methodologies in RWE to generate insights on healthcare, with a special focus on the pharmaceutical industry. (2) Examined timely topics of high relevance to industry such as bioethical considerations, regulatory

standards, and compliance requirements. (3) Highlighted emerging and current trends and provides guidelines for best practices. (4) Illustrated methods through examples and use-case studies to demonstrate impact. (5) Provided guidance on software choices and digital applications for successful analytics. [10]

Table 1 lists the topics in the entire book. Through extensive multi-year efforts, this cutting-edge project is a timely development for the biopharmaceutical medical researchers, health technology innovators, data scientists, epidemiologists, population health analysts, health economists, outcomes researchers, policymakers, and analysts in the healthcare industry.

This book was jointly written by several authors and co-edited by Drs. Kelly H. Zou, PhD, PStat®, FASA, Lobna A. Salem, MD, MBA, and Amrit Ray, MD, MBA. [10] The authors have won Reuters Events Pharma USA’s Most Valuable Data & Insights Initiative Award. Announcements by Reuters Events Pharma USA also included being the finalists of two additional awards, including digital innovation and partnerships in emerging markets. [11]

4. AI Algorithms and Applications in Pharma

In three recent publications, we have explored commonly used algorithms in the data science community,

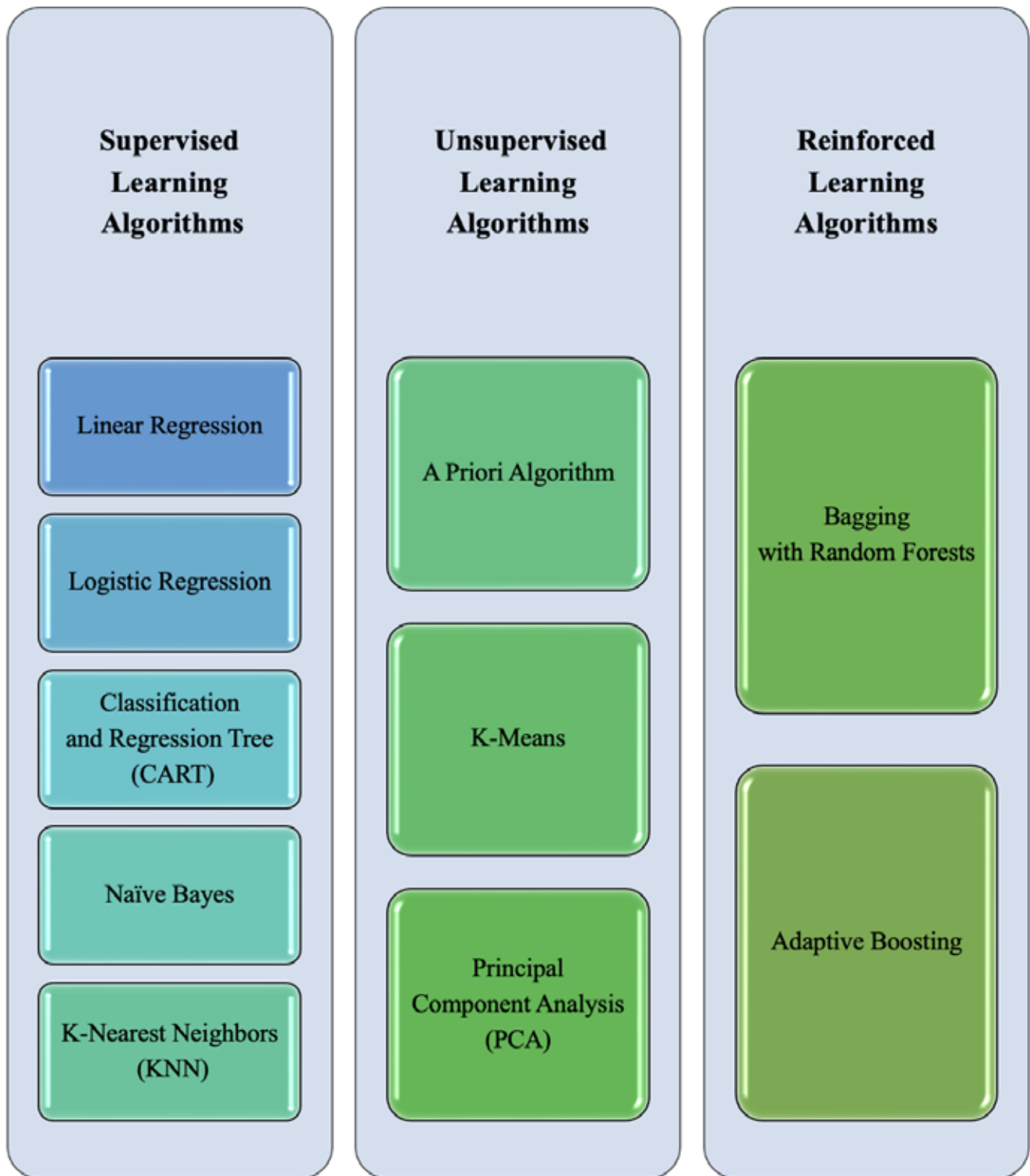


Figure 1. Commonly used ML algorithms. [12]

as well as the type of AI examples in the pharm industry, as well as best practices in digital innovations through “bring your own device” (BYOD) designs. [12-14]

5. Benefits and Impact via Patient-Centricity

In terms of benefits through innovative patient-centric statistical and data science methodologies and algorithms,

practitioners may evaluate medication adherence, patient preference, patient voice, patient journey, and precision medicine, and patient engagements to better understand the complex set of predictors and behaviors of patients within the healthcare system. To adequately explore and make inferences on these outcomes, a high-quality data

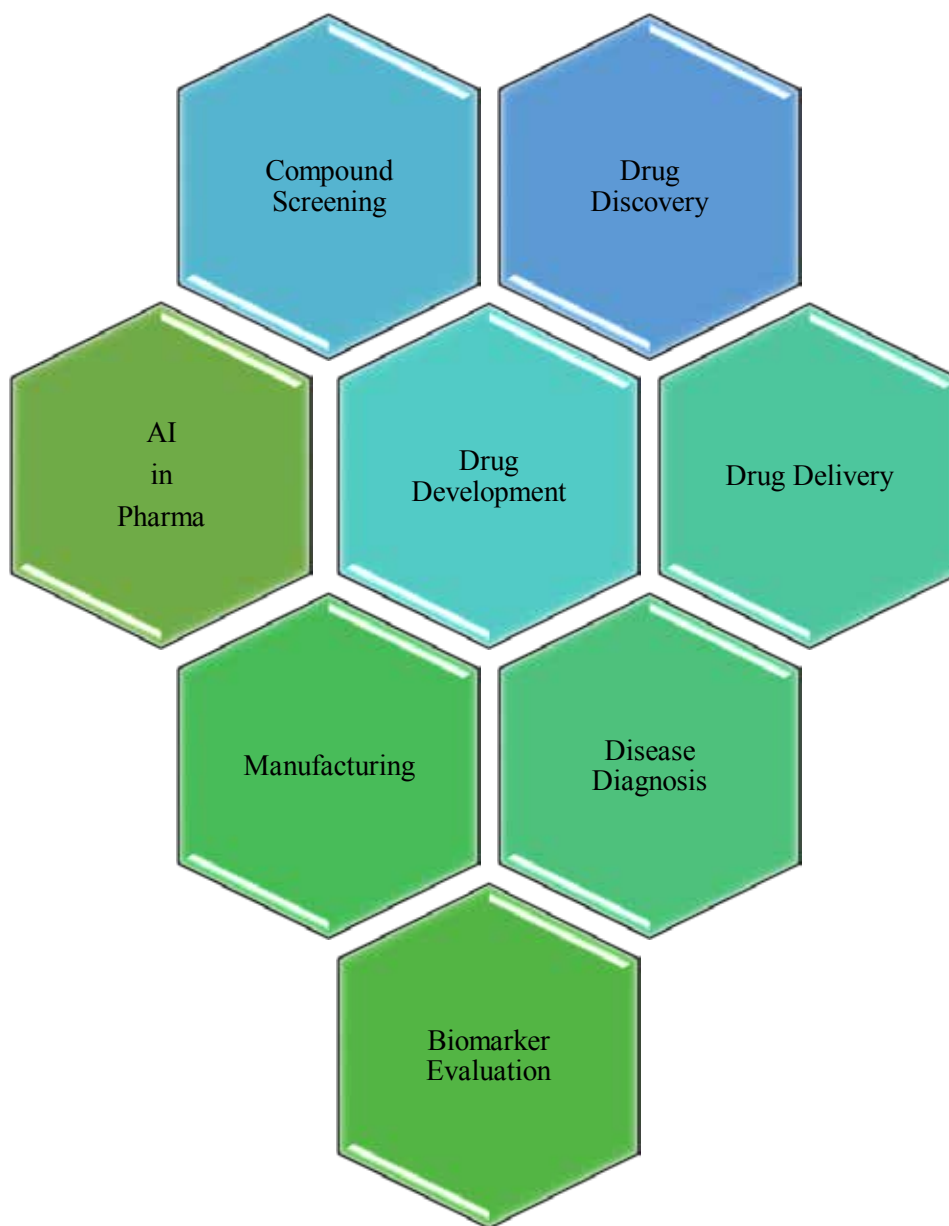


Figure 2. Commonly used ML algorithms in the pharmaceutical industry. [13]

framework is necessary, especially for regulatory use, given the fact that RWD is typically from routine clinical practice with varying degrees of “messiness” within such data. [15] For example, adherence to medication is one of the most complex behaviors of patients. Strategies for measuring and improving adherence require innovative and sophisticated “beyond-pill” solutions, which new technologies may help. [16]

Finally, RCTs alone may not adequately address the complex intersection of many diseases and comorbid conditions, which are patient centric and require us to find alternate ways of getting evidence to support such gaps.

[17] Big data, RWD, digital and AI can support patients for generalizability across the spectrum of various characteristics and comorbid conditions, by taking into account the tradeoff between potential benefits and risks, as well as the data privacy rules such as Health Insurance Portability and Accountability Act of 1996 (HIPAA) in the United States, [18] General Data Protection Regulation (GDPR) in the European Union, [19] cross-border data transfers, [20] and an AI bill of rights. [21] Most recently, The U.S. White House’s executive order on AI reshapes the uses of AI in healthcare, ranging from drug/therapy research to hospital/clinical care. AI will play increasing roles in the

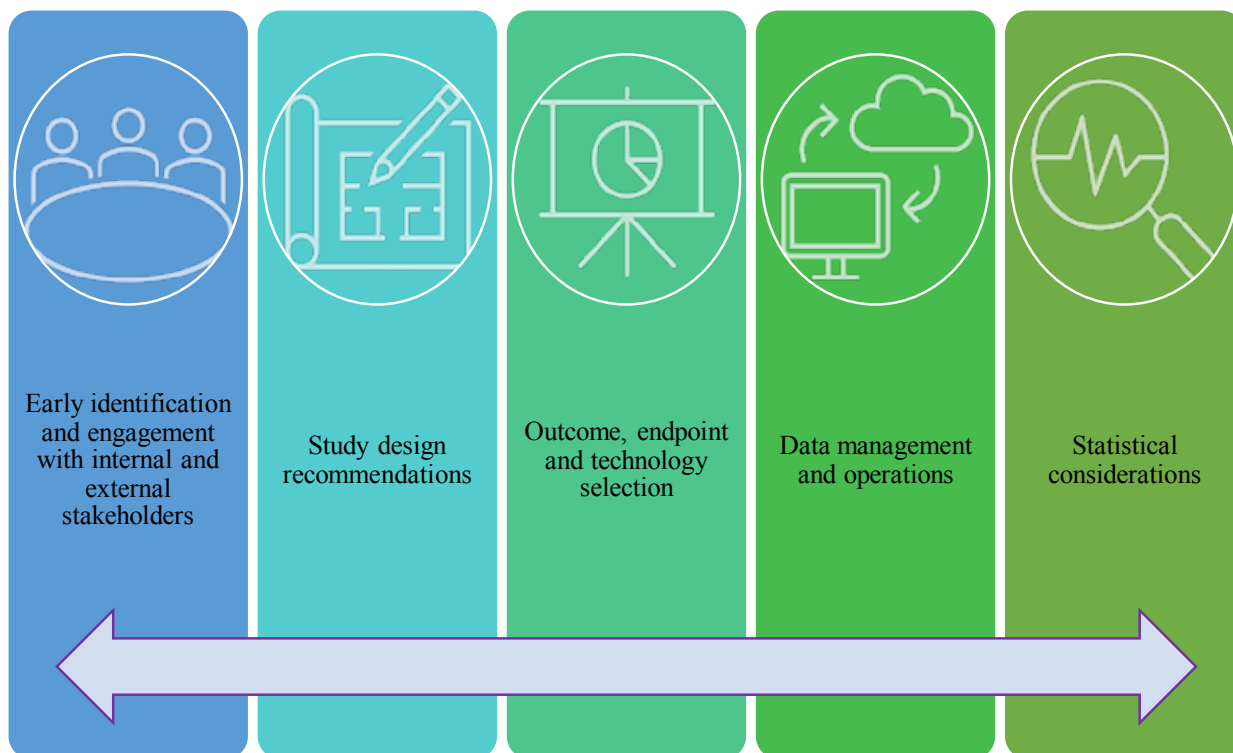


Figure 3. Best practices in digital applications via BYOD designs. [14]

biopharma industry, along with issues associated with ethics, transparency and trustworthiness. [22]

References

1. U.S. Department of Health and Human Services. 2023. Artificial Intelligence (AI) at HHS. <https://www.hhs.gov/about/agencies/asa/ocio/ai/index.html> (Accessed on October 18, 2023)
2. Congressional Research Service. 2021. Artificial Intelligence: Background, Selected Issues, and Policy Considerations. <https://crsreports.congress.gov/product/pdf/R/R46795> (Accessed on October 18, 2023).
3. EUR-Lex. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS. 2021. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206> (Accessed on October 18, 2023).
4. European Parliament (News). AI Act: a step closer to the first rules on Artificial Intelligence. 2023. <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence> (Accessed on October 18, 2023).
5. European Parliament (News). EU AI Act: first regulation on artificial intelligence. 2023. <https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence> (Accessed on October 18, 2023).
6. European Parliament (News). Artificial intelligence: threats and opportunities . 2023. <https://www.europarl.europa.eu/news/en/headlines/society/20200918STO87404/artificial-intelligence-threats-and-opportunities> (Accessed on October 18, 2023).
7. U.S. Food and Drug Administration. 2022. Real-World Evidence. <https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence> (Accessed on October 18, 2023).
8. European Commission. 2022. European Health Data Space. 2023. https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en (Accessed on October 18, 2023).

9. Digital Medicine Society. Defining Digital Medicine. 2023. <https://dimesociety.org/about-us/defining-digital-medicine> (Accessed on October 18, 2023).
10. Zou KH, Salem LA, Ray A (Eds). Real-World Evidence in a Patient-Centric Digital Era. 2022. Chapman and Hall/CRC. Taylor and Francis Group. New York, NY, USA. <https://www.routledge.com/Real-World-Evidence-in-a-Patient-Centric-Digital-Era/Zou-Salem-Ray/p/book/9780367861810> and <https://www.taylorfrancis.com/books/edit/10.1201/9781003017523/real-world-evidence-patient-centric-digital-era-kelly-zou-lobna-salem-amrit-ray> (Accessed on October 18, 2023).
11. American Statistical Association. Co-editors publish award-winning book. 2023. <https://magazine.amstat.org/blog/2023/02/01/asa-members-award-winning-book> (Accessed on October 18, 2023).
12. Zou KH, Li JZ, Imperato J, et al. Harnessing Real-World Data for Regulatory Use and Applying Innovative Applications. *J Multidiscip Healthc.* 2020;13:671-679.
13. Zou KH, Li JZ. Enhanced Patient-Centricity: How the Biopharmaceutical Industry Is Optimizing Patient Care through AI/ML/DL. *Healthcare (Basel).* 2022;10(10):1997.
14. Demanuele C, Lokker C, Jhaveri K, Georgiev P, Sezgin E, Geoghegan C, Zou KH, Izmailova E, McCarthy M. Considerations for Conducting Bring Your Own "Device" (BYOD) Clinical Studies. *Digit Biomark.* 2022;6(2):47-60.
15. European Medicines Agency. Multi-stakeholder workshop on Real World Data (RWD) quality and Real World Evidence (RWE) use. 2023. <https://www.ema.europa.eu/en/events/multi-stakeholder-workshop-real-world-data-rwd-quality-real-world-evidence-rwe-use> (Accessed on October 18, 2023).
16. Hassan TA, Sáenz JE, Ducinskiene D, Cook JP, Imperato JS, Zou KH. New Strategies to Improve Patient Adherence to Medications for Noncommunicable Diseases During and After the COVID-19 Era Identified via a Literature Review. *J Multidiscip Healthc.* 2021;14:2453-2465.
17. U.S. Food and Drug Administration. FDA Releases Two Discussion Papers to Spur Conversation about Artificial Intelligence and Machine Learning in Drug Development and Manufacturing. 2023. <https://www.fda.gov/news-events/fda-voices/fda-releases-two-discussion-papers-spur-conversation-about-artificial-intelligence-and-machine> (Accessed on October 18, 2023).
18. U.S. Department of Health and Human Services. Health Information Privacy. 2023. <https://www.hhs.gov/hipaa/index.html> (Accessed on October 18, 2023).
19. GDPR.EU. Complete guide to GDPR compliance. 2023. <https://gdpr.eu> (Accessed on October 18, 2023).
20. U.S. Department of Commerce. Data privacy framework program launches new website enabling U.S. companies to participate in cross-border data transfers. 2023. <https://www.commerce.gov/news/press-releases/2023/07/data-privacy-framework-program-launches-new-website-enabling-us> (Accessed on October 18, 2023).
21. The White House. Blueprint for an AI Bill of Rights. 2023. <https://www.whitehouse.gov/ostp/ai-bill-of-rights> (Accessed on October 18, 2023).
22. The White House. Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. 2023. <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence> (Accessed on November 1, 2023).

Disclosure

Dr. Zou is an employee of Viatrix. The views expressed are her own and do not necessarily represent her employer's. Viatrix did not provide any editorial support for this article.

Dr. Zou would like to thank the award agency Reuters Events, especially Liz Swift, Awards Director; Pharma, as well as the publisher Rob Calver, Senior Publisher – Mathematics, Statistics, and Physics, Chapman & Hall/CRC, Taylor & Francis Group. All authors and editors of the published book have declined any royalties associated with this project to be impartial. ■

USING BAYESIAN NETWORK TO IDENTIFY STATISTICALLY CAUSAL CLINICAL BIOMARKERS IN HIGH DIMENSIONAL DATASETS

John Kang¹, Wei Wei¹ | Merck & Co., Inc.

Abstract

Biomarkers are critical in predicting patient's clinical response, and therefore are an integral part of rational development of drug and medical devices. While the advancement of technology creates unprecedented opportunity to perform biomarker discovery using novel data modalities such as single cell RNA seq, proteomics, etc., innovative statistical methods are often also required to analyze these complex datasets. In this paper, we present a graphical model (Bayesian network) based causal inference framework for high dimensional clinical biomarker discovery. In addition, under the Bayesian network analysis paradigm, we also present the metric CCR (Conditional Causal Relative Risk) and the inference procedure to determine its statistical significance, which can be easily implemented using standard software packages such as R.

Introduction

Biomarkers are critical to the rational development of drug and medical devices [1]. A predictive biomarker is defined as a molecular, histologic, radiographic, or physiologic characteristics that is predictive of the propensity of an individual or group of individuals to experience a favorable or unfavorable effect from the exposure to a medical product or environmental agent; whereas a prognostic biomarker is used to identify the likelihood of a clinical event, disease recurrence, or disease progression in patients with a disease or medical condition of interest [2].

Predictive biomarkers can play an important role in the devising enrichment strategies in the design and conduct of clinical trials, particularly in the pre-registration stage of drug development. Prioritizing enrollment of participants with desirable baseline levels of a predictive biomarker enables detection of a clearer treatment effect by enriching subgroups of subjects that are more likely to benefit from the treatment [3].

Prognostic biomarkers are regularly used to set trial entry and exclusion criteria to identify higher-risk populations. Incorporation of prognostic biomarkers into clinical trial design could increase the statistical power of a trial by increasing the number of events rather than the sample size.

While the importance of a biomarker strategy during clinical development is widely recognized, robust identification of candidate biomarkers remains a challenging task. First, the complexity of human disease such as cancer is reflected in the diverse molecular and phenotypic fingerprints of individual patients [4]. As a result, a large number of possible covariates, such as patient demographic information and laboratory assay results (e.g. high dimensional genomic data) are often required to be explored for the discovery of novel clinical biomarkers. The high dimensionality of feature space, compounded by the small sample size (N) for the available translational/clinical trial datasets used by the pharmaceutical industry to discover clinical biomarkers, poses tremendous challenges in the proper control of statistical type I and type II errors.

To overcome the analytical challenges introduced by the small N of the clinical biomarker discovery cohort, observational datasets such as Real-world evidence (RWE), have gained much momentum in the scientific community for biomarker identification. However, observational studies are often subject to confounding factors if the expression of biomarkers is significantly associated (due to random chance or due to underlying biology) with demographic variables that are also predictive of clinical response. In the presence of confounding factors, statistical methods such as propensity score matching [5] are required to select the subset of candidate biomarkers that are directly influencing the response of interest independent of the confounding variables, and hence more likely to be replicated in the future validation exercise.

Traditional confounder adjustment methods such as propensity score are widely used for confirmatory biomarker studies, where there is only one single well defined treatment/exposure variable (i.e. the biomarker of interest). However, in high dimensional biomarker discovery studies, where every feature should be treated as a potential exposure variable and needs to be adjusted for the associated confounders, methods such as propensity score adjustment often can not be directly applied. Alternatively, graphical model based de-confounding strategies such as Bayesian networks can be used as the basis for an alternative set of non-experimental, statistical techniques for causal inference in the high dimensional biomarker discovery setting.

First formalized and developed by [6], Bayesian networks have now become widely applied in the social and natural science applications. By definition, a Bayesian network is a representation of a joint probability distribution, which consists of two components: E , which is a directed acyclic graph (DAG) whose vertices correspond to the random variables X_1, \dots, X_n ; and θ , which describes a conditional distribution for each variable, given its parents in E . Together, these two components specify a unique distribution on X_1, \dots, X_n . The graph E represents conditional independence assumptions that allow the joint distribution to be factorized, economizing the number of parameters. The graph E encodes the Markov assumption, which states that each variable X_i is independent of its non-descendants, given its parents in E [7].

To fully specify a joint distribution, we also need to specify each of the conditional probabilities in product form. If each variable X and its parents U_1, \dots, U_K are treated as continuous variables, a natural choice for multivariate continuous distributions is Gaussian distributions, which can be represented in a Bayesian network by using linear Gaussian conditional densities. Alternatively, features can also be encoded into ordinal variables, in which case, a multinomial distribution will be used. Although Bayesian networks are generally considered to encapsulate conditional dependencies (and independencies) between variables rather than necessarily implying causal relationships between them, a causal interpretation can be ascribed to a Bayesian network under certain assumptions [7-11].

In this paper, we apply a Bayesian network based approach to perform de-novo biomarker discovery using the Accelerating Medicines Partnership-Alzheimer's Disease [12] proteomics dataset with the aim of identifying Alzheimer's Disease associated prognostic protein biomarkers in the presence of confounders.

Data and Methods

AMP-AD Proteomics Dataset

516 dorsolateral prefrontal cortex (DLPFC) tissues from control, asymptomatic AD (AsymAD), and AD brains from the Religious Orders Study and Memory and Aging Project (ROSMAP) [13-15] and the Banner Sun Health Research Institute [16] by TMT-MS-based quantitative proteomics [17]. After data processing and outlier removal, a total of 488 subjects (n=182 AD, 200 AsymAD, and 106 controls) and 7863 proteins are used to construct a Bayesian network for protein biomarker discovery in predicting

AD severity score, CERAD (Consortium to Establish a Registry for Alzheimer's) score. [18]

The set of patient-level confounding variables included in this study include: APOE genotype [19] (categorical), gender (binary), postmortem interval (PMI, continuous), study cohort (binary), and age at death (continuous).

Network Construction

The nodes fed into the Bayesian network contain the following variables: patient level confounding variables (APOE genotype, gender, PMI, study cohort, and age at death), the response variable to be predicted (CERAD score), and the expression values for 7863 proteins. The conditional likelihood of the variables given their parents is represented in a Bayesian network by using linear Gaussian conditional densities. To avoid biologically uninterpretable directional edges in the network, we ban the following edges from appearing in the network: (1) edges that point from CERAD score to proteins, and (2) edges that connect CERAD score to confounding variables.

Network Optimization

We optimize the Bayesian network using a Monte Carlo Markov chain. The steps are as follows: First, a random network structure using all the variables is initialized. Next, a node from the network is randomly selected. Then, one of the following three operations is performed on the selected node: (1) adding an edge between the selected node and a potential parent node if the selected node has no parents; (2) deleting the edge from an existing parent; or (3) reversing the direction of the edge between the selected node and one of its existing parents. Finally, the post-operational likelihood for the selected node is calculated. To do this, a random number from the continuous uniform distribution (0, 1) is chosen; if the random number is smaller than the Metropolis-Hasting criterion, then the new network configuration is accepted; otherwise, we revert back to the original configuration. After the initialization step, the process is repeated many times until the network likelihood stabilizes.

Network Confidence Score Derivation and Consensus Network Generation

First, an ensemble of 200 networks are generated through bootstrap sampling of the original data. Next, a list of all possible edges and the associated nodes (i.e. in the form of Node A \rightarrow Node B) across all the bootstrap networks is created. The statistical confidence of each unique edge (i.e. a directional relationship

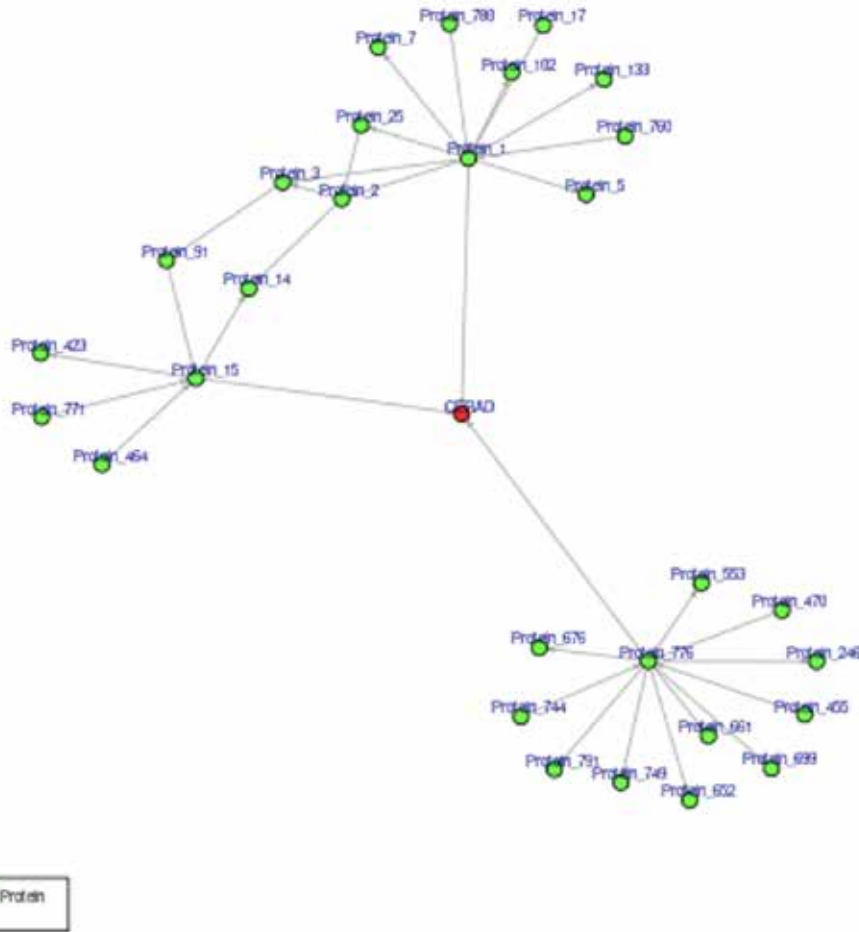


Figure 1: the consensus network generated from the AMP-AD proteomics dataset based on 200 bootstrap samples and the edge confidence threshold of 30%. Green nodes represent proteins, red node represents outcome variable (CERAD score), and blue nodes represent patient level confounding variables. In this dataset, none of the confounding variables survives the 30% edge confidence threshold.

exists between the two nodes) on the list is estimated by counting the number of times it appears among the 200 bootstrap networks. More formally, the confidence score for an edge f is calculated as:

where $f(G_i) = 1$ if and only if edge f can be extracted from the network constructed from bootstrap data set G_i . In our analysis, a cutoff of 30% (i.e., edges that appear in at least 60 out of the 200 replicates) is applied to the confidence score to select the confident edges. Finally, a consensus network is constructed by concatenating all the edges that survive the confidence score thresholding procedure described above. (Figure 1)

Biomarker Prioritization by Conditional Causal Relative Risk (CCRR)

Although the consensus network provides a visually appealing representation of the complex interplay among predictor variables (proteins and patient level

confounding variables) in jointly driving the outcome of interest; additional statistical methods need to be implemented to query the network and to extract a subset of proteins from the network for downstream validation, based on the evidence of their causal association with the outcome variable (e.g. CERAD score).

While it is tempting to select only proteins that are directly connected to the outcome variable as candidate biomarkers for the downstream validation experiments, this approach suffers from several major limitations. First, the cut point to define the edge confidence score threshold (e.g. 30% used for the example described in this paper) in constructing the final consensus network can be subjective, resulting in a somewhat arbitrary selection of candidate biomarkers. Second, proteins that are indirectly connected to the outcome variable could still exert significant effect, and sometimes, might have more desirable properties (e.g. easier to design targeting

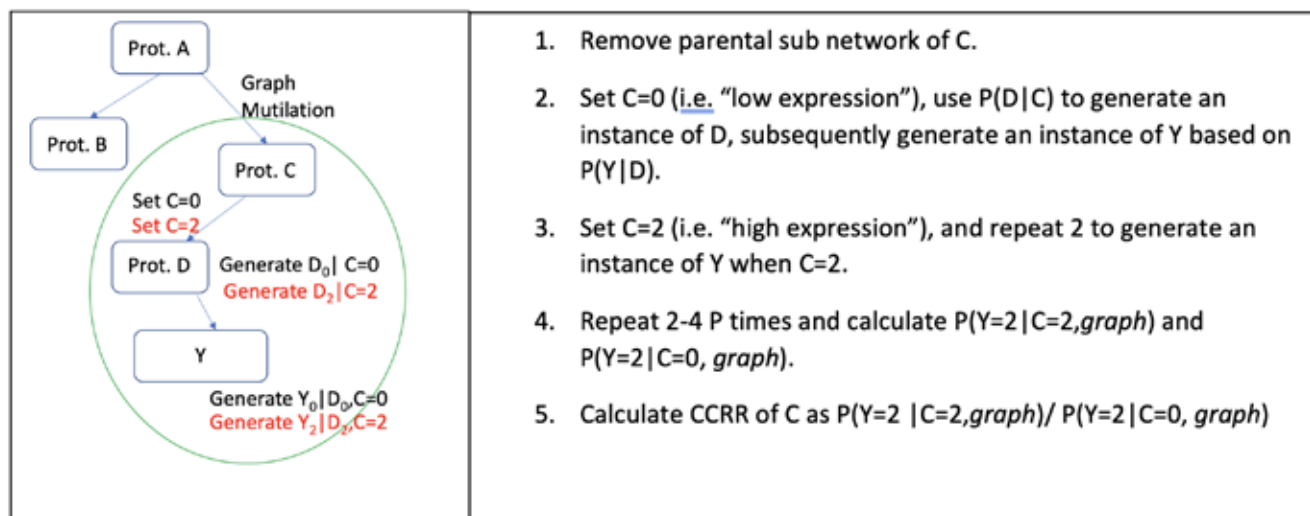


Figure 2: illustration of CCRR calculation for protein C (Prot. C) on the outcome variable Y in one of the bootstrap samples. Left panel: given the estimated Bayesian network structure in this bootstrap sample, graph mutation is first performed to remove subgraph upstream of protein C, forward propagation simulation is then carried out in the remaining subgraph (enclosed in the green oval) to calculate CCRR of protein C on Y. Right panel: pseudo code for CCRR calculation for the toy Bayesian network on the left.

probes from an assay development perspective) than those proteins that are directly connected to the outcome variable. Therefore, it is important to define a statistical metric to quantify the causal effect for any protein to the response of interest regardless of its distance to the outcome variable on the network.

We implement a graph based Conditional Causal Relative Risk (CCRR) metric to rank proteins based on their statistically causal relationship with outcome variable. More specifically, an ensemble of N (e.g. 200) networks are first generated through bootstrap sampling of the original data. Then, within each of the bootstrap network, a total number of P *in-silico* “knock-in” and “knock-out” experiments are carried out on the target protein, to quantify its causal effect (e.g. CCRR) on the outcome variable, leveraging the local conditional dependencies of this protein with its neighbor proteins as well as with the outcome variable (as captured by the Bayesian network),

To illustrate how CCRR metric is calculated, consider the Bayesian network presented in Figure 2, which represents an estimated Bayesian network from one of the N bootstrap samples. Prot. A to Prot. D are the four protein nodes, and Y is the outcome variable of interest. In this toy example, the objective is to estimate the CCRR of protein C on the outcome variable Y. We first perform graph mutilation [23] to remove the subgraph that sits upstream of protein C, as perturbing the expression of protein C will have no impact on its upstream

proteins. In the remaining sub network (as represented by the green oval), a total of P simulations are conducted. In each of the P simulations, we first manually set the expression level of protein C to “low”. Next, based on the conditional dependencies captured by the estimated Bayesian network, we sample an instance of the protein expression D from the posterior conditional distribution $P(D_0|C=low)$. Propagating forward, the same procedure will be used to generate an instance of outcome variable Y based on the previously simulated instance of D_0 and the posterior distribution of $P(Y|D_0)$ (again from the estimated Bayesian network). Following these steps, $P(Y=CERAD\ high|C=low)$ can be calculated by simply counting the number of times the simulated Y equals to high CERAD score, divided by the total number of simulations P. Similarly, $P(Y=CERAD\ high|C=high)$, or the probability of having high CERAD score when protein C is “knocked in” can also be generated.

CCRR is then defined as

$$\frac{P(Y=CERAD\ high|C=high),}{P(Y=CERAD\ high|C=low)}$$

Finally, the distribution of CCRR for protein C can be estimated by aggregating the CCRR values across all N bootstrap samples, from which we can calculate the mean and the standard error (SE) parameters.

We apply the procedures outlined above to calculate the mean CCRR and the associated standard error for all proteins in the AMP-AD data. To prioritize the proteins

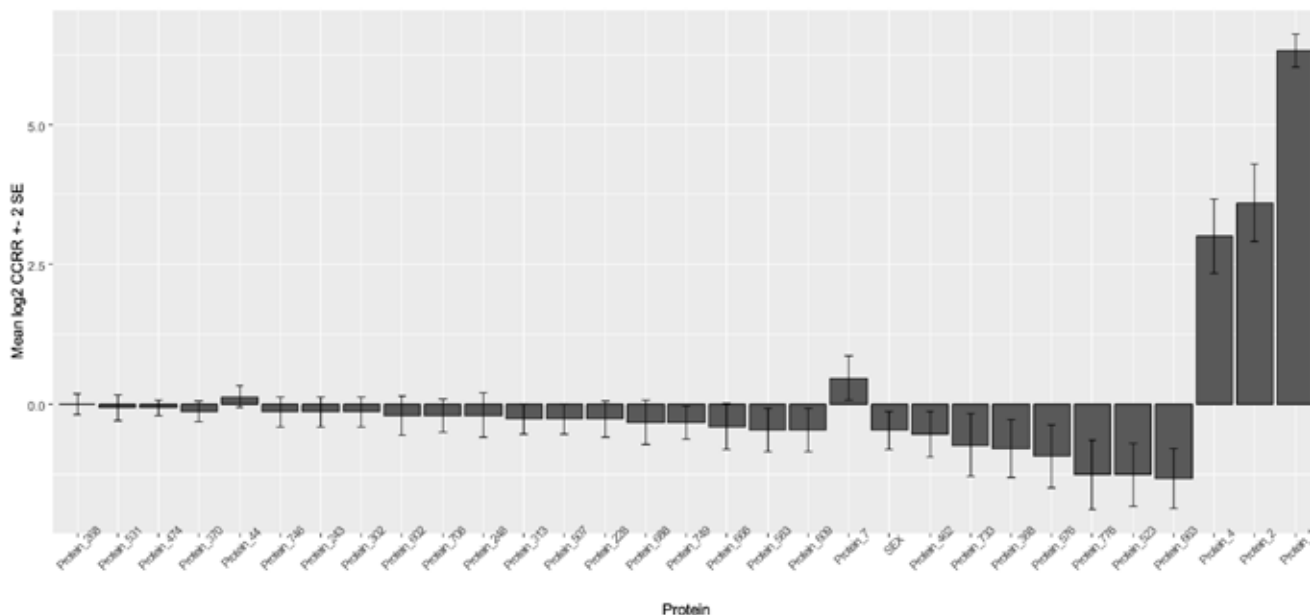


Figure 3: summary of the top ranked proteins based on the absolute value of mean log₂ (CCRR) in the AMP-AD dataset. Negative log₂ (CCRR) indicates negative association between protein expression and AD status. On the y axis, both mean and the standard errors for each protein's log₂ (CCRR) are plotted.

for downstream experimental validation, the candidate proteins are ranked by (1) the absolute value of mean log₂ (CCRR) and (2) the absolute value of the CCRR z score (i.e. mean / standard error). Figure 3 below summarizes the top ranked proteins based the mean log₂ (CCRR) value.

Finding Literature Support for Candidate Biomarkers Identified from Bayesian Network Analysis

It is worth noting that the “causal” relationships identified from the Bayesian network analysis merely capture statistical causality (i.e. conditional independence), and not necessarily the true biological causality. Therefore it is strongly recommended to establish solid literature support on the list of candidate biomarkers prior to executing expensive experimental validations using technologies such as CRISPR [20].

Discussion

Identification of prognostic and predictive biomarkers has become an integral part of drug development as the importance of precision and personalized medicine is increasingly recognized by the scientific community. A number of clinical biomarkers have been approved as companion diagnostics for cancer treatments, such as the PD-L1 protein expression [21] for pembrolizumab (Keytruda).

While the advancement of technology creates unprecedented opportunity to perform biomarker discovery using novel data modalities such as single cell RNA seq, proteomics, etc., innovative statistical methods are often required to analyze these complex datasets. In this paper, we present a graphical model (Bayesian network) based causal inference framework for high dimensional clinical biomarker discovery. Unlike the traditional propensity score method which requires the specification of one exposure variable before confounder adjustments can be applied, and therefore is more suitable for biomarker confirmation studies; Bayesian network adopts a data driven approach to automatically perform multivariate de-confounding using all variables simultaneously, and hence is appropriate for de-novo discovery. One common challenge of using Bayesian network for biomarker discovery is the lack of robust metrics that can quantify the degree of causal association between candidate biomarkers and the clinical response. In this paper, we also describe a metric CCRR (Conditional Causal Relative Risk) and its inference procedure to determine statistical significance. CCRR can be easily implemented using standard software packages such as R.

As retrospectively collected multi-omics datasets gradually became the mainstream data source for

clinical biomarker discovery, we also anticipate the rise of popularity in Bayesian network type of causal inference analysis. However, it is important to remember that the causal relationships between biomarkers and response variable identified via Bayesian network are strictly statistical; therefore, experimental validation of these candidate biomarkers remains as the critical next step.

Reference

1. Robb, M.A., P.M. McInnes, and R.M. Califf, *Biomarkers and Surrogate Endpoints: Developing Common Terminology and Definitions*. JAMA, 2016. 315(11): p. 1107-8.
2. BEST, *BEST (Biomarkers, Endpoints, and other Tools) FDA Workshop*. 2016: Silver Spring (MD)
3. Califf, R.M., *Biomarker definitions and their applications*. Exp Biol Med (Maywood), 2018. 243(3): p. 213-221.
4. Polyak, K., *Heterogeneity in breast cancer*. J Clin Invest, 2011. 121(10): p. 3786-8.
5. Austin, P.C., *An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies*. Multivariate Behav Res, 2011. 46(3): p. 399-424.
6. Pearl, J. *Bayesian networks: A model of self-activated memory for evidential reasoning*. in Proceedings, Cognitive Science Society. Irvine, CA. 1985.
7. Friedman, N., et al., *Using Bayesian networks to analyze expression data*. J Comput Biol, 2000. 7(3-4): p. 601-20.
8. Spirtes, P., *Introduction to Causal Inference*. Journal of Machine Learning Research, 2010(11): p. 1643-1662.
9. Spirtes, P.G., C; Scheines, R, *Causation, prediction, and search*. 1993: Springer.
10. Cooper, G.H., D; Meek, C, *A Bayesian approach to causal discovery*. 1997: Springer.
11. Pearl, J.V., TS, *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*. 1991.
12. AMP-AD. *Accelerating Medicines Partnership - Alzheimer's Disease*. 2016; Available from: <https://www.nia.nih.gov/alzheimers/amp-ad>.
13. Bennett, D.A., et al., *Overview and findings from the religious orders study*. Curr Alzheimer Res, 2012. 9(6): p. 628-45.
14. Bennett, D.A., et al., *Overview and findings from the rush Memory and Aging Project*. Curr Alzheimer Res, 2012. 9(6): p. 646-63.
15. Bennett, D.A., et al., *The Rush Memory and Aging Project: study design and baseline characteristics of the study cohort*. Neuroepidemiology, 2005. 25(4): p. 163-75.
16. Beach, T.G., et al., *Arizona Study of Aging and Neurodegenerative Disorders and Brain and Body Donation Program*. Neuropathology, 2015. 35(4): p. 354-89.
17. Johnson, E.C.B., et al., *Large-scale deep multi-layer analysis of Alzheimer's disease brain reveals strong proteomic disease-related changes not observed at the RNA level*. Nat Neurosci, 2022. 25(2): p. 213-225.
18. Mirra, S.S., et al., *The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease*. Neurology, 1991. 41(4): p. 479-86.
19. Karch, C.M. and A.M. Goate, *Alzheimer's disease risk genes and mechanisms of disease pathogenesis*. Biol Psychiatry, 2015. 77(1): p. 43-51.
20. Moore, J.D., *The impact of CRISPR-Cas9 on target identification and validation*. Drug Discov Today, 2015. 20(4): p. 450-7.
21. Doroshov, D.B., et al., *PD-L1 as a biomarker of response to immune-checkpoint inhibitors*. Nat Rev Clin Oncol, 2021. 18(6): p. 345-362. ■

Brief Overview of Dose Optimization in Oncology

Kentaro Takeda (Astellas), Yusuke Yamaguchi (Astellas)

1 | INTRODUCTION

In recent years, oncology drug development has moved toward molecular-targeted agents and immunooncology therapies. The primary purpose of a dose-finding trial for novel anticancer agents is to identify an optimal dose (OD), defined as the tolerable dose that has adequate efficacy under unpredictable dose-toxicity and dose-efficacy relationships. This goal differs from a dose-finding trial for traditional cytotoxic agents, which aims to determine the maximum tolerated dose (MTD). Unlike cytotoxic agents, whose efficacy and toxicity monotonically increase with dose, novel anticancer agents can exhibit nonmonotonic patterns in their dose–efficacy relationships. For example, due to the mechanism of action, the dose–efficacy relationship often plateaus at an intermediate dose when exposure reaches a saturation level in the body; therefore, a further increase in the dose level may not improve the efficacy¹. In some cases, the efficacy of the new anticancer agent can even decrease at higher dose levels and exhibit a bell-shaped dose–efficacy relationship².

Sachs et al. (2016) provided evidence that non-MTD development strategies have been successfully implemented, including strategies for several drugs with doses substantially lower than MTD³. Figure 1 shows a summary of the results of the review of MTDs and approved doses. The use of these strategies can be motivated, for example, by (i) the MTD being above a pragmatic dose for many new targeted therapies or

(ii) safety considerations. Examples of how development can succeed without using MTD, such as through modeling and simulation-based methods, illustrate the impact of such strategies. Zirkelbach et al. (2022) reviewed the initial approvals by the US FDA (2019–2021) of small molecules and antibody–drug conjugates for oncologic indications to determine the proportion with a recommended dose at the maximum tolerated dose or the maximum administered dose, highlighted strategies to integrate dose optimization into the development of premarketing drugs, and discuss the underlying statistical principles⁴.

The FDA oncology center of excellence initiated the project Optimus to reform the paradigm of dose optimization and dose selection in the development of oncology drugs in 2022⁵ and issued a draft guidance 'Optimizing the Dosage of Human Prescription Drugs and Biological Products for the Treatment of Oncologic Diseases'⁶. The project aims to ensure that cancer drug doses are optimized to maximize efficacy, safety and tolerability and to understand the pharmacokinetics (PK), pharmacodynamics (PD), toxicity, and efficacy at each dose level. For this purpose, the project Optimus recommends that sponsors plan their development programs such that the identification of the optimal dose(s) can occur prior to or concurrently with the establishment of the drug's safety and effectiveness. This article provides a brief review of dose optimization approaches in oncology.

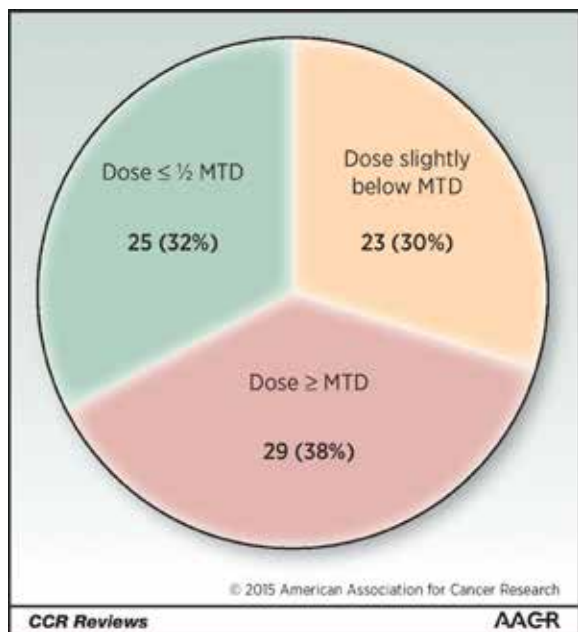


Figure 1: Summary of results from the review of MTDs and approved doses. Roughly two thirds (48/77) of the compounds have approved doses less than MTD, with roughly one third of them being dosed at less than one half of MTD.

2 | DOSE-FINDING BASED ON EFFICACY AND TOXICITY

To accommodate this paradigm shift, various model-based designs and model-assisted designs have been proposed to identify ODs by incorporating both efficacy and toxicity responses in early phase dose-finding trials. These designs can be categorized as: 1) fully sequential approaches and 2) two-stage approaches.

2.1 | Fully sequential approach

Fully sequential approaches continuously update the estimate of the toxicity and efficacy profile of each dose after each cohort to determine the dose assignment and directly find the OD. Therefore, fully sequential approaches could be more

efficient and often require a smaller sample size than two-stage designs. Fully sequential approaches usually do not have separated randomization part even though they implicitly randomize patients among promising doses.

Various model-based designs have been proposed assuming dose-toxicity and dose-efficacy relationship models in early phase dose-finding trials. Braun (2002) extended the continual reassessment method by constructing a joint probability model⁷. Thall and Cook (2004) proposed the efficacy-toxicity (EffTox) trade-off design that uses logistic models with copula to jointly model toxicity and efficacy and selects the dose based on toxicity-efficacy desirability contours⁸. Bekele and Shen (2005) proposed a dose-finding approach for correlated bivariate binary toxicity and continuous efficacy outcomes⁹. Zhang et al. (2006) proposed the flexible continuation ratio model and the straightforward OD selection criteria. Yin et al. (2006) proposed a curve-free method to jointly model bivariate binary data and conducted a risk-benefit trade-off using odds ratio contours¹⁰. Riviere et al. (2015) proposed a Bayesian phase I/II design incorporating a plateau parameter in a proportional hazard model for time-to-efficacy¹¹.

To simplify the implementation of phase I/II trials, several model-assisted designs have been proposed. Lin and Yin (2017) proposed a simple toxicity and efficacy interval (STEIN) design to find a safe dose with the highest efficacy¹². Takeda et al. (2018) developed the BOIN-ET design to extend the Bayesian optimal interval (BOIN) design to phase I/II trials¹³. Lin et al. (2020) proposed the utility-based Bayesian optimal interval phase I/II (BOIN12) design to find OD for immunotherapies and targeted therapies¹⁴. Li et al. (2017) proposed a Bayesian adaptive toxicity and efficacy probability

interval (TEPI) design, which took into account both the toxicity and efficacy outcomes within the mTPI framework¹⁵. Lin and Ji (2021) modified the TEPI design and proposed a probability interval of toxicity and efficacy (PRINTE) design¹⁶. Shi et al. (2021) extended the keyboard design and proposed a utility-based toxicity probability interval (uTPI) design¹⁷.

For rule-based design, Lin and Ji (2020) proposed the joint i3+3 (Ji3+3) design to incorporate both the simplicity and transparency of rule-based designs¹⁸.

2.2 | Two-stage approach

Two-stage approaches first detect admissible doses defined a set of doses that satisfy the lowest prespecified safety (and efficacy) requirements at stage I, followed by the identification of an OD at stage II. Two-stage approaches often randomize patients among admissible doses at stage II.

Pan et al. (2014) proposed a seamless dose escalation/expansion with adaptive randomization scheme (SEARS) to identify an OD¹⁹. Zang and Lee (2017) proposed a two-stage design to simultaneously evaluate the dose-toxicity and dose-efficacy relationships. Zhou et al. (2019) proposed the two-stage U-BOIN design, in which the trade-off between efficacy and toxicity is measured by utility elicited from clinicians²⁰. Han et al. (2021) proposed a two-stage nonparametric (TSNP) phase I/II clinical trial design to identify OBD for immunotherapy²¹.

3 | LATE-ONSET

Late-onset efficacy and toxicity outcomes, where the occurrences of some of the treatment-related efficacy responses and

adverse events are delayed due to the nature of the treatments, are another different toxicity characteristic for some of the new noncytotoxic anticancer agents. It has been reported that 57% of the toxicities of grades 3-4 occur after cycle 1 of treatment from 36 clinical trials of molecularly targeted agents²². For immune checkpoint inhibitors, it is well known that some of immune-related toxicity can have delayed onset; for example, endocrinopathies occur late and have been observed between weeks 12 and 24²³. Ignoring the feature of late-onset toxicity in modern oncology trials will cause potential dose reductions or interruptions in subsequent trials, adding unnecessary uncertainties in dose optimization and delay in treatment development. While the importance of accelerating the drug development process is ever greater, other factors could also delay early stage dose-finding trials; for example, most traditional designs used in early stage oncology trials require that the efficacy and toxicity results of those patients who have already been enrolled in the trial should be available at the interim decision time point, and therefore do not allow sequential enrollment when previous patients have not completed the required efficacy and toxicity assessments. This suspension of enrollment could cause the slowdown of the trials, especially when multiple adaptive decisions on dose-escalation or de-escalation need to be made at different interim decision-making time points. Note that when the accrual rate is fast, the waiting time needed for those enrolled patients to complete their efficacy and toxicity assessments could delay the immediate access to treatment for the new patients, and therefore could further prolong the duration of the trial. Furthermore, the difference in the evaluation period between efficacy and toxicity outcomes could delay decision-making. This problem

is exacerbated in many immunooncology trials, where frequently late-onset outcomes require significantly longer observation windows for a robust evaluation of efficacy and toxicity. To efficiently develop new anticancer agents, novel adaptive designs for early stage trials are needed to evaluate efficacy and toxicity outcomes while allowing sequential enrollment and minimizing treatment interruptions or pauses so that the duration of the trial can be reduced. To solve the problem, several approaches are proposed. Jin et al. (2014) extended the EffTox design and proposed the LO-EffTox design by using data augmentation to impute missing outcomes from posterior predictive distributions computed from partial follow-up times and complete outcome data²⁴. Takeda et al. (2020) extended the BOIN-ET design and proposed the TITE-BOIN-ET design using a partial likelihood with pending data²⁵. Zhou et al. (2022) extended the BOIN12 design and proposed the TITE-BOIN12 design using a Bayesian data augmentation and a partial likelihood with pending data²⁶. Zhang and Zang (2021) proposed the conditional weighted likelihood (CWL) method to address the delayed effect and applied the method in U-BOIN (CWL-U-BOIN)²⁷.

4 | ORDINAL GRADED OUTCOME

Most of the trial designs for the OD finding assume that the efficacy and toxicity data are summarized as binary outcomes, scored as a response or not a response for efficacy, and scored as a dose-limiting toxicity (DLT) or not a DLT for toxicity. However, unlike cytotoxic agents, molecular-targeted agents and immuno-oncology therapies appear more likely to induce multiple low or moderate-grade toxicities than DLTs^{28,29,30}.

Additionally, for efficacy, it is preferable to evaluate overall response (ORR) and long-term stable disease (SD) in solid tumors and consider the difference between complete remission (CR) and partial remission (PR) in lymphoma. Therefore, it is crucial to incorporate the toxicity grade and multiple efficacy levels in dose-finding and decision making to determine the OD for developing novel anticancer agents. Thall and Nguyen (2012) proposed a sequentially outcome-adaptive Bayesian design based on the elicited utility of bivariate ordinal efficacy and toxicity outcomes with adaptive randomization³¹. Takeda et al. (2022, 2023) proposed a generalized Bayesian optimal interval design for dose finding that takes into account ordinal graded efficacy and toxicity outcomes (gBOIN-ET design and TITE-gBOIN-ET design)^{32,33}.

5 | DOSE-FINDING DESIGN WITH PK/PD

Pharmacologically, PK information is considered an appropriate indicator to assess the degree of drug intervention in humans. Several significant exposure–efficacy relationships have been identified in several immune checkpoint inhibitors, and some PK information has been considered important predictors of efficacy endpoints, e.g., overall response, immune response, and overall survival³⁴. Therefore, including PK evaluation in oncology dose-finding trials may enable us to determine OD with greater precision. Guo and Yuan (2023) proposed a dose-ranging approach to optimizing dose (DROID) for oncology clinical trials with targeted agents based jointly on three endpoints: a toxicity endpoint, an efficacy endpoint, and an efficacy surrogate endpoint such as a

PD biomarker³⁵. DROID is a two-stage approach and randomizes patients among admissible doses in stage II. Takeda et al. (2023) proposed a Bayesian optimal interval design for dose optimization with a randomization scheme based on PK outcomes³⁶. The decision to allocate the dose for the proposed design is guided by toxicity and PK outcomes. The PK outcome is considered a surrogate indicator of the efficacy outcome. After the trial is completed, the OD is determined on the basis of all toxicity, efficacy, and PK outcomes. These approaches, in addition to the efficacy and toxicity outcomes, incorporate the PK/PD outcome into the trial design for decision making under a randomization scheme.

6 | INCORPORATING HISTORICAL DATA

The sample size in oncology dose-finding trial is often limited. Therefore, it should be valuable to consider incorporating historical information into dose-finding designs. Thall et al. (2014) discussed an algorithm to determine prior hyperparameters using least squares penalized by effective sample size for the EffTox design³⁷. Zhao et al. (2022) proposed the iBOIN-ET design that incorporates historical study information into the BOIN-ET design through the concepts of skeleton and prior effective sample size using Bayesian inference³⁸.

7 | SELECTION OF OD

Most of the model-assisted designs determine the dose of the next cohort based on a prespecified decision rule and select OD based on some measures derived using all data on the tox-

In some designs, the OD is simply determined by selecting a dose that is tolerable and has the highest estimated efficacy probability. Some designs used utility functions to measure the trade-off between toxicity and efficiency and chose a dose that maximized utility, where the definition of utility function varied between designs. For example, the STEIN design used a weighted function of the estimated toxicity and efficacy probabilities. The TEPI design and the PRINTE design use truncated linear functions of the estimated probabilities of toxicity and efficacy. The U-BOIN design and the BOIN12 design use the efficacy-toxicity trade-off directly scoring the toxicity and efficacy responses. These utilities are flexible and affect the operating characteristics of the final OD selection. Yamaguchi et al. (2023) have reported the operating characteristics of OD selection approaches of model-assisted designs through comprehensive simulation studies³⁹.

8 | EXAMPLES

Dose optimization approaches introduced in the previous sections are implemented in real clinical trials. For example, the EffTox design is used in a phase I/II trial of CAR-transduced natural killer cells in CD19 positive lymphoid tumors⁴⁰. The LO-EffTox design is implemented in a phase I/II trial of sitravatinib and nivolumab in clear cell renal cell carcinoma after progression of antiangiogenic therapy⁴¹ and a phase I/II trial in advanced pancreatic cancer⁴². The BOIN12 design is used in a phase I/II study of enhanced CD33 CAR T cells in subjects with relapsed or refractory acute myeloid

I/II trial of genetically engineered cells (COH06) with or without Atezolizumab for the treatment of non-small cell lung cancer previously treated with immune checkpoint inhibitors (NCT05334329) and a phase II extension trial for the evaluation of safety and efficacy in patients with breast cancer (NCT05334329). The TITE-BOIN-ET design is used in a randomized phase I/ II multicenter study evaluating combination of luspatercept in LR-MDS without RS having failed or being ineligible to ESA (NCT05181735).

9 | SOFTWARE

Some dose optimization approaches provide Web application/R package/Codes as listed in Table 1.

10 | DISCUSSION

We have briefly reviewed dose optimization approaches in oncology. Model-based designs assume complicated dose-toxicity and dose-efficacy models with a potential risk of

misspecification and require real-time model fitting and estimation for each decision making. Therefore, model-based optimal dose-finding designs are rarely used in practice, except for the EffTox design. Compared to model-based designs, model-assisted designs can pre-tabulate decision tables before the trial starts, are simple and transparent to implement, and are easier to understand by the clinical community. Yuan et al. (2017) provide a comprehensive review of phase I-II trials, in particular model-based designs⁴³. Yuan et al. (2022) provide a comprehensive review of model-assisted design, including phase I-II trial designs⁴⁴.

The appropriate designs are different in each dose-finding trial. Massive preliminary simulation considering the mechanism of action, indication, endpoints, target efficacy and toxicity probabilities, evaluation period, accrual rate, and multiple settings is quite important to implement the novel designs in real clinical trials. In this regard, physicians, biostatisticians, and other stakeholders must work closely together, for example, considering the purpose of the trial, realistic settings, and operational issues.

TABLE 1 Available Software List

Design	Web application/R Package/Codes
EffTox	https://biostatistics.mdanderson.org/SoftwareDownload/SingleSoftware/Index/2
BOIN12	https://www.trialdesign.org/
TITE-BOIN12	https://www.trialdesign.org/
U-BOIN	https://www.trialdesign.org/
TEPI	EAST Bayes
PRINTE	EAST Bayes
Ji3+3	EAST Bayes
BOIN-ET	R package boinet: https://cran.r-project.org/web/packages/boinet/index.html
gBOIN-ET	R package boinet: https://cran.r-project.org/web/packages/boinet/index.html
TITE-BOIN-ET	R package boinet: https://cran.r-project.org/web/packages/boinet/index.html
TITE-gBOIN-ET	R package boinet: https://cran.r-project.org/web/packages/boinet/index.html
SEARS	R package sears: https://cran.r-project.org/web/packages/SEARS/index.html
uTPI	https://github.com/haoluns/uTPI
TSNP	https://github.com/yongzang2020/TSNP

References

1. Druker BJ. Perspectives on the development of a molecularly targeted agent. *Cancer Cell* 2002; 1(1): 31–36.
2. Postel-Vinay S, Arkenau HT, Olmos D, et al. Clinical benefit in Phase-I trials of novel molecularly targeted agents: does dose matter?. *Br. J. Cancer* 2009; 100(9): 1373–1378.
3. Sachs JR, Mayawala K, Gadamsetty S, Kang SP, Alwis dDP. Optimal Dosing for Targeted Therapies in Oncology: Drug Development Cases Leading by Example. *Clin. Cancer Res.* 2016; 22(6): 1318–1324.
4. Fourie Zirkelbach J, Shah M, Vallejo J, et al. Improving Dose-Optimization Processes Used in Oncology Drug Development to Minimize Toxicity and Maximize Benefit to Patients. *J. Clin. Oncol.* 2022; 40(30): 3489–3500.
5. U.S. Food and Drug Administration . Project Optimus. 2022.
6. U.S. Food and Drug Administration . Optimizing the Dosage of Human Prescription Drugs and Biological Products for the Treatment of Oncologic Diseases. 2023.
7. Braun TM. The bivariate continual reassessment method. extending the CRM to phase I trials of two competing outcomes. *Control. Clin. Trials* 2002; 23(3): 240–256.
8. Thall PF, Cook JD. Dose-finding based on efficacy-toxicity trade-offs. *Biometrics* 2004; 60(3): 684–693.
9. Bekele BN, Shen Y. A Bayesian approach to jointly modeling toxicity and biomarker expression in a phase I/II dose-finding trial. *Biometrics* 2005; 61(2): 343–354.
10. Yin G, Li Y, Ji Y. Bayesian dose-finding in phase I/II clinical trials using toxicity and efficacy odds ratios. *Biometrics* 2006; 62(3): 777–784.
11. Riviere MK, Yuan Y, Dubois F, Zohar S. A Bayesian dose finding design for clinical trials combining a cytotoxic agent with a molecularly targeted agent. *J. R. Stat. Soc. Ser. C Appl. Stat.* 2015; 64(1): 215–229.
12. Lin R, Yin G. STEIN: A simple toxicity and efficacy interval design for seamless phase I/II clinical trials. *Stat. Med.* 2017; 36(26): 4106–4120.
13. Takeda K, Taguri M, Morita S. BOIN-ET: Bayesian optimal interval design for dose finding based on both efficacy and toxicity outcomes. *Pharm. Stat.* 2018; 17(4): 383–395.
14. Lin R, Zhou Y, Yan F, Li D, Yuan Y. BOIN12: Bayesian Optimal Interval Phase I/II Trial Design for Utility-Based Dose Finding in Immunotherapy and Targeted Therapies. *JCO Precis Oncol* 2020; 4: 1393–1402.
15. Li DH, Whitmore JB, Guo W, Ji Y. Toxicity and Efficacy Probability Interval Design for Phase I Adoptive Cell Therapy Dose-Finding Clinical Trials. *Clin. Cancer Res.* 2017; 23(1): 13–20.

16. Lin X, Ji Y. Probability intervals of toxicity and efficacy design for dose-finding clinical trials in oncology. *Stat. Methods Med. Res.* 2021; 30(3): 843–856.
17. Shi H, Cao J, Yuan Y, Lin R. uTPI: A utility-based toxicity probability interval design for phase I/II dose-finding trials. *Stat. Med.* 2021; 40(11): 2626–2649.
18. Lin X, Ji Y. The Joint i3+3 (Ji3+3) design for phase I/II adoptive cell therapy clinical trials. *J. Biopharm. Stat.* 2020; 30(6): 993–1005.
19. Pan H, Xie F, Liu P, Xia J, Ji Y. A phase I/II seamless dose escalation/expansion with adaptive randomization scheme (SEARS). *Clin. Trials* 2014; 11(1): 49–59.
20. Zhou Y, Lee JJ, Yuan Y. A utility-based Bayesian optimal interval (U-BOIN) phase I/II design to identify the optimal biological dose for targeted and immune therapies. *Stat. Med.* 2019; 38(28): 5299–5316.
21. Han Y, Liu H, Cao S, Zhang C, Zang Y. TSNP: A two-stage nonparametric phase I/II clinical trial design for immunotherapy. *Pharm. Stat.* 2021; 20(2): 282–296.
22. Postel-Vinay S, Gomez-Roca C, Molife LR, et al. Phase I trials of molecularly targeted agents: should we pay more attention to late toxicities?. *J. Clin. Oncol.* 2011; 29(13): 1728–1735.
23. Weber JS, Yang JC, Atkins MB, Disis ML. Toxicities of Immunotherapy for the Practitioner. *J. Clin. Oncol.* 2015; 33(18): 2092–2099.
24. Jin IH, Liu S, Thall PF, Yuan Y. Using Data Augmentation to Facilitate Conduct of Phase I-II Clinical Trials with Delayed Outcomes. *J. Am. Stat. Assoc.* 2014; 109(506): 525–536.
25. Takeda K, Morita S, Taguri M. TITE-BOIN-ET: Time-to-event Bayesian optimal interval design to accelerate dose-finding based on both efficacy and toxicity outcomes. *Pharm. Stat.* 2020; 19(3): 335–349.
26. Zhou Y, Lin R, Lee JJ, et al. TITE-BOIN12: A Bayesian phase I/II trial design to find the optimal biological dose with late-onset toxicity and efficacy. *Stat. Med.* 2022; 41(11): 1918–1931.
27. Zhang Y, Zang Y. CWL: A conditional weighted likelihood method to account for the delayed joint toxicity–efficacy outcomes for phase I/II clinical trials. *Stat. Methods Med. Res.* 2021; 30(3): 892–903.
28. Brahmer JR, Drake CG, Wollner I, et al. Phase I study of single-agent anti-programmed death-1 (MDX-1106) in refractory solid tumors: safety, clinical activity, pharmacodynamics, and immunologic correlates. *J. Clin. Oncol.* 2010; 28(19): 3167–3175.
29. Le Tourneau C, Diéras V, Tresca P, Cacheux W, Paoletti X. Current challenges for the early clinical development of anticancer drugs in the era of molecularly targeted agents. *Target. Oncol.* 2010; 5(1): 65–72.
30. Penel N, Adenis A, Clisant S, Bonnetterre J. Nature and subjectivity of dose-limiting toxicities in contemporary phase 1 trials: comparison of cytotoxic versus non-cytotoxic drugs. *Invest. New Drugs* 2011; 29(6): 1414–1419.

31. Thall PF, Nguyen HQ. Adaptive randomization to improve utility-based dose-finding with bivariate ordinal outcomes. *J. Biopharm. Stat.* 2012; 22(4): 785–801.
32. Takeda K, Morita S, Taguri M. gBOIN-ET: The generalized Bayesian optimal interval design for optimal dose-finding accounting for ordinal graded efficacy and toxicity in early clinical trials. *Biom. J.* 2022; 64(7): 1178–1191.
33. Takeda K, Yamaguchi Y, Taguri M, Morita S. TITE-gBOIN-ET: Time-to-event generalized Bayesian optimal interval design to accelerate dose-finding accounting for ordinal graded efficacy and toxicity outcomes. *Biom. J.* 2023; 65(7): e2200265.
34. Centanni M, Moes DJAR, Trocóniz IF, Ciccolini J, Hasselt vJGC. Clinical Pharmacokinetics and Pharmacodynamics of Immune Checkpoint Inhibitors. *Clin. Pharmacokinet.* 2019; 58(7): 835–857.
35. Guo B, Yuan Y. DROID: dose-ranging approach to optimizing dose in oncology drug development. *Biometrics* 2023.
36. Takeda K, Zhu J, Li R, Yamaguchi Y. A Bayesian optimal interval design for dose optimization with a randomization scheme based on pharmacokinetics outcomes in oncology. *Pharm. Stat.* 2023.
37. Thall PF, Herrick RC, Nguyen HQ, Venier JJ, Norris JC. Effective sample size for computing prior hyperparameters in Bayesian phase I-II dose-finding. *Clin. Trials* 2014; 11(6): 657–666.
38. Zhao Y, Liu R, Takeda K. Incorporating historical information to improve dose optimization design with toxicity and efficacy endpoints: iBOIN-ET. *Pharm. Stat.* 2023; 22(3): 440–460.
39. Yamaguchi Y, Takeda K, Yoshida S, Maruo K. Optimal biological dose selection in dose-finding trials with model-assisted designs based on efficacy and toxicity: a simulation study. *J. Biopharm. Stat.* 2023; in press: 1–15.
40. Liu E, Marin D, Banerjee P, et al. Use of CAR-Transduced Natural Killer Cells in CD19-Positive Lymphoid Tumors. *N. Engl. J. Med.* 2020; 382(6): 545–553.
41. Msaouel P, Goswami S, Thall PF, et al. A phase 1-2 trial of sitravatinib and nivolumab in clear cell renal cell carcinoma following progression on antiangiogenic therapy. *Sci. Transl. Med.* 2022; 14(641): eabm6420.
42. Tidwell RSS, Thall PF, Yuan Y. Lessons Learned From Implementing a Novel Bayesian Adaptive Dose-Finding Design in Advanced Pancreatic Cancer. *JCO Precis Oncol* 2021; 5(5): 1719–1726.
43. Yuan Y, Nguyen HQ, Thall PF. *Bayesian Designs for Phase I-II Clinical Trials*. Chapman & Hall/CRC Biostatistics Series London, England: Taylor & Francis . 2017.
44. Yuan Y, Lin R, Jack Lee J. *Model-assisted Bayesian Designs for Dose Finding and Optimization: Methods and Applications*. C&Hall/CRC Press . 2022.

2023 JSM SESSION ON “CLINICAL AND NONCLINICAL STATISTICS ROLES IN THE PHARMACEUTICAL INDUSTRY: HOW KNOWLEDGE OF ONE HELPS THE OTHER”

By Stan Altan, W. Scott Clark, Imola Fodor, Cyrus Hoseyni, John Kolassa, Eve Pickering, and Shanthi Sethuramen

1. Introduction

A career development session, sponsored by the Career Development committee of the Biopharmaceutical Section of the ASA, was held on Tuesday August 8, 2023, at the Joint Statistical Meetings, in Toronto. The session was organized by Stan Altan (Janssen), Scott Clark (Eli Lilly) and John Kolassa (Rutgers University), with the following senior leaders of statistics departments in the pharmaceutical industry serving as panelists: Imola Fodor (Genentech), Cyrus Hoseyni (Janssen), Eve Pickering (Pfizer), and Shanthi Sethuraman (Eli Lilly). The session abstract is given in Appendix 1.

2. Session chair’s introductory remarks

Session chair, Scott Clark, began by introducing the main topics for discussion, highlighting the need for a closer

linkage between the clinical and nonclinical spaces. This is occurring as “patient centric specifications” considerations in the nonclinical space intersects with “Patient-focused drug development (PFDD)” in the clinical space. Translational medicine has traditionally been at the juncture of nonclinical and clinical pursuits. But more recently, scientific and regulatory developments have motivated companies to study product quality attributes with clinical relevance considerations in mind. This convergence is expected to expand into the future. Given this convergence, the question was raised as to how career development benefits from knowledge and experience across the two areas. As further background for the relevance of the session topics, the subject matter spheres pertinent to the clinical and nonclinical areas were reviewed as summarized in Figure 1.

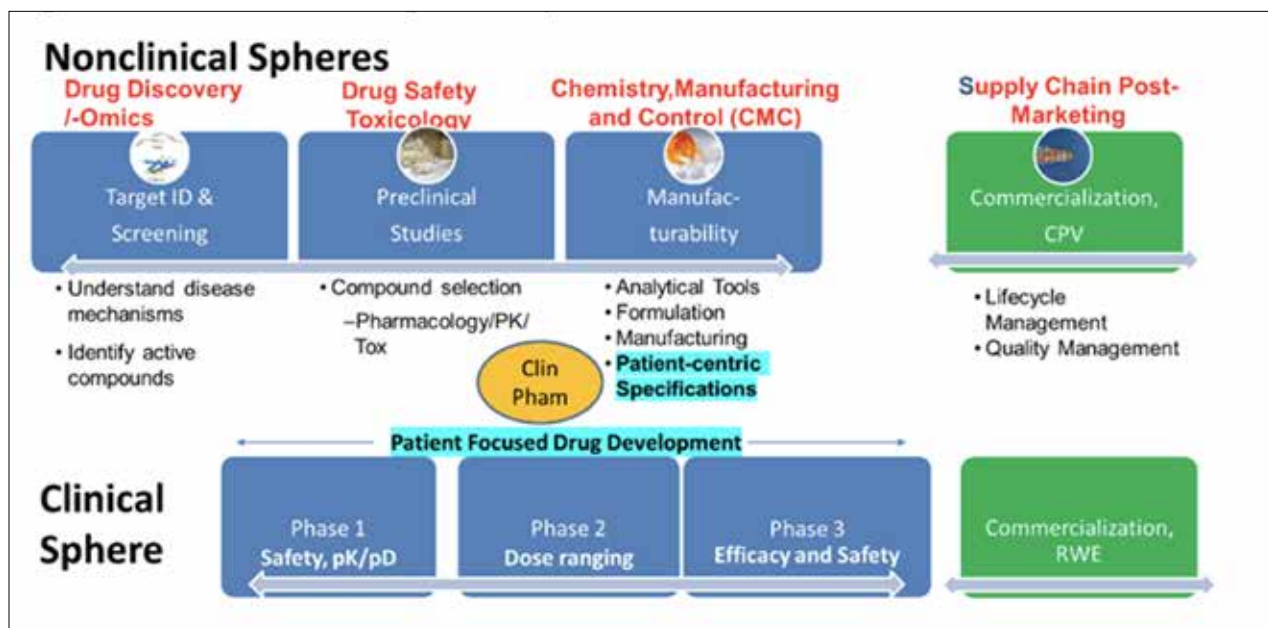


Figure 1 Statistics in Drug Development

Two aspects of the current technological landscape having potential to dramatically impact statisticians in the pharmaceutical industry were highlighted.

1. The continuing emphasis on patient-focused drug development in clinical studies and the parallel development of patient centric specifications in the nonclinical space. These initiatives imply a need for greater collaborations between the clinical and nonclinical spaces moving forward. Specifically Figure 2 shows graphically the organic linkages between manufacturing process, the pharmaceutical product and the patient outcomes, and product quality.



Figure 2 Process-Product-Patient Linkages and Quality

2. Rapidly evolving technologies and science driving drug development is another important driver of the expected convergence. Closer collaborations are expected to accelerate the integration of these technological advances into the drug development process (Figure 3). Setting the stage for further discussions, another question was posed “Are nonclinical to clinical connections ready for the coming revolution?” with the goal of illustrating the advantages of acquiring a basic understanding of the role of nonclinical statistics to the clinical statistician, and the reverse, advantages to acquiring a basic understanding of the role of clinical statistics to the nonclinical statistician. In the discussions, one important goal was to address

the question of how this helps career development. Ideas on how to bridge any gaps in the connections is also sought. What are the connections points, what needs to be improved, and how does product quality impact on clinical outcomes?

3. Panelists’ opening statements.

Scott Clark introduced each of the panelists, followed by the panel discussion led by John Kolassa. It kicked off with opening statements by each of the panelists. The panelists gave a brief perspective on the main topic of the session. These are summarized as follows.

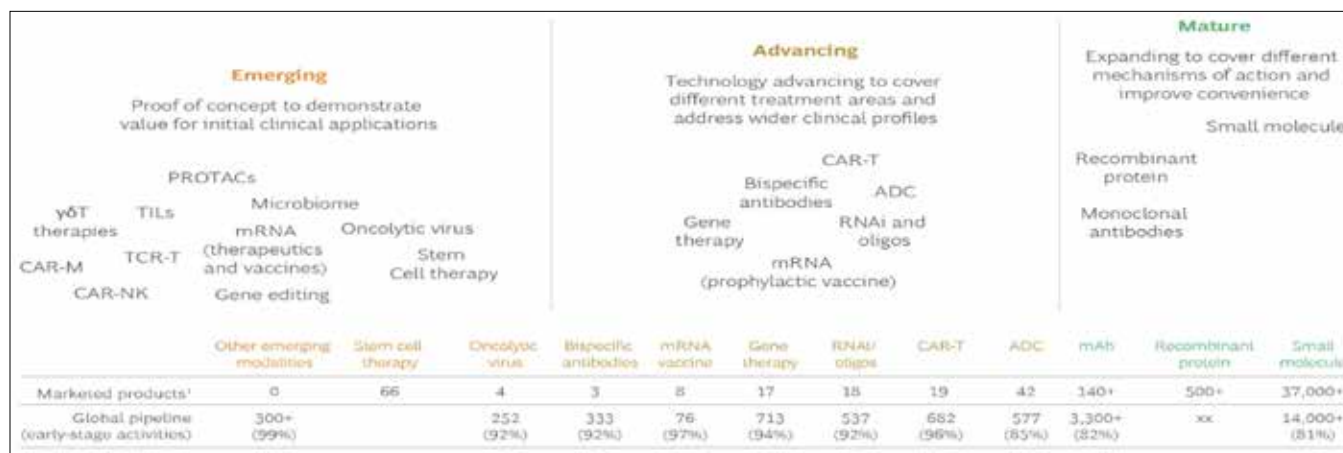


Figure 3 New Drug Modalities developed since 2000 <https://www.bcg.com/publications/2023/benefits-and-risks-of-new-drug-modalities>

Imola Fodor (Vice President Data and Statistical Sciences, Hematology and Early Development, Genentech)

My professional journey started in nonclinical statistics for 10 years, and then I moved to clinical. I very much appreciate the diversity of statistical tools and approaches used across the drug development process, and the impact that statisticians can have on the business. Having that experience in early development and nonclinical statistics was a major advantage to appreciate better the compound's late development challenges.

Cyrus Hoseyni (Global Head, Statistics and Decision Sciences, Janssen)

The first 3 years post-graduation I spent supporting early development, toxicology, and in parallel, late development. Observing up close the end to end process provided me with valuable experience to help connect the parts of the entire process of drug development. This was not serendipitous; it was helped by having a sense of curiosity and seeking opportunities to learn and engage. Leaving such opportunities to chance is not sufficient, a proactive attitude, being ready to help, is essential to learning and growing.

Eve Pickering (Vice President, Nonclinical Statistics, Pfizer)

I had an undergraduate degree in physics, spending time in a lab generating scientific data. I moved to statistics at Rutgers. Although it was a theoretical department, I gained practical experience supporting agricultural experiments. I moved to Wright State, was on the consulting center and the IACUC committee, supporting animal protocols. Then I moved to Pfizer. I was initially involved in a late development project. An early experience with the statistical issues of a biomarker and its analytical performance motivated me to return to the lab and nonclinical studies. All the work in early development feeds into later development stages, and knowing how the data is used in clinical development, helped inform my work in nonclinical.

Shanthi Sethuramen (R&D Global Analytics Officer and Sr.Vice President, Eli Lilly)

In my experience, taking short term assignments, getting outside your comfort zone and spending time in both non-clinical and clinical areas is the best way to learn and to see the bigger picture of the discovery, drug development process and post launch. Getting

experience from outside of statistics also has benefits. When I was part of project management, it allowed me to look at a molecule from birth to late stage. I gained a deeper insight into what it takes to make the molecule a product, and to optimize it later. Having an appreciation for the lifecycle of the molecule helps me to be a better critical thinker, understanding the importance of sharing knowledge across the silos and be a better drug developer. It's very importance to learn from one another and other functions.

4. Discussion questions

John Kolassa moderated the discussions and posed questions to the panelists. This was intended to be a wide-ranging discussion on a set of questions directed to the panelists. The following are the series of questions and salient points made in response to the questions (4 questions are given in Appendix 2 which were not discussed due to lack of time).

1. What are the important connections between early development/Discovery and implications on clinical studies?
 - a. From the perspective of safety, understanding toxicology data and its role in setting the "no regrets" human dose is important, especially with new modalities., and having a basic understanding of the PK/PD profile and pharmacology, mechanistic modeling helps with implications on clinical measures and outcomes.
 - b. There is a flow from discovery to the clinic, but there is also a flow back. For example, understanding immunogenicity issues in the clinic can feed back to the design of molecules in early nonclinical development, with the goal of minimizing toxicity issues in clinical trials.
 - c. We have huge databases of molecules with known physical and chemical properties, with possibly thousands of attributes. Seeking the one(s) that meets certain requirements for safety and efficacy, knowing the clinical implications informs the search for the right compound. While bioinformatics plays an important part, statistical models/methodologies help with potential multiplicity issues during the analyses and plays a significant role in the design of the molecules.

2. How do we contrast the diversity of statistical applications in nonclinical vs clinical and how does this impact on career trajectories?

a. The beauty of nonclinical and clinical statistics is that there are so many interesting questions requiring statistical approaches. Being less regulated in some cases in nonclinical, there is more flexibility in statistical approaches and innovative models. The nature of the variety of challenging problems in areas of nonclinical allow more creative and interesting statistical approaches. Experience with diverse statistical tools is important no matter which direction a statistician takes.

b. In clinical, patient safety is paramount. However, there is plenty of room on the clinical side to develop clinical development plans (CDPs). This is not a typical statistical question, but using our training in statistics and quantitative skills can help to inform the writing of the CDPs. This is a basic skill important to career development.

c. The ratio of statisticians to scientists in nonclinical might be 1:400, so the duration of projects is shorter, and multiple projects in parallel is the norm. Therefore, project management becomes important. This is a skill that can be taken forward into the clinical side, a learning from non-clinical.

3. What is the potential and current role of AI and ML in the clinical and nonclinical spaces?

a. AI and ML are best applied in the hands of experts who exercise a balance between augmentation and automation. Some caution is needed in placing such tools into the hands of novices, especially in the application of off the shelf applications. The statistician's role in helping understand the appropriate methodologies, modeling and statistical issues such as multiplicity are critical to getting to the right hypotheses and solutions.

b. Image analysis in cancer studies, and augmenting pathology studies.

c. Opens up opportunities for deeper collaborations with data science colleagues and related approaches.

4. What kind of academic training can prepare statisticians to pursue clinical or nonclinical statistics roles in the industry. Would some kind of continuing education courses or certification mitigate lack of preparedness for an industry career.

a. Having direct hands-on experience generating and handling data, translates to more effective understanding of data collection and generation by our collaborators. Taking classes in biology or chemistry can help to provide the language of our scientific collaborators, to enhance communication.

b. A statistics degree is not sufficient, interdisciplinary, or additional training in the sciences relevant to drug development makes you a better drug developer. Knowing the scientific language enhances collaboration.

c. Academic programs could consider internship programs, or having visiting lectures from industry statisticians to bring to students the experiences of statisticians on the job as one way to ameliorate the lack of awareness, for example different aspects of the drug development process. We should look for ways to collaborate more between industry and academia to address this.

5. What is the practice for making sure that nonclinical data are summarized and presented accurately?

a. Approaching this not as policemen of the process, but asking what the goal is, are there biases in the data, making objective data-based decisions.

b. Data presentation and communicating findings clearly.

c. Education should be part of the role. Due diligence requires looking at the totality of the data, how we come into the process, sharing the common goal of making the right decision based on the data. We must draw valid conclusions based on the data.

6. How can the statistical stakeholders add clarity to the dialogue? What Collaborations are desirable?

a. Stakeholders may include both internal partners, e.g., clinical statistics, biomarker groups, PK/Pharmacology/Pharmacometrics, commercial supply chain; external partners,

IQ consortium, DIA, AAPS, academic institutions, regulatory agencies.

b. Understanding the key factors that could influence the outcome, understanding the underlying assumptions, informs the proper design of experiments, statistical analysis and modeling.

c. Expanding your network across the industry, with regulators, scientific colleagues, not just statisticians. Consider working on industry working groups, especially those that are comprised of different disciplines, different companies, universities, all working on a research effort together. This experience will provide greater understanding of the big picture, enhances innovation and the ability to work more effectively.

7. Can we incorporate QbD principles as a background for pursuing a patient centric specification? What are the hurdles?

a. Linking quality attributes to clinical outcomes is an important goal. QbD is fundamentally defining an operating space, whether it's manufacturing, pKpD, or clinical. In manufacturing the goal is to justify expansion of the operating space, but in clinical, it is the opposite, product variability in clinical trials is reduced. This basic conflict needs some resolution, but finding an approach efficiently, is still a challenge.

b. The question of how CMC attributes relate to ADME properties is an important question, the concept is appealing but can this question be brought into the clinic using a DoE subject to time and cost constraints? This is still a challenge.

c. This is an emerging topic that will require deeper engagement by statisticians. We must embrace the technological advances, and in the process, this may open new possibilities for this to be studied in an efficient way.

8. Do you agree that patient value is enhanced by clinical statisticians recognizing the purpose of

pharmaceutical product quality, nonclinical statisticians recognizing connections of product quality attributes with clinical endpoints?

a. Having experience in both areas helps us see the bigger picture.

b. Understanding the business, what are the important questions, why it is important, allows statisticians to play a more influential role. Experience in both areas enables a broader perspective and allows for a better appreciation for the relevant questions of drug development.

c. People can pigeonhole statisticians, but seeing the end to end picture and demonstrating your knowledge of the early development data, the toxicology data, the acquisition data, when you talk like a drug developer, the attitude changes. So, it is important to have business acumen and see the whole process end to end. This has large dividends.

Questions from the audience

1. How can statisticians hired into one area acquire experience in the entire end to end drug development process?

a. The most important factor is the awareness of the importance of getting this kind of experience. Some organizations have rotation options, but the individual statistician can work with their management and ask to put it into their development plan.

b. "Challenge" or "Stretch" assignments may be an option at some companies to work on short term projects in other areas. Be proactive in finding multiple assignments that bring you new opportunities, possibly outside your comfort zone, to learn and grow. Be curious.

c. Senior Managers frequently encourage their staff to get experience on both sides. If your management does not support your interest in doing this, it might be a good time to seek other opportunities.

2. I am an isolated statistician, lacking a PhD, working for a smaller organization, with no or

few statistical colleagues, with a heavy workload. I find balancing statistical rigor with business needs a challenge. How can someone in this position handle this?

a. Being a one person department means you have to make more of an effort to demonstrate your value to the organization. Look for opportunities to get more deeply engaged. It can be simple, maybe a data presentation strategy, how to help your colleagues make their case in the best way. Start with smaller opportunities to show your value to the organization.

b. Being alone, the primary goal should be to avoid wrong interpretations of the data that could lead to incorrect decisions. Framing the data, the interpretation and risks such that a data supported decision strategy can be established should be the critical objective.

c. Understand what the decision maker's priorities are, see things from their side, discuss their thoughts and let them speak more. This can lead to good negotiation to get to a satisfactory decision.

3. How can we foster more collaboration between clinical and nonclinical at our companies?

a. Statisticians speak the same language, so the statisticians can often be the conduit to enhance and build bridges between different stakeholders. Breaking down the silos will naturally lead to more collaboration.

b. Leverage your experience and knowledge, and look for opportunities to share that knowledge across the areas.

c. Have conversations about the molecule, know the animal data, help to inform the downstream clinical implications.

d. The more each group knows about the other, the better they can appreciate their technical issues. This may not happen organically, so some effort has to be made. Pharmacology/Pharmacometrics groups can

be a good start for this kind of communication and knowledge sharing.

4. Collaborators sometimes ask that statistical results to be reduced down to a single number. But when a spectrum of results is the necessary interpretation, how do we bridge this tension between complexity and the demand for a simple answer.

a. This is where the job of the statistician to walk the collaborator to a decision is extremely important. Walk through the background to the research problem and the data collected to address the problem. Approach it as first providing the bottom line, and then follow up with how you got there. Then discuss concerns and caveats.

b. Communicating the topline result is an art, but necessary when communicating to decision makers. Takes practice. See it as an opportunity and challenge to make the right decision.

c. You're facilitating the decision; you're not making the decision. There should be a back and forth, ask what you think this means. The collaborator must have a sense of the data as well, what is their conclusion. Get their perspective. Help them make the decision you want.

4 Panelist answers led to a side discussion on Opportunities to apply "soft skills" in the evaluation of potential licensing and acquisition agreements.

a. As applied scientists, soft skills related to influence, collaboration, and engagement will further the statistician's effectiveness. Tukey said, "We play in everybody's back yard", so these soft skills are essential. Curiosity and showing interest in the collaborator's research questions is important. It's also important to avoid a skeptical attitude and to approach communications from the perspective of trying to understand.

b. The goal is making the right decisions, using the right data, the right design, paying

attention to bias and sources of variability. This ability makes the statistician's role extremely valuable and allows statisticians to engage in preparation for internal and external (regulatory) questions and dialogue.

c. Our collaborators want to do good science, our role is to help them pursue the science as partners and to enable the right decisions. It's important to remember to be helpful, think about consequences, be tactful, unbiased, and calm under fire. This is very much a fine art, honed by those who are successful in the business.

Final remarks

John Kolassa ended the session with thanks to the panelists and the audience of 51 people, with "We play in everyone's back yard, but remember, it's not our back yard. We have to understand it's their back yard, so we have to be diplomatic in how we negotiate what happens in their back yard".

References

1. S. Altan, T. Schofield, D. LeBlond (2023). A Framework to Achieving Quality in Pharmaceutical Therapeutics and Vaccine Development. *Biopharmaceutical Report*, Vol.30, No. 1 Spring 2023 <https://community.amstat.org/biop/biopharmreport>

Acknowledgements

Industry reviewers

- Don Bennett, Pfizer
- Ray Buck, Consultant
- Jaya Natarajan, Janssen

Appendix I Session abstract

This is intended as a career development session targeted to clinical and nonclinical statisticians in the pharmaceutical industry. It is expected that early in their career statisticians, as well as more experienced statisticians would benefit from this session. The session will start with a broad overview of the

nonclinical and clinical statistics applications area by way of introduction, and then enter a broad discussion of the advantages to the pharmaceutical statistician of acquiring a basic understanding of the role of nonclinical statistics in preparing them for a more insightful and productive career as a clinical biostatistician. The reverse is also true, there are distinct advantages to learning about the clinical aspects of the business, in pursuing a career in nonclinical biostatistics. This topic is especially timely today, as the industry is moving in a direction of closer linkages between the two areas. "Patient-focused drug development (PFDD)" in the clinical space and "Patient-centric" specifications in the nonclinical space, are examples of this trend. Translational medicine considerations that seek to identify biomarkers early in development is another example. The panel session will consist of experienced statisticians with broad areas of experience or responsibility in both areas who will speak to the advantages of career development benefitting from knowledge and experience from both spaces.

Appendix 2 – Time did not permit discussion of these questions.

1. Beyond effective designs, what are key learnings across the nonclinical/clinical spectrum that you have seen impact statistical innovation?
2. How does domain knowledge or business acumen that you learned across roles affect your ability to influence business partners and/or company leaders?
3. The explosion of available data from genomics, imaging, single-cell RNA, etc. present new challenges to statisticians and data scientists but differing skills necessary in the clinical vs nonclinical roles. How should statisticians best prepare for the fast-changing science and data explosion?
4. In thinking of insights and productivity between nonclinical and clinical roles, how might these be impacted by the expansion of artificial intelligence and machine learning tools? ■

REPORT ON PHARMACEUTICAL INDUSTRY SURVEY OF NONCLINICAL STATISTICS RESOURCES

NCBLF Communications Committee

Members: Stan Altan, W. Scott Clark (Chair), Katherine Giacoletti, Kristi Griffiths, John Kolassa, Katja Remlinger, Chi-Hse Teng, Donghui Zhang

Introduction

The Nonclinical Biostatistics Leaders' Forum (NCBLF) has been conducting resource surveys of Nonclinical Statistics (NCS) departments across a range of pharmaceutical companies. The purpose of these surveys is to track changes in nonclinical statistics resources within the industry, and to identify organizational trends. The intention is to help senior managers assess their organizational allocations relative to the industry averages and ranges. In addition, the surveys provide insights into the industry's current human resource landscape and support decision-making processes among companies.

History of Surveys of Nonclinical Statistics Departments

Surveys were conducted in 1997, 2008, 2015, and 2022. Over this quarter century of conducting surveys, the target population of companies changed as companies merged, and reorganized and new companies were formed. Given this background of a changing target population of companies, the surveys have revealed structural evolution of nonclinical statistics departments over time, with changes in sub-areas, and programming support varying across both time and companies. Of particular note is the most recent survey conducted in 2022, which showed that data science as a discipline has emerged as an organizational interest, in addition to traditional programming resources. We interpret this as a shift towards more sophisticated data engineering tools and approaches being brought to bear on the drug development process. Additionally, while only large pharma companies participated in the 1997 survey, the 2022 survey saw contributions from both large and mid to large-size biotech companies, highlighting the expansion of nonclinical statistical support across a wider range of companies in the industry. The findings of these surveys provide valuable insights into

the evolving landscape of nonclinical statistics in the pharmaceutical industry and can inform decision-making among companies managing nonclinical statistical resources. Summary statistics of resource allocations by survey year and nonclinical areas are given in Table 1. The number of respondents was generally half of the number polled, ranging from 13-15 companies over the period 1997-2022.

Industry Trends in Nonclinical Statistics Resources

The 2022 survey revealed that nonclinical statistics remains an important area across companies, underscoring its relevance to the drug development process. The 2022 survey indicated an increase in resources compared to the 2015 survey, mainly due to the emergence of data science within non-clinical statistics groups.

The numbers in the Discovery sub-area seem to have declined since the earlier surveys. We believe this was due to further specialization and categorization of functions. For instance, biomarkers (molecular modelling) were separated from Discovery in the 2022 survey, representing the growth in the biomarker/genomics area. It is worth noting that during the early days of the Human Genome Project (i.e., in the 2008 and 2015 surveys), genomics/biomarkers were not separate from Discovery, despite having many colleagues working in these areas. The official separation of biomarkers from Discovery in the 2022 survey reflects the growth and development of this area in the industry.

This restructuring of the Statistics functions into smaller, specialized areas creates challenges for career development, by possibly placing barriers to cross-learning and cross-fertilization. This suggests that management may have to more actively encourage collaboration and knowledge-sharing among specialized areas to ensure that team members have access to the full range of expertise and knowledge within the organization.

The Chemistry, Manufacturing, and Controls (CMC) area of nonclinical statistics resources has experienced strong growth from the earliest 2 surveys

Table 1: Median (Range) of Resources (FTE) in Nonclinical Statistics Departments by Year

Table 1 – Median (Range) of resources (FTE) in nonclinical statistics departments by year							
Year	Number Of Companies	Department					Total FTEs
		Discovery	Biomarkers	Preclinical Tox Biomarkers	CMC	Other*	
1997	15	2.5 (0-5.5)		4 (1-5)	2 (0-10.5)	1 (1-15)	9.5 (0-25)
2008	13	2 (0-30)		3 (0-10)	6 (1-27.5)	1.5 (0-5)	15 (0-60)
2015	13	2.2 (2-19)		2 (0-7)	9 (0.4-29)	-	11 (3-60)
2022	15	3 (1-24)	8 (1-11)	2 (1-4)	9 (0-30)	5 (0-25)	13.5 (2-67)

*: Other includes, programmers, data scientists, pre-clinical PK/Drug Metabolism and Efficacy, Animal PK/PD, phase I, vary from year to year.

through the 2015-2022 period, likely due to advancements in manufacturing and regulatory requirements. This growth is indicative of the increased importance of nonclinical statistics to the CMC subject matter areas in general, and the impact of new and complex dosage forms and other scientific and engineering advances. Furthermore, the operating model for non-clinical statistics (NCS) in the CMC area appears to be evolving, with a shift observed between the Discovery and CMC areas. In most companies surveyed, Discovery is still operating primarily in consultation mode, with statisticians providing advice and guidance to project teams. However, in the CMC area, it appears to be moving towards an embedded resources model where statisticians are being embedded in project teams, similar to clinical statistics.

This shift in operational model in the CMC area has important implications for the industry, particularly with regard to collaboration and knowledge-sharing. As nonclinical statisticians become more integrated into project teams, they will have the opportunity to work more closely with scientists and other team members, sharing expertise and knowledge more readily and directly, and subsequently, leading to better decision-making and more efficient product development. With the recent initiatives related to patient focused drug development and

patient centric specifications, the industry will derive increased efficiencies from closer collaborations and knowledge sharing.

The results of the 2022 survey also provide an interesting insight into the future of nonclinical statistical support in the pharmaceutical/biotechnology industry. The general sentiment among survey respondents leans towards the future expansion of nonclinical statistical support within their respective companies. This finding highlights the growing importance of nonclinical statistics in the industry, as well as the increasing demand for data-driven decision-making.

Summary and Conclusions

As the industry continues to evolve and address competitive forces, companies will seek to leverage the power of data to drive innovation, increase efficiency, and improve decision-making and outcomes. Nonclinical statistical support will play a key role in achieving these goals, providing valuable insights into the safety, efficacy, and quality of products throughout their lifecycle.

The expansion of nonclinical statistical support in pharmaceutical/biotechnology companies indicates a growing recognition of the established benefits of statistical design, modelling, and analysis in various areas such as drug discovery, safety, pre-clinical

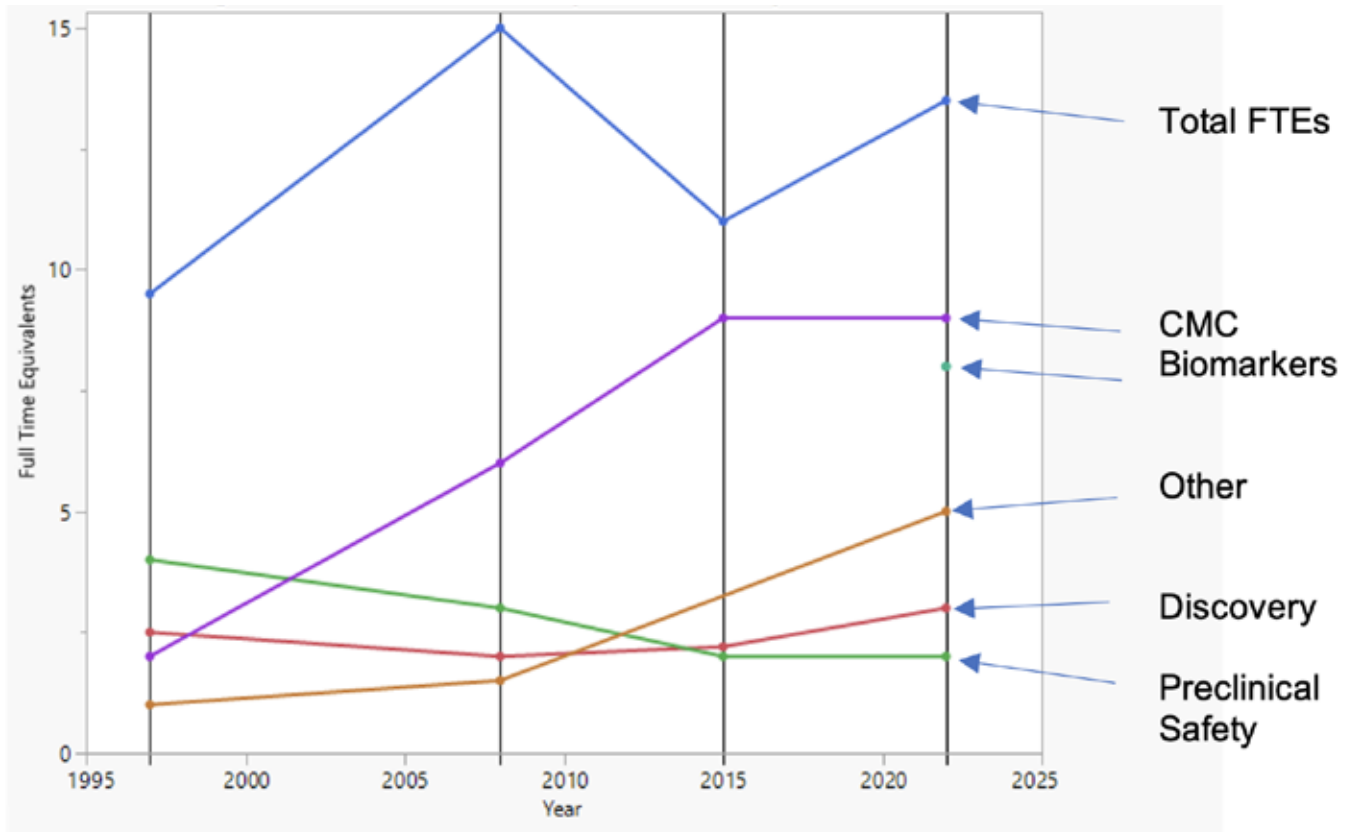


Figure 1: Median Nonclinical Department FTEs by Year

pharmacology, manufacturing, and regulatory compliance. However, this expansion of nonclinical statistical support also presents challenges, such as the need for specialized expertise and resources. The NCBLF can play an important role in helping senior leaders in nonclinical organizations address these challenges. It can provide a platform for knowledge-sharing and collaboration among industry professionals and support opportunities for professional growth and networking. The biennial Nonclinical Biostatistics (NCB) conference, and its corresponding NCS conference in Europe are examples of the active role of the NCBLF in making such opportunities possible. The NCBLF can also help to identify areas for improvement and support the development of best practices in nonclinical statistics.

In conclusion, the results of the 2022 survey conducted by the NCBLF underscore the continued importance of nonclinical statistical support in the pharmaceutical/biotechnology industry. The median (and maximum) nonclinical full-time equivalents (FTEs) increased rapidly from 1997 to 2008. The 2015 survey showed a drop of approximately 25% in median FTEs from 2008, probably due to company consolidations. The 2022 survey showed some recovery of median FTEs with a 20% increase from 2015, but still not at the 2008 level. The largest nonclinical group reported a staff of 67 FTEs in 2022. As there is a need for decision-making and effective pre-competitive collaboration, respondents indicated that they anticipate continued expansion of their nonclinical statistical capabilities to ensure that the potential benefits of statistical approaches are fully realized. ■

SUMMARY OF ASA BIOP SECTION'S VIRTUAL DISCUSSION WITH REGULATORS ON CONSIDERATION OF CRITERIA FOR EVALUATION OF SURROGATE ENDPOINTS

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

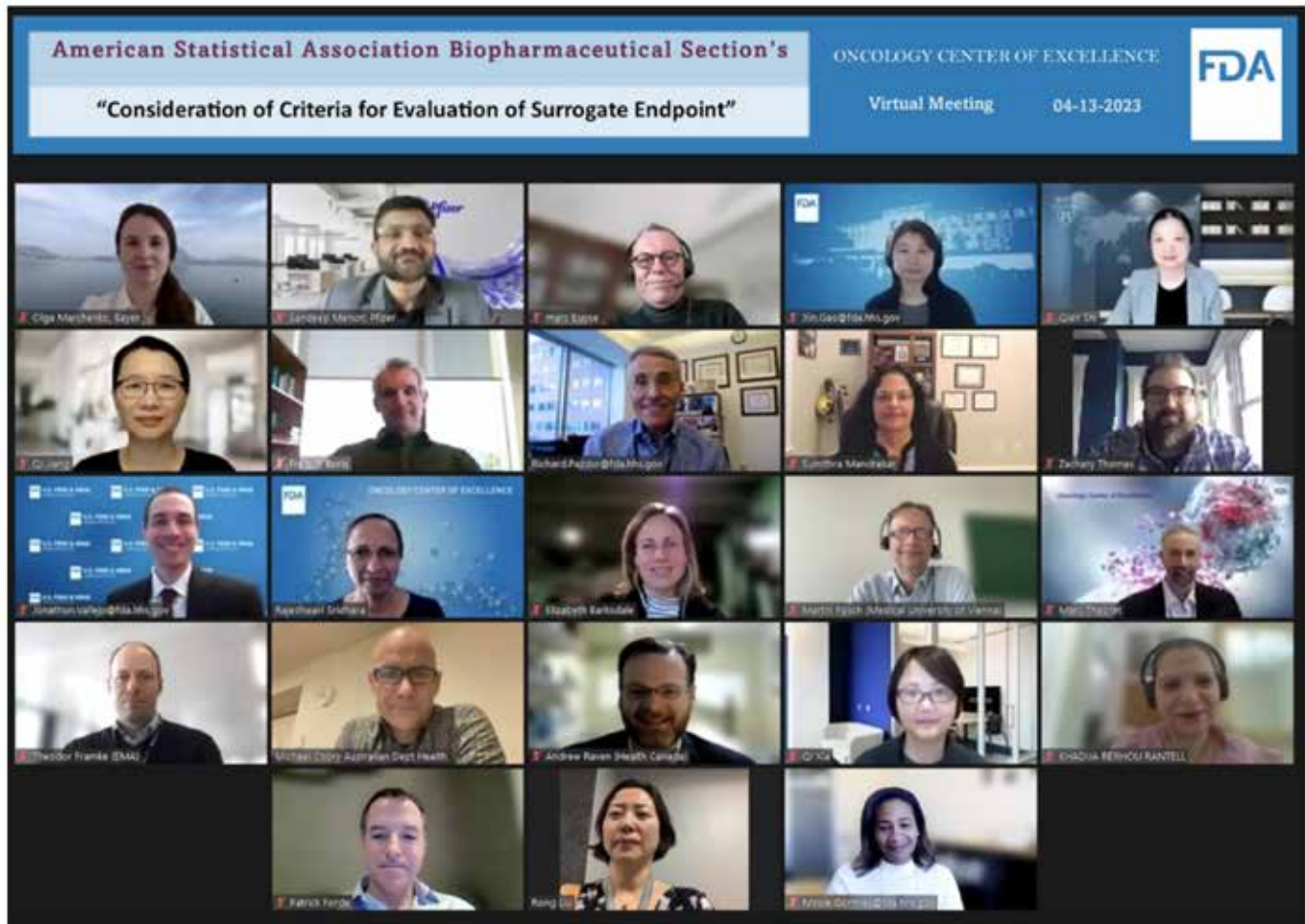
On April 13th, 2023, the American Statistical Association (ASA) Biopharmaceutical Section (BIOP) and LUNGeivity Foundation hosted a virtual forum to discuss Consideration of Criteria for Evaluation of Surrogate Endpoints. This forum was part of a series conducted under the guidance of the U.S. FDA Oncology Center of Excellence's Project SignifiCanT (Statistics in Cancer Trials). The goal of Project SignifiCanT is to advance cancer drug development through collaboration and engagement among various stakeholders in the design and analysis of cancer clinical trials. The discussion was organized jointly by the ASA BIOP Statistical Methods in Oncology Scientific Working Group, the FDA Oncology Center of Excellence (OCE), and LUNGeivity Foundation.

Individual patient data from randomized clinical trials have been used to evaluate surrogate endpoints (FDA-NIH, 2017) in cancer trials at both patient and trial levels, with some endpoints being validated as surrogates to predict clinical benefit. For example, in adjuvant colon cancer, 3-year disease-free survival (DFS) has been validated as a surrogate endpoint for 5-year overall survival (OS) (Sargent DJ et al <https://bit.ly/3sHPDYZ>), and in follicular lymphoma, 30-month complete response (CR) rate has been validated as surrogate endpoint for progression-free survival (PFS; Shi Q et al, <https://bit.ly/3QH0n1E>). There are also examples of intermediate endpoints which could not be validated as surrogate endpoints, such as pathologic complete response as a surrogate for event-free survival and overall survival in early-stage breast cancer (Cortozar P et al, <https://bit.ly/40FaqJb>, Buyse M et al, <https://bit.ly/3MN9GMv>). Shi Q et al used a pre-specified statistical analysis plan to establish the 30-month CR rate (CR30) as a surrogate endpoint by using correlation of the CR30 odds ratio with the PFS hazard ratio evaluated by both linear regression (R2WLS) and bivariate copula (R2Copula) models. Prespecified criteria for surro-

gacy required either R2WLS or R2Copula ≥ 0.80 , with a lower bound of 95% CI > 0.60 and neither estimate < 0.7 . With advances in science and technology, new biomarker based intermediate endpoints, such as minimal residual disease negativity and ctDNA measurements, are being proposed as surrogate endpoints for long term clinical benefit endpoints. This discussion among multi-disciplinary experts focused on whether the criteria used previously by Shi Q et al or a modified criteria may be needed for evaluation of future surrogate endpoints.

The speakers/panelists* for the discussion included members of the BIOP Statistical Methods in Oncology Scientific Working Group representing pharmaceutical companies, representatives from international regulatory agencies (Food and Drug Administration (FDA), Health Canada (HC), Medicines and Healthcare products Regulatory Agency (MHRA), European Medicines Agency (EMA), Therapeutic Goods Administration (TGA), and Brazilian Health Regulatory Agency (ANVISA)), clinicians, academicians, patient advocacy groups, and expert statisticians. In addition, over 100 participants attended the virtual meeting, including representatives from other international regulatory agencies (Health Sciences Authority (HAS), Singapore; Ministry of Health, Israel; Pharmaceuticals and Medical Devices Agency (PMDA), Japan). 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 LUNGeivity Foundation; and Dr. Rajeshwari Sridhara, consultant from OCE, FDA.

In the introductory presentation, the OCE leadership reviewed differences in how surrogate endpoints are used and evaluated in FDA's two approval pathways (i.e., traditional and accelerated). A validated surrogate endpoint can support marketing approval without additional studies demonstrating direct clini-



cal benefit. Correlation with the clinical outcome is necessary, but correlation by itself is not sufficient for approval, as the surrogate endpoint must also capture the net effect of treatment which is almost impossible to demonstrate. Modified conditions to capture most of the net effect have been proposed in literature. Questions posed to academic, industry, and regulatory panelists included how to select criteria for establishing trial-level associations, challenges in sharing individual patient data, and experience with different criteria used to establish surrogacy.

The first speaker, from academia, emphasized the importance of surrogate endpoints in accelerating regulatory approval for oncology treatments, while acknowledging the risks of relying on unvalidated surrogate endpoints. Surrogate endpoints must meet two fundamental requirements for validation: 1) show strong evidence as a direct causal pathway to a disease outcome, and 2) have compelling biological support. The current statistical methodology for surrogacy evaluation is

the meta-analytic approach, evaluating both individual patient-level and trial-level correlations. The importance of stringent validation criteria in surrogate endpoint analyses was highlighted due to the challenges posed by data limitations, clinical relevance, novel biomarkers, and heterogeneities in trial designs.

The second speaker, from industry, focused on the potential confounding effect in using surrogate endpoints for indolent cancers. Using pathological complete response (pCR) in early breast cancer as an example, the speaker demonstrated that pCR may not be a reliable surrogate for overall survival because of a lack of trial-level associations. The presentation concluded that addressing confounding is crucial but difficult due to the limitations of statistical methods and data availability.

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

- Establishing feasible surrogate endpoints in early-stage cancer is important to patients, clini-

cians, investigators and regulators for making new effective therapies more accessible because traditional clinical endpoints require long follow-up periods.

- Extrapolating surrogate endpoint validation from past trials to future trials may not be appropriate with changing patient populations and new therapies with different mechanism of action. Introduction of effective salvage therapies is likely to change relationship between a surrogate endpoint and survival making overall survival surrogacy challenging. While in principle the same criteria for evaluation of surrogacy may be used for evolving biomarkers, based on the patient population under study and the mechanism of action of the treatment, different criteria may need to be considered.
- Validating potential endpoints requires broad collaboration among stakeholders, large databases, and tailored approaches. Cross-company collaborations and data sharing are crucial for developing a comprehensive database to establish surrogacy.
- Discussing validation plans with regulators before performing analyses is important. Preparatory work, such as harmonizing definitions across trials, is crucial to the interpretability of results.

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

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

*** Speakers/ Panelists:** Dr. Elizabeth Barksdale (LUNGevity Foundation), Dr. Marc Buyse (International Drug Development Institute), Dr. Michael Coory (TGA, AU), Dr. Leonardo Filho (ANVIS, BR), Dr. Patrick Forde (Johns Hopkins Medicine), Dr. Theodor Framke

(EMA), Dr. Boris Freidlin (NCI/ NIH), Dr. Xin (Cindy) Gao (FDA), Dr. Nicole Gormley (FDA), Dr. Qi Jiang (Seagen), Dr. Rong Liu (BMS), Dr. Sumithra Mandrekar (Mayo Clinic), Dr. Olga Marchenko (Bayer), Dr. Sandeep Menon (Pfizer), Dr. Richard Pazdur (FDA), Prof. Martin Posch (Center for Medical Statistics, Informatics, and Intelligent Systems at the Medical University of Vienna), Dr. Khadija Rerhou Rantell (MHRA, UK), Mr. Andrew Raven (Health Canada), Prof. Qian Shi (Mayo Clinic), Dr. Rajeshwari Sridhara (FDA), Dr. Marc Theoret (FDA), Dr. Zachary Thomas (Eli Lilly), Dr. Jonathon Vallejo (FDA), Dr. Qi Xia (AbbVie).

References:

- FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource. Validated Surrogate Endpoint (2017): <https://www.ncbi.nlm.nih.gov/books/NBK453484/#:~:text=Definition,predicts%20a%20specific%20clinical%20benefit>
- Buyse M, Burzykowski T, Saad ED. Neoadjuvant as Future for Drug Development in Breast Cancer—Letter. *Clin Cancer Res* (2016) 22 (1): 268.
- Cortazar P, Zhang L, Untch M, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis, *Lancet* 2014; 384:164–72.
- Sargent DJ, Wieand HS, Haller DG, et al. Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials. *Journal of Clinical Oncology* 2005; 23: 8664–70.
- Shi Q, Flowers CR, Hiddemann W, et al. Thirty-Month Complete Response as a Surrogate End Point in First-Line Follicular Lymphoma Therapy: An Individual Patient-Level Analysis of Multiple Randomized Trials. *Journal of Clinical Oncology* 2017; 35:552-560. ■

SUMMARY OF ASA BIOP SECTION'S VIRTUAL DISCUSSION WITH REGULATORS ON CANCER CLINICAL TRIAL DESIGN AND ANALYSIS CONSIDERATIONS IN EVALUATING TREATMENT EFFECT IN MARKER NEGATIVE POPULATION – PART 2

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

On May 11th, 2023, the American Statistical Association (ASA) Biopharmaceutical Section (BIOP) and LUNGeivity Foundation hosted a virtual forum to discuss *Cancer Clinical Trial Design and Analysis Considerations in Evaluating Treatment Effect in Marker Negative Population*. This forum was part of a series conducted under the guidance of the U.S. FDA Oncology Center of Excellence's 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 LUNGeivity Foundation. This was a continuation of the discussion held in December of 2021.

Advances in precision medicine and development of molecularly targeted therapies have led to many effective therapies for targeted populations in the past 10 years. However, many randomized clinical trials evaluating molecularly directed therapies are conducted in the overall population, which includes both marker positive and negative populations. Indeed, it is not uncommon that the hypothesis is tested in the overall population first and if the treatment effect is found to be statistically significant, then the hypothesis is tested either as a preplanned or exploratory analysis in the marker positive subgroup. In general, hypothesis testing in a marker negative subgroup is not prespecified and conducted only as an exploratory analysis. If the treatment effect in the marker positive subgroup is large and contributes to a significant effect in the overall population, it may be challenging to infer treatment effect in the marker negative subgroup. There are examples of

cancer drug approvals indicated in both restricted and overall populations stemming from this type of situation.

This open forum discussion among multi-disciplinary experts focused on understanding and measuring uncertainties in the evaluation of treatment effect in the marker negative subgroup. The speakers/panelists* for the discussion included members of the BIOP Statistical Methods in Oncology Scientific Working Group representing pharmaceutical companies, representatives from regulatory agencies (Food and Drug Administration (FDA), Health Canada (HC), Medicines and Healthcare products Regulatory Agency (MHRA), and Therapeutic Goods Administration (TGA)), clinicians, academicians, patient advocacy groups, and expert statisticians in industry. In addition, over 100 participants attended the virtual meeting, including representatives from other international regulatory agencies (European Medicines Agency (EMA), Pharmaceuticals and Medical Devices Agency (PMDA), Health Sciences Authority (HAS), Israel Ministry of Health (MOH)). The discussions were moderated by the BIOP Statistical Methods in Oncology Scientific Working Group co-chairs, Dr. Olga Marchenko from Bayer and Dr. Qi Jiang from Seagen; and Dr. Rajeshwari Sridhara, consultant from OCE, FDA.

In the introductory presentation, the OCE leadership reviewed design and analysis considerations in evaluating treatment effects in marker negative populations in cancer clinical trials. It was noted that most phase III randomized clinical trial (RCT) designs that evaluate biomarker effects include testing in overall population and biomarker-positive subpopulation, but not in biomarker negative subpopulation. Examples were given where indications were granted to the overall



population, and cases where approval was restricted to a subpopulation. Panelists were asked to consider how to set up subgroup hypothesis testing to avoid false conclusions, quantify uncertainty in subgroup analyses, measure the strength of association between the biomarker and the treatment, and when to include testing in the biomarker negative subpopulation.

The first speaker, from academia, critiqued the prevalent use of biomarker positive/overall (BM+/O) design in randomized controlled trials. Problems with BM+/O design have been discussed in literature (Rothmann et al. *Drug Inf J.* 2012; Freidlin et al. *Nat Rev C Onc* 2014; Tannock et al. *Ann Onc* 2020; Kim et al. *EJC* 2021). Using several trials (e.g., KEYNOTE-119, KEYNOTE-048, CHECKMATE-648, BELLE-2) as illustrative examples, the speaker pointed out that the BM+/O design may lead to problematic recommendations for

biomarker negative (BM-) subgroups. The speaker recommended using biomarker-stratified designs and reporting the treatment effects in BM- subgroups to inform the choice of treatment for individual patients.

The second speaker, from industry, highlighted the imperfect nature of biomarkers and the value of broad population experience in Phase 3 trials when biomarker effects are uncertain. The presenter also discussed the benefits of testing multiple hypotheses by taking advantage of correlated hypotheses and emphasized the importance of considering treatment effect differences between BM+ and BM- subpopulations during trial design (Anderson et al., 2022).

The third speaker, from academia, pointed out that under the assumption that the treatment effect in BM+ subpopulation is the same or larger than in BM- subpopulation, trials can test the hypothesis in both BM+

and BM- subpopulations more efficiently. The type I error rate will be inflated if the assumption is violated. The speaker also explored different trial designs, including targeted and biomarker-stratified approaches, and their implications for sample size and error control.

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

- While p-values from subgroup analyses may not be meaningful due to lack of power, calculating posterior probabilities of survival improvement for biomarker groups can aid decision making.
- Pre-planning and setting up the analysis at design stage is crucial. If the treatment is an add-on, a pre-planned analysis for the BM- subpopulation should be included due to the potential for increased toxicity.
- Assuming that the prevalence rate for BM+ is higher than the prevalence rate for BM- and the treatment effect in BM- is smaller than that in BM+ subpopulation, studies evaluating biomarker effects should be powered within the BM+ subpopulation but also should quantify treatment effects within the BM- subpopulation using a confidence interval with some relaxed level of testing. It is important not to exclude subjects who are BM- from being evaluated for the treatment effect, if there is uncertainty regarding drug's target population.
- Subgroup hypothesis testing should be set up using well-established strategies to control the overall type I error rate as a result of multiple testing. If a treatment effect is possible for the BM- subpopulation, continued enrollment to collect data for testing within the BM- subpopulation is important.
- The approach to testing depends on the nature of the biomarker. Different biomarkers may interact with each other, adding complexity to identifying treatment effect within biomarker-defined subpopulations.
- The decision to grant approval in the overall population or to restrict to BM+ subpopulation is made on a case-by-case basis. The decision depends on scientific rationale, reliability of the biomarker, and whether the treatment effect

in BM+ population drives overall results. The decision on approval goes beyond statistics and involves various considerations such as study design, endpoints, maturity of data, biomarker prevalence and unmet needs.

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

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

*** Speakers/ Panelists:** Dr. Anup Amatya (FDA), Dr. Keaven Anderson (Merck Research Laboratories), Dr. Frank Bretz (Novartis), Dr. Michael Coory (TGA, AU), Dr. Lola Fashoyin-Aje (FDA), Dr. Leonardo Filho (ANVISA, BR), Dr. Boris Freidlin (NCI/NIH), Prof. Anastasia Ivanova (University North Carolina, Chapel Hill), Dr. Qi Jiang (Seagen), Dr. Erin Larkins (FDA), Dr. Nicole Li (Beigene), Dr. Olga Marchenko (Bayer), Dr. Richard Pazdur (FDA), Dr. Khadija Rerhou Rantell (MHRA, UK), Mr. Andrew Raven (Health Canada), Prof. Mary Redman (Fred Hutch Cancer Center), Dr. Sunhee Ro (Sierra Oncology), Dr. Rajeshwari Sridhara (FDA), Dr. Daniel Suzman (FDA), Dr. Marc Theoret (FDA), Dr. Qing Xu (FDA), Prof. Ying Yuan (MD Anderson Cancer Center), Dr. Weidong Zhang (Sana Biotechnology).

References:

- Rothmann MD, et al. *Drug Inf J*, 46(2):175-179, 2012
- Freidlin B, Korn EL. *Nat Rev Clin Oncol*, 11(2):81-90, 2014
- Tannock IF, Templeton AJ. *Ann Oncol* 31(3):331-333, 2020
- Kim MS, Prasad V. *Eur J Cancer* 155:163-167, 2021
- Anderson, Keaven M, Zifang Guo, Jing Zhao, and Linda Z Sun. 2022. "A Unified Framework for Weighted Parametric Group Sequential Design." *Biometrical Journal* 64 (7): 1219–39 ■

REPORT FROM THE CELL AND GENE THERAPY SCIENTIFIC WORKING GROUP

Alan Y Chiang (Lyell Immunopharma), Daniel Li (Bristol Myers Squibb), Zhenzhen Xu¹ (FDA)

Introduction

According to the American Society of Gene and Cell Therapy, gene therapy is the use of genetic material to treat or prevent a disease, and cell therapy is the transfer of a specific cell type(s) into a patient to treat or prevent a disease. These definitions are very broad and can cover a wide range of medical products. For example, cell therapy may include products that are stem or progenitor cell derived, mature or functionally differentiated cell derived, or tissue engineering based; gene therapy can include products that incorporate viral or nonviral vectors, microbial vectors, oncolytic viruses, or ex vivo genetically modified cells. In recent years, the research and development of cell and gene therapy (CGT) products have been making rapid progress, offering potential effective treatment options for various serious diseases and damaged tissues or organs. Many of the initial clinical outcomes have led to considerable investment in such innovative therapies around the world. A recent survey by Alliance for Regenerative Medicine (alliancerm.org) shows that there were approximately 2,200 ongoing Phase 1-3 CGT clinical trials worldwide in December 2022. The early stages of therapeutic development of CGT focused on treating serious and life-threatening diseases, such as cancers, genetic diseases, severe burns, and infectious diseases. With the recent advancement of science and technology, therapeutic applications of CGT have expanded to treating patients with arthritis, lupus, neurological disorder, diabetes, cutaneous ulcer, and various chronic diseases. CGT products may require laboratory procedures or surgical operations, which can be invasive for delivery to the target site.

The world's first commercial gene therapy, Gen-

dicine, was approved in 2003 by China's State Food and Drug Administration to treat head and neck squamous cell carcinoma (Pearson et al., 2004). Gendicine utilizes recombinant adenoviral vectors to deliver wild-type p53 gene, the expression of which is known to transfer antitumor abilities into the nucleus of tumor cells. The approval and subsequent clinical experience helped pave a path for success in making the technology become more accepted (Zhang et al., 2018). In 2012, the first commercial gene therapy product Glybera was approved in Europe for the treatment of lipoprotein lipase (LPL) deficiency, an ultra-rare genetic disorder (Kassim and Somerville, 2013). Glybera employs adeno-associated virus as a vector to convey a functional replica of the LPL gene to skeletal muscle. In 2017, the manufacturer of Glybera announced it would not seek renewal of the European Union market authorization and the product was subsequently withdrawn from the market. However, 2017 also marked two important milestones for CGT development in the United States. Kymriah was the first genetically modified cell-based gene therapy to receive approval from the U.S. FDA for any indication (FDA, 2017; Braendstrup et al., 2020). Yescarta was also approved later in 2017 for patients with large-B-cell lymphomas. Both Kymriah and Yescarta are autologous Chimeric Antigen Receptor (CAR) T cell therapies that are tailored to treat each individual patient. Immune cells are extracted from the patient's body, genetically modified in a lab to target specific cancer cells, and then reintroduced back into the patient's body through infusion. As of June 1, 2023, there have been 29 CGT products approved by the FDA (2023a), including 6 CAR T therapies (Table 1), and an autologous active cellular immunotherapy

¹ **Disclaimer:** The findings and conclusions in this article have not been formally disseminated by the Food and Drug Administration and should not be construed to represent any Agency determination or policy.

often referred to as a vaccine, Provenge, approved in 2010 for the treatment of asymptomatic or minimally symptomatic metastatic castration-resistant prostate cancer (Cheever and Higano, 2011).

CAR T Cell Therapy

CAR T cell therapy is a human gene therapy product where T cells are genetically modified to enable them to recognize desired target antigen(s) more effectively. In oncology, it is an emerging form of cancer immunotherapy, which involves supercharging a patient's T cells to recognize and attack certain cancer cells. Administering and manufacturing CAR T cell therapy products are a complex process that can take a few weeks. The steps generally include collecting the T cells from the patient or healthy donors, engineering the T cells in a laboratory by adding a manufactured CAR and allowing the CAR T cells to multiply and grow, and infusing the CAR T cells to the patient once the laboratory or manufacturing facility has enough CAR T cells. Lymphodepletion chemotherapy is commonly used prior to the product infusion to help increase the treatment's effectiveness. See Figure 1 for an illustration of autologous CAR T therapies patients' journey.

FDA has issued several guidance documents on the development of CGT therapies, including CAR T cell products (FDA, 2023b). These guidance documents provide important information on the chemistry, manufacturing, and control (CMC), preclinical, and clinical study design considerations of CAR T cell therapy development. The CMC guidance provides recommendations on the following topics:

- Vector manufacturing and testing
- Collection, handling and testing of starting material
- Manufacturing process and analytical testing
- Manufacturing changes and assessing comparability
- Single-site or multisite manufacturing

The preclinical guidance provides recommendations on the following topics:

- Design of vector component and transgene

delivery process

- Characterization of transduced cell expression
- In vivo testing

The clinical study of guidance provides information on the following topics:

- Considerations for study population selection
- Treatment plan, including dose finding and situations when there is a manufacturing delay or failure
- Pharmacokinetics, pharmacodynamics, and immunogenicity
- Safety evaluation and monitoring
- Long term follow-up plan
- Additional considerations for allogeneic CAR T products

In addition to the FDA's guidance for industry, there are also several resources available to sponsors engaged in CAR T cell product development. They are available through the web site of:

- The International Society for Cell and Gene Therapy (ISCT)
- The American Society of Gene and Cell Therapy (ASGCT)
- The National Cancer Institute (NCI)
- The National Institutes of Health (NIH)

The development of CAR T cell products is a complex and challenging process. The FDA's guidance documents and other resources can help sponsors develop safe and effective products that have the potential to revolutionize the treatment of cancer and other diseases. Several statistical related issues have emerged and methodological solutions to address these unique challenges have become an area of active research.

Cell and Gene Therapy Scientific

Working Group

In late 2022, a statistics-focused scientific working group (SWG) was formed to address some of the critical and unique questions arisen from CGT development, and in January 2023, the SWG was subsequently endorsed and supported by the Biopharmaceutical Section. The proposed SWG works to strengthen the use of appropriate analytics by sharing collective experience, facilitating the development of promising statistical methods, and promoting the adoption of those methods by the statistical community. The initial efforts focus on issues related to genetically modified autologous cell-based gene therapies. The term modified cell therapies includes a variety of immune therapies, such as CAR, including CAR T cells and CAR Natural Killer (NK) cells, or T cell receptor (TCR) based, tumor infiltrating lymphocytes, and other adaptive immune cell-based therapies. Current members include representatives from pharmaceutical and biotech companies, contract research organizations, and health authorities.

Statistical issues that are pertinent to CGT development include the followings:

- Design of multiple versions of CGT in early or late phase clinical trial
- Determination of dose-range, and approaches for dose escalation and dose optimization
- Implementation of estimand framework appropriate to CGT special considerations
- Use of real-world data and real-world evidence (RWD/RWE) to support or supplement CGT product registration or reimbursement
- Design of de-centralized long-term follow-up study (LTFU)
- Prediction of CGT related adverse events such as cytokine release syndrome and neurotoxicity or efficacy endpoints
- Approaches to establish CMC critical quality attributes and justifications of specification
- Performing comparability analysis for multiple versions of CGT

- Addressing statistical issues related to clinical trial design of next generation CGT
- Addressing the issue of small sample size in rare disease trials

While some of the topics are also relevant to other therapeutic applications, Table 2 summarizes the features of statistical challenges and opportunities distinctive to CGT development.

Summary

CGT's are at the forefront of therapeutic innovation and transform how we treat and potentially cure certain diseases. The new era of CGT development shares a profile of astonishing efficacy, complex production and manufacturing process, and unique statistical challenges, aimed at very limited patient populations. In the spirit of scientific collaboration, the CGT SWG was formed to collectively address and overcome some of these challenges. Currently the SWG core team members include representatives from BMS (Daniel Li and Revathi Ananthakrishnan), FDA (Zhenzhen Xu), ICON (Patricia Anderson), Kite (Jim Whitmore), Lyell (Alan Chiang and Yeonhee Kim), MHRA (Khadija Rantell), and Novartis (Shihua Wen). As our understanding of CGT continues to improve, the SWG has planned to have several scientific disclosures in 2023 and beyond. It is anticipated that there are opportunities to collaborate with other statisticians and SWGs within the Biopharmaceutical Section. The CGT SWG is also actively recruiting statistical experts and subteam members to help tackle some of these difficult problems and identify new opportunities. For more information, please see the CGT SWG website at <https://community.amstat.org/biop/workinggroups/cellandgenetherapy>.

References

- Pearson S, Jia H, Kandachi K. (2004). China approves first gene therapy. *Nat Biotechnol*, 22, pp. 3-4.
- Zhang W-W, Li L, Li D, Liu J, Li X, Li W, Xu X, Zhang MJ, Chandler LA, Lin H, Hu A, Xu W, Lam DM. (2018). The first approved gene therapy product for cancer Ad-p53 (Gendicine): 12 years in the clinic. *Hum Gene Ther*, 29, pp. 160-179.

- Kassim SH, Somerville RPT. (2013). From lipoprotein to chondrocytes: a brief summary of the European Medicines Agency's regulatory guidelines for advanced therapy medicinal products. *Hum Gene Ther*, 24, pp. 568-570.
- FDA News Release. (2017). <https://www.fda.gov/news-events/press-announcements/fda-approval-brings-first-gene-therapy-united-states>. Accessed June 9, 2023.
- Braendstrup P, Levine BL, Ruella M. (2020). The long road to the first FDA-approved gene therapy: chimeric antigen receptor T cells targeting CD19. *Cytotherapy*, 22, pp. 57-69.
- FDA Approved Cellular and Gene Therapy Products. (2023a). <https://www.fda.gov/vaccines-blood-biologics/cellular-gene-therapy-products/approved-cellular-and-gene-therapy-products>. Accessed June 9, 2023.
- Cheever MA, Higano CS. (2011). Provenge (sipuleucel-T) in prostate cancer: the first FDA-approved therapeutic cancer vaccine. *Clin Cancer Res*, 17, pp. 3520-3526.
- FDA Cellular & Gene Therapy Guidances. (2023b). <https://www.fda.gov/vaccines-blood-biologics/biologics-guidances/cellular-gene-therapy-guidances>. Accessed June 9, 2023.
- Li DH, Whitmore JB, Guo W, Ji Y. (2017). Toxicity and efficacy probability interval design for phase I adaptive cell therapy dose-finding clinical trials. *Clin Cancer Res*, 23, pp. 13-20.
- Lin R, Zhou Y, Yan F, Li D, Yuan Y. (2020). BOIN12: Bayesian optimal interval phase I/II trial design for utility-based dose finding in immunotherapy and targeted therapies. *JCO Precision Oncology*, 4, pp. 1393-1402.
- Devlin SM, Iasonos A, O'Quigley J. (2021). Phase I clinical trials in adaptive T-cell therapies. *J R Stat Soc Series C*, 20, pp. 815-834.
- Better M, Fuchs M, Gidwani S, Goessl C, Healy J, Jadowsky J, Jazayeri J, Kalos M, Liu W, Pearce L, Pilon-Thomas S, Snehal N, Veldman M, Vong J, Weinbach S, Yohrling J. (2023). Accelerating the development of engineered cellular therapies: a framework for extrapolating data across related products. Friends of Cancer Research, https://friendsofcancerresearch.org/wp-content/uploads/Accelerating_The_Development_of_Engineered_Cellular_Therapies.pdf. Accessed June 9, 2023.
- Lin X, Lee S, Sharma P, George B, Scott J. (2022). Summary of US Food and Drug Administration chimeric antigen receptor T-cell biologics license application approvals from a statistical perspective. *J Clin Onc*, 40, pp. 3501-3509.
- Casadei B, Argnani L, Guadagnuolo S, Pellegrini C, Stefoni V, Broccoli A, Nanni L, Morigi A, Lolli G, Guarino M, Spinardi L, Pierucci E, Fanti S, Bartoletti M, Dicaldo M, Sabattini E, Bonifazi F, Zinzani PL. (2021). Real world evidence of CAR T-cell therapies for the treatment of relapsed/refractory B-cell non-Hodgkin lymphoma: a monocentric experience. *Cancers*, 13, pp. 4789.
- Derman BA, Belli AJ, Battiwalla M, Hamadani M, Kansagra A, Lazarus HM, Wang C-K. (2022). Reality check: real-world evidence to support therapeutic development in hematologic malignancies. *Blood Rev*, 53, pp. 100913.
- Swanson C, Do T, Merrigan S, Lonning S, Merriam K, Prentiss J, Sridar J, Sutherland CL, Xie B, Pham M, Li D, Albertson T, Stern HM, Moulis S. (2017). Predicting clinical response and safety of JCAR017 in B-NHL patients: potential importance of tumor microenvironment biomarkers and CAR T-cell tumor infiltration. *Blood*, 130, pp. 194.

Stein-Thoeringer CK, Saini NY, Zamir E, Blumenberg V, Schubert ML, Mor U, Fante MA, Schmidt S, Hayase E, Hayase T, Rohrbach R, Chang CC, McDaniel L, Flores I, Gaiser R, Edinger M, Wolff D, Heidenreich M, Strati P, Nair R, Chihara D, Fayad LE, Ahmed S, Iyer SP, Steiner RE, Jain P, Nastoupil LJ, Westin J, Arora R, Wang ML, Turner J, Menges M, Hidalgo-Vargas M, Reid K, Dreger P, Schmitt A, Müller-Tidow C, Locke FL, Davila ML, Champlin RE, Flowers CR, Shpall EJ, Poeck H, Neelapu SS, Schmitt M, Subklewe M, Jain MD, Jenq RR, Elinav E. (2023). A non-antibiotic-disrupted gut microbiome is associated with clinical responses to CD19-CAR-T cell cancer immunotherapy. *Nat Med*, 29, pp. 906-916.

FDA Guidance for Industry. (2020). Long term follow-up after administration of human gene therapy products. <https://www.fda.gov/media/113768/download>. Accessed June 9, 2023.

FDA Guidance for Industry. (2022). Studying multiple versions of a cellular or gene therapy product in an early-phase clinical trial. <https://www.fda.gov/media/152536/download>. Accessed June 9, 2023.

Rochigneux P, Chanez B, De Rauglaudre B, Mitry E, Chabannon C, Gilibert M. (2021). Adoptive cell therapy in hepatocellular carcinoma: biological rationale and first results in early phase clinical trials. *Cancers*, 13, pp. 271.

Table 1. FDA approved CAR T-cell therapies

Generic Name	Brand Name	Target Antigen	Year of First Approval	Targeted Disease	Patient Population
Tisagenlecleucel	Kymriah	CD19	2017	B-cell acute lymphoblastic leukemia (ALL)	Children and young adults with relapsed or refractory B-cell ALL
			2018	B-cell non-Hodgkin lymphoma (NHL)	Adults with relapsed or refractory B-cell NHL
			2022	Follicular lymphoma (FL)	Adults with relapsed or refractory FL
Axicabtagene ciloleucel	Yescarta	CD19	2017	B-cell non-Hodgkin lymphoma (NHL)	Adults with relapsed or refractory B-cell NHL
			2021	Follicular lymphoma (FL)	Adults with relapsed or refractory FL
Brexucabtagene autoleucel	Tecartus	CD19	2020	Mantel cell lymphoma (MCL)	Adults with relapsed or refractory MCL
			2021	B-cell acute lymphoblastic leukemia (ALL)	Adults with relapsed or refractory B-cell ALL
Lisocabtagene maraleucel	Breyanzi	CD19	2021	B-cell non-Hodgkin lymphoma (NHL)	Adults with relapsed or refractory B-cell NHL
Idecabtagene vicleucel	Abecma	BCMA	2021	Multiple myeloma (MM)	Adults with relapsed or refractory MM
Ciltacabtagene autoleucelw	Carvykti	BCMA	2022	Multiple myeloma (MM)	Adults with relapsed or refractory MM

Table 2. Statistical challenges in CGT development

Key Areas of Focus	Unique Challenges	References
Dose Escalation	<ol style="list-style-type: none"> 1. Traditional maximum tolerated dose may not be suitable for these therapies; approaches to address toxicity-efficacy trade-off during dose escalation can help optimize the recommended Phase 2 dose. 2. Leveraging data extrapolation from prior products of identical CAR construct with similar features to help accelerate dose escalation. 3. Manufacturing autologous products may fail to achieve the desired dose level or meet the specified release criteria during ex vivo expansion of the T-cell, resulting non-conforming products in dose escalation. 	Li et al. (2017); Lin et al. (2020); Devlin et al. (2021); Better et al. (2023)
Estimand	<ol style="list-style-type: none"> 1. Clear understanding of scientific questions and treatment of interest. 2. Intercurrent events occurring after surgery or leukapheresis, such as the use of lymphodepleting therapies and/or bridging therapies. 3. Intercurrent events due to manufacturing failure or starting new therapies without progression. 	Lin et al. (2022)
RWD/RWE	Real world data (e.g., data sources from clinical sites, registries and research databases) that are derived and set up to generate an external comparison arm for ancillary analysis of efficacy endpoints.	Casadei et al. (2021); Derman et al. (2022)
Predictive Biomarkers	Identification of biomarkers, including product cell composition, tumor burden, T cell infiltration, and immunosuppressive factor expression, that may be associated with or predictive of CART safety and efficacy.	Swanson et al., (2017); Stein-Thoeringer et al. (2023)
Master Protocol; Novel Trial Design	<ol style="list-style-type: none"> 1. Utilization of a master protocol design to conduct CGT LTFU studies 2. Studying multiple versions of CGT product in early phase using umbrella trial designs 3. Basket trial designs implemented in targets with lower prevalence cancer subtypes 	FDA LTFU (2020); FDA multiple versions of CGT (2022); Rochigneux et al. (2021)

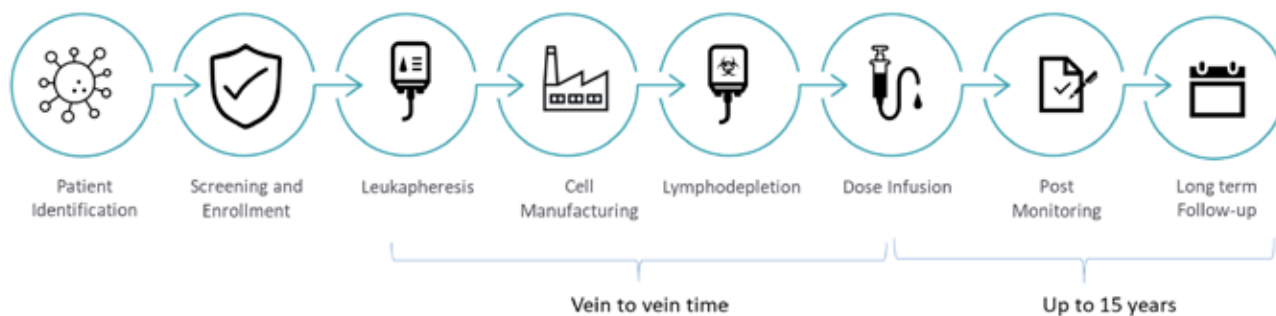


Figure 1. A patient journey from an autologous CART cell therapy treatment

IMPORTANT INFORMATION FOR PROSPECTIVE ASA FELLOWS

Original written by Paul Gallo, Brenda J Crowe, Bruce Binkowitz, Ilya Lipkovich, and Amy Xia, and updated by Paul Gallo, Bruce Binkowitz, Weili He, Inna Perevozkaya

September 2023

Selection as a Fellow of the American Statistical Association is a high honor to which many members of the ASA aspire. Each year, new Fellows are chosen based upon their record of achievements and contributions to the field, summarized in nomination packages submitted to the ASA Committee on Fellows, announced in the spring, and recognized in ceremonies at the Joint Statistical Meetings. Biopharmaceutical Section (BIOP) members have been well represented among those honored. To help BIOP members who are considering being put forth for selection or would like to assist others in achieving this honor, a Fellows Nomination Committee has been operating within the section. Its members are ASA Fellows experienced in successfully supporting others in the nomination process. The committee does not prepare packages for potential nominees, but can offer general advice and, importantly, send a proposed nomination package to an independent expert reviewer for comments and suggestions for improvements. Nominators who would like to take advantage of this service should send their draft packages to the current committee chair (information can be found at <https://community.amstat.org/biop/aboutus/sub-committees/fellows141>) at least 4-5 weeks in advance of the planned submission date in order for feedback to be received and potentially

acted upon (the submission deadline each year is March 1). Those planning a nomination should thoroughly familiarize themselves with the process, along with suggestions for an effective nomination. There are a few good sources of information readily available to prospective candidates and nominators and can be found below:

- ASA Fellows website (<https://www.amstat.org/your-career/awards/asa-fellows>) contains tips for Nominators, frequently asked questions, rating of nominees, nomination form preview, example of letters of support, and other useful information.
- In addition, a helpful article with perspectives and tips from several BIOP ASA Fellows, entitled “Nomination for ASA Fellowship” (Dmitrienko et al), appeared in the Spring 2020 Biopharmaceutical Report: *BIOPSpring2020_FINAL.pdf*.
- Finally, an ASA-sponsored webinar was presented in 2020, “Biopharmaceutical Section Offers Advice on Strategic Planning for ASA Fellow Nomination”, containing presentations and panel discussions featuring a large group of BIOP members with experience in the Fellows process, and can be viewed at https://www.youtube.com/watch?v=YLkXund_p7I.

Good luck!

CHECK OUT THE LATEST EPISODES OF THE ASA BIOPHARMACEUTICAL SECTION PODCAST



Christina Nurse (Takeda), Rebecca Wilson (J&J)

It's been an exciting year for the ASA Biopharmaceutical Section Podcast. We welcomed a new co-host, Rebecca Wilson, earlier this year to help while Christina Nurse welcomed her son, Kaleb. Speaking of family additions, Amy Lalonde also welcomed her son over the summer. As we like to cover a variety of topics on the podcast, Amy and Christina will discuss what it's like to be a working mom and statistician in the industry in an upcoming episode.

Please visit the ASA website to listen to all episodes: <https://community.amstat.org/biop/media-contents/podcasts>

Upcoming episodes include a conversation with Godwin Yung, PhD (Genentech/Roche) on overall

survival in oncology trials. Cesar Torres, PhD (FDA) will share his perspective on safety estimands in clinical trials. Satrajit Roychoudhury, PhD (Pfizer) will highlight the work of the ASA Statistical Partnerships Among Academe, Industry & Government (SPAIG) committee. He will also discuss the ASA SPAIG award that recognizes outstanding partnership or collaborative efforts across different career sectors.

We look forward to having you tune into these episodes. If you have suggestions for a topic or want to be a guest, please reach out to Christina Nurse, PhD (christina.nurse@takeda.com), Amy Lalonde, PhD (lalonde_amy@lilly.com), and Rebecca Wilson, DrPH (rwilso12@its.jnj.com).



A SUCCESSFUL CONFERENCE SEASON

Brian Millen, Chair, ASA Biopharmaceutical Section. (Biogen)

We've recently wrapped up what has become known in my house as conference season – that most busy and energizing time of the year spanning from JSM through RISW. Fittingly for a time that includes kids' returning to school and launches of academic years, I enjoy the opportunity to reunite with friends and colleagues, to share, and learn during this season. This year, both JSM and RISW provided great opportunities for learning and reflected the continued growth in interest from members of our Biopharm Section (BIOP). I share a few highlights below.

Elena Polverejan (J&J), 2023 BIOP Program Chair, led the selection of BIOP sessions for JSM. This year, BIOP sponsored 6 invited sessions. This includes our 4 allotted sessions plus two additional sessions earned through competition. In addition to the invited sessions, BIOP sponsored 16 topic-contributed sessions, which were selected out of 49 proposals. BIOP also





sponsored 20 contributed paper sessions, 48 contributed poster presentations and 19 contributed speed presentations. This program of presentations, talks, and panels offered an abundance of opportunity for colleagues to share their work and all of us to engage on topics of interest. The short course, “Causal inference in Randomized Controlled Trials,” rounded out the scientific offerings by BIOP at JSM.

Of course, the annual BIOP business meeting and mixer is a highlight of the conference for many of us. This event offers a chance to network, reconnect with friends and make new ones. This is all done, of course, with a backdrop of fine appetizers and drinks, and BIOP officers sharing updates on behalf of the executive committee. This year, as Chair, I had the privilege of presenting multiple awards to deserving recipients, including our student scholarship awards which grew to eight this year, due to the large

“This year, as Chair, I had the privilege of presenting multiple awards to deserving recipients, including our student scholarship awards which grew to eight this year...”

numbers of really deserving high-quality applicants. (Thank you to the Scholarship Award Committee led by Jared Lunceford (Merck)).



Just a few weeks after returning from Toronto, I landed in Washington, D.C., to participate in the BIOP Regulatory Industry Statistics Workshop, affectionately known as RISW. This conference is unique in its very relevant focus for our BIOP community as well as its strong participation from statistician colleagues at FDA and industry alike. Conference Co-Chairs, Fanni Natanegara (Lilly Japan) and Erik Bloomquist (Merck; formerly at FDA) and the 2023 Steering Committee put on a stellar program. The global theme of the conference, Statistical Thinking and Innovation with Global Impact, was felt prominently in the plenary sessions and throughout the conference. In addition to the expansion of emphasis to global ideas, the number of session offerings was increased and an additional day of roundtable lunches was added. As usual, the conference kicked off with short course offerings on Wednesday. These were

well attended, with the ten offerings covering a wide range of relevant topics for conference attendees.

Prior to leaving RISW, I had the opportunity to attend the kickoff meeting for the 2024 RISW steering committee. Energy and engagement were high. I look forward to what is to come next year to continue the strong legacy of the RISW conference.

Now in the post-conference season, I look forward to the end of the year and the holiday season. I wish you and your loved ones a great season to come. Thank you, BIOP members, for trusting me to serve you in this role. I will write again for the Spring 2024 Biopharm Report, sharing reflections on the year. I will be joined by current Chair-elect, Ted Lystig. I look forward to all the Section will accomplish going forward.

Be sure to follow our BIOP LinkedIn profile for regular updates on the Section.



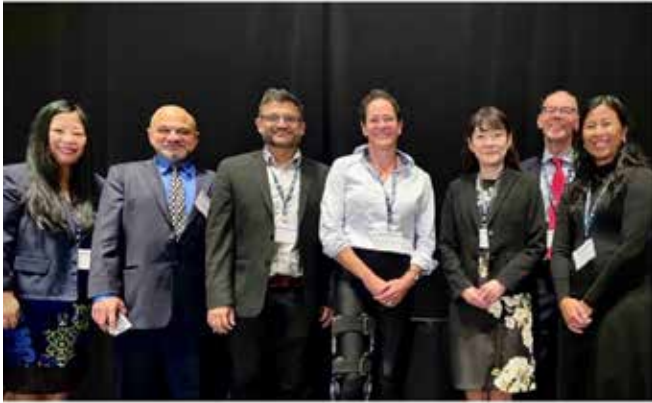
**THE ASA BIOPHARMACEUTICAL SECTION
REGULATORY-INDUSTRY STATISTICS WORKSHOP**
September 27–29, 2023 • Rockville, MD
Statistical Thinking and Innovation with Global Impact

REGULATORY INDUSTRY STATISTICS WORKSHOP 2023 RECAP

Erik Bloomquist, 2023 RISW Co-Chair (Merck)

The 2023 ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop (RISW) was held in North Bethesda, MD from September 27-29 and once again resulted in a highly successful event. The workshop had a theme of “Statistical Thinking and Innovation with Global Impact” and featured over 1,150 statisticians participating in 10 short courses, 49 parallel sessions, and 2 plenary sessions that focused on ICH harmonization and digital health technologies. The session on ICH harmonization

featured Amy Xia from Amgen as moderator and featured a panel of statistical leaders from FDA, EMA, PMDA, Novartis, and UNC Chapel Hill discussing efforts to develop international guidance for drug development. The second plenary session was moderated by Kelly Zou from Viartis and featured talks by Vinay Pai from the FDA Digital Health Center of Excellence and Digital Medicine Society (DiME) CEO Jennifer Goldsack on digital innovation and outlooks for this area. Due to high demand, both



talks were recorded and are available on the 2023 workshop homepage.

As part of the RISW 2023 global theme, a special outreach effort was conducted to statisticians outside of North America. During this outreach, a collaboration was established between the European Federation of Statisticians in the Pharmaceutical Industry (EFSPI) and ASA BIOP to promote future regulatory-industry statistics workshops held both in North America and Europe. Stay tuned for information on the 9th EFSPI workshop on regulatory statistics workshop to be held September 11-12, 2024 in Basel Switzerland.

For RISW speakers and participants, the peer-reviewed statistics journal *Statistics in Biopharmaceutical Research* is pleased once again to offer a special issue devoted to the 2023 workshop. Authors who presented at the workshop are welcome to submit full papers based on the results presented at RISW2023,

and the journal welcomes submissions beyond the workshop as well. The SRB special issue editorial team includes Erik Bloomquist, Fanni Natanegara, Weili He, Zachary Micah Thomas, and Hana Lee. Deadline for submission is January 31, 2024.

It's never too early to make plans to attend RISW 2024 to be held September 25-27, 2024 at the North Bethesda Marriott. Zhiheng Xu from FDA and Jianchang Lin from Takeda will be co-chairs for the workshop and have a wonderful program planned. Stay tuned to LinkedIn and the RISW 2024 website (<https://ww2.amstat.org/meetings/biop/2024/>) for upcoming participation, submission, and registration deadlines.

We would like to one last time thank all the members of our steering committee, workshop sponsors, and ASA event staff for all their dedication and work planning the 2023 workshop. We hope to see you all once again at RISW 2024.

RECAP OF 2023 BOSTON PHARMACEUTICAL SYMPOSIUM

Jianjun Hua¹, Zhaoyang Teng², Gautier Paux³, Xihao Li⁴, Kristin Baltrusaitis⁵, Wenting Cheng⁶

¹ Dartmouth College, Hanover, NH, ² Servier, Boston, MA, ³ Sanofi, Cambridge, MA, ⁴ UNC-Chapel Hill, NC, ⁵ Center for Biostatistics in AIDS Research, Harvard T.H. Chan School of Public Health, Cambridge, MA, ⁶ Biogen, Cambridge, MA

The 2023 Boston Pharmaceutical Symposium, organized by the Boston Chapter of the American Statistical Association (BCASA), hosted by Sanofi, was a resounding success. The event marked a significant milestone in the history of the symposium, with a record-breaking attendance of 157, the most representative Scientific Committee from both industry and academia, and a wide variety of presentations in different sessions since its inception in 2017. The attendees came from the New England area and other regions with diverse backgrounds, representing various sectors of the pharmaceutical industry, clinical research organization, academia, and university students.

Our Symposium transcended expectations, thanks to the remarkable contributions of our ten distinguished oral presentation speakers and poster presenters. Their wealth of insights, knowledge, and expertise elevated the event, providing attendees with exceptional learning opportunities and making this gathering an unequivocal success. One of the highlights was the morning keynote by Prof. Mark Chang, an Adjunct Professor at Boston University and the visionary Founder of AGInception. In a mesmerizing address, he delved deep into the intricate landscape of Artificial Intelligence and Machine Learning (AI/ML) in Drug Development and Healthcare. Prof. Chang not only illuminated the challenges and regulatory aspects associated with AI applications but also painted a vivid picture of the benefits that stem from shifting from traditional statistics to AI/ML methodologies. His presentation left an indelible mark on everyone present, offering a glimpse into the future of innovation in these critical fields. Furthermore, the afternoon keynote by Rui (Sammi) Tang, Vice President and Global Head of Biometrics at Servier, was nothing short of inspirational. With unwavering passion, Sammi embarked on a journey exploring the theme of strategic innovations in clinical development. Drawing from real-life cases, she masterfully illustrated how statistical innovations can profoundly impact drug development, ultimately leading to tangible benefits for patients. Sammi's compelling narrative underscored the significance of statistical teams as leaders, their indispensable role in

This summary was prepared by the Scientific Committee Leadership Team:

Andrew (Jianjun) Hua, Co-chair, Program Chair of BCASA, Dartmouth College

Zhaoyang Teng, Co-chair, Servier

Gautier Paux, Vice-chair and Venue host, Sanofi

Xihao Li, Vice-chair, UNC-Chapel Hill, formerly Harvard T.H. Chan School of Public Health

Kristin Baltrusaitis, Past Chair, Center for Biostatistics in AIDS Research, Harvard T.H. Chan School of Public Health

Wenting Cheng, the President of BCASA, Biogen

shaping the future of healthcare. Sammi's keynote was a resounding call to action, reminding us all to cherish the joy of being statisticians and our power to make a genuine impact on the lives of patients. The Symposium, enriched by the insights and wisdom of our exceptional speakers, will leave an enduring legacy in the realm of AI/ML and clinical development. The full event agenda and presentation slides can be found on BCASA website: <https://community.amstat.org/bostonchapter/upcoming-events/new-page2>

The Scientific Committee, comprising 24 members, played a crucial role in organizing the symposium. This includes the leaders of the Scientific Committee (Andrew (Jianjun) Hua, Co-chair, Program Chair of BCASA; Zhaoyang Teng, Co-chair; Gautier Paux, Vice-chair and Venue host; Xihao Li, Vice-chair; Kristin Baltrusaitis, Past Chair). Special thanks to the Poster Subcommittee (Xihao Li, Disa Yu, Ina Jazić, Hailu Chen), the Invited Speaker Session Subcommittee (Zhaoyang Teng, Xihao Li, Jianchang Lin, Brooks Clark), Student Award Subcommittee (Kosalaram Goteti, Hrishikesh Kulkarni, Xihao Li) and our Venue Host Representative (Gautier Paux) for their outstanding leadership and unwavering dedication to the Symposium.

The event was made possible by the generous support of sponsors, including Venue Sponsor: Sanofi; Platinum Sponsor: ASA Biopharmaceutical (BIOP)



Section; Gold Sponsors: Phastar, Servier, and ASA BIOP Leadership-in-practice Committee (LipCom); and Silver Sponsors: Cytel, MaxisIT, Algorics, and ClinChoice.

This symposium served as a unique platform for sharing insights into statistical applications and research within the pharmaceutical industry, fostering connections among colleagues involved in statistical practices in the Greater Boston area and nurturing future innovations in this field. The event featured the strong local

community of statisticians, promoting research and innovation in pharmaceutical statistics, highlighting the dynamic and ever-evolving landscape of statistical practices within the pharmaceutical industry. Much positive feedback was received from the attendees.

In conclusion, the 2023 Boston Pharmaceutical Symposium was a landmark event that showcased the collective efforts of many dedicated colleagues and organizations. We look forward to continuing this tradition of excellence in the years to come.

ISBS TO HOLD SYMPOSIUM IN BALTIMORE

Dear ASA BIOP colleagues,

We are proudly announcing that the 7th International Symposium on Biopharmaceutical Statistics will take place on March 6 – 9, 2024, at Hilton Baltimore Inner Harbor. The symposium is organized by the International Society for Biopharmaceutical Statistics with a theme on “Statistical innovation in the era of integrated evidence for medical product development”. The purposes of this symposium are

- To bring together worldwide statisticians and drug development professionals who are involved in quantitative biopharmaceutical research, development, and regulations to share and exchange information, experience, and research findings, and
- To improve and promote the harmonization of statistical practice in the industry at the international front.

Prominent statisticians and drug development professionals from regulatory agencies, academia, and industry will deliver keynote speeches on various emerging/evolving fields. Invited and contributed presentations will cover a wide range of topics from non-clinical statistics, preclinical discovery, clinical development, post-licensure evidence generation, to regulatory science, data science and statistics. A series of short courses will be given by experts in

their respective professional fields. *The Statistics in Biopharmaceutical Research* will publish a special issue of high-quality papers presented (including poster presentation) at the Symposium via a peer-review process according to the policy and principles of the journal.

This symposium is co-sponsored by several organizations:

- American Statistical Association (ASA) Biopharmaceutical Section
- Center for Innovative Study Design, Stanford University
- Department of Biostatistics, Bioinformatics, & Biomathematics, Georgetown University
- Department of Biostatistics and Bioinformatics, The George Washington University
- Department of Mathematics and Statistics, University of Maryland Baltimore County
- UMBC-Stanford Workshop on Clinical Trials and Regulatory Science
- DahShu
- ASA-DahShu IDSWG Multidisciplinary Master Protocol Working Group

The preliminary program will be available online on Nov 15 and the online registration will start on the same day. We sincerely invite you to join us at this event!

ISBS 2024 7th Symposium Organizing Committees



THE INTERNATIONAL SOCIETY FOR BIOPHARMACEUTICAL STATISTICS

THE 7th SYMPOSIUM ON STATISTICAL INNOVATION FOR MEDICAL PRODUCT DEVELOPMENT

MARCH 6 – 9, 2024, HILTON BALTIMORE INNER HARBOR, MARYLAND, USA

CONGRATULATIONS TO THE NEW

ASA Fellows

Big congratulations to the following BIOP members who become ASA fellows in 2023! Your hard work and influence are really valuable to the BIOP community and you continue to be our inspiration.



Margaret Gamalo

Pfizer

For exceptional impact to drug development, benefitting millions of patients worldwide; for promoting statistical methods to expedite access of drugs in children and high unmet need diseases; and for sustained and impactful service to the statistical profession.



Inna T. Perevozskaya

GSK

For outstanding contributions to the statistical profession through development, dissemination, and application of innovative designs in the pharmaceutical industry; for engagement in highly impactful scientific working groups; and for service to the ASA.



Bo Huang

Pfizer

For outstanding statistical leadership and consulting in biopharmaceutical industry, extraordinary research and collaboration, exceptional promotion of novel statistical methods to oncology clinical trials, and exemplary service to the profession.



Xiaofeng Wang

Cleveland Clinic

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.



Kalyan Ghosh

Inference

For statistical application in clinical pharmacology, outstanding organizational leadership in statistical consultancy and services, and excellent leadership and service to the statistical profession.

UPCOMING CONFERENCES

The 79th Annual Deming Conference on Applied Statistics

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

- **Key dates:** registration opens on August 14, 2023.



The 7th International Symposium on Biopharmaceutical Statistics

This conference will take place on March 6-9, 2024, at the Hilton Baltimore Inner Harbor. The symposium is organized by the International Society for Biopharmaceutical Statistics with a theme on “Statistical innovation in the era of integrated evidence for medical product development”. For more information please visit <https://www.isbiostat.org/7th-international-symposium-on-biopharmaceutical-statistics/>.

- **Key dates:** registration opens on November 15, 2023