Pfizer/ASA/Columbia University Symposium on Risks and Opportunities of AI in Clinical Drug Development

May 18, 2020
8:30 a.m. – 5:00 p.m.
Pulitzer Hall|Lecture Hall
Columbia University
New York, NY 10027

The *Pfizer/ASA/Columbia University Symposium on Risks and Opportunities of AI in Clinical Drug Development* is an event jointly sponsored by Pfizer Inc., the American Statistical Association (ASA), and the Statistics Department and Data Science Institute at Columbia University.

Our world increasingly relies on data and computing to create knowledge, to make critical decisions, and to better predict the future. Data science has emerged to support these data-driven activities by integrating and developing ideas, concepts, and tools from computer science, engineering, information science, statistics, and domain fields. Data science now drives fields as diverse as biology, astronomy, material science, political science, and medicine—not to mention vast tracts of the global economy, key government activities, and quotidian social and societal functions.

The pharmaceutical enterprise has been slower to respond, especially to the rapid developments in AI, but tectonic shifts are underway in approaches to the discovery, development, evaluation, registration, monitoring, and marketing of medicines for the benefit of patients and the health of the community.

While there is much discussion about the potential of AI and modern machine learning tools to transform the drug development paradigm, there is a growing recognition of the paucity of research about the inevitable pitfalls and unintended consequences of the digital revolution in this important area of application. As we move toward personalized and truly evidence-based medicine, the use of AI and machine learning to optimize drug deployment raises a whole different set of challenges.

This forum is, therefore, expected to serve as a platform for distinguished statisticians, data scientists, regulators, and other professionals to address the challenges and opportunities of AI in pharmaceutical medicine; to foster collaboration among industry, academia, regulatory agencies, and professional associations; and to propose recommendations with policy implications for proper implementation of AI in promoting public health.

PFIZER/AMERICAN STATISTICAL ASSOCIATION/COLUMBIA UNIVERSITY SYMPOSIUM ON RISKS AND OPPORTUNITIES OF ARTIFICIAL INTELLIGENCE IN CLINICAL DRUG DEVELOPMENT

PULIZER HALL|LECTURE HALL, COLUMBIA UNIVERSITY
MAY 18, 2020

Program Co-Chairs

Demissie Alemayehu, Pfizer Inc. and David Madigan, Columbia University

7:30 AM – 8:25 AM:	Registration and Breakfast
8:25 AM – 8:30 AM:	Welcome Address, David Madigan, Columbia University
8:30 Am – 9:00 AM:	Opening Plenary Session (Chair D. Madigan) • *A Perspective on the Promise and Pitfalls of Artificial Intelligence in Clinical Drug Development.* Kannan Natarajan, Pfizer Inc.
9:00 AM – 10:00 AM:	Keynote Address (Chair D. Madigan) • *Machine Learning: Changing the future of healthcare - from next generation clinical trials to individualized treatment effects. Mihaela van der Schaar, University of Cambridge* • Discussant: Patrick Ryan, Janssen Research and Development
10:00 AM - 11:00 AM:	Plenary Session I (Chair, D. Alemayehu, Pfizer Inc) • *ML and AI in Clinical Development and Statistical Innovation*. Wang Li, AbbVie • *Opportunities for AI in Healthcare*. Lawrence Carin, Duke University
11:00 AM – 11:15 AM:	Break
11:15 AM – 12:15 PM:	Plenary Session II (Chair, John Doyle, Pfizer Inc.) • *Bias from Treatment Contamination in Clinical Decision Rules*. Noémie Elhadad, Columbia University • *Artificial Intelligence in Pharmaceutical R&D at AstraZeneca*. Faisal Khan, AstraZeneca
12:15 PM –1:15 PM:	Lunch
1:15 PM – 2:00 PM:	Fireside Chat (Moderator K. Natarajan) • Amy Abernethy, U.S. Food and Drug Administration
2:00 PM – 3:00 PM:	Plenary Session III (Chair, Javier Cabrera, Rutgers University) • *Bias from Treatment Contamination in Clinical Decision Rules*. Amol Navathe, University of Pennsylvania • *Finding Natural Experiments in Observational Data*. Adler Perotte, Columbia University • Discussant: Jesse Berlin, Johnson & Johnson
3:00 PM – 3:15 PM:	Break
3:15 PM – 4:50 PM:	Panel Discussion: “Ethical and Regulatory Issues with AI in Clinical Drug Development”. (Moderator, Jeannette M. Wing, Columbia University) • Graeme Archer, GlaxoSmithKline • Igor Jurisica, Krembil Research Institute & Univ of Toronto • Mary Ann Slack (U.S. Food and Drug Administration) • Simon Tavaré, Columbia University
4:50 PM – 5:00 PM:	Closing Remarks (Demissie Alemayehu, Pfizer Inc.)

Speakers

Amy P. Abernethy, US Food and Drugs Administration (FDA)

Dr. Amy P. Abernethy is Principal Deputy Commissioner of the US FDA, where she oversees the agency's day-to-day functioning and directs special and high-priority initiatives. Prior to joining the FDA, Abernethy was chief medical officer, chief scientific officer and senior vice president of oncology for Flatiron Health. Earlier, she was a professor of medicine at Duke University School of Medicine.

Graeme Archer, GSK R&D

Dr. Graeme Archer is VP & Head of the Non-clinical and Translational Statistics (NCTS) group in GSK R&D. NCTS ensures integrity in the design, analysis and interpretation of models and experiments developed by GSK's Research Units – Graeme is particularly focused on 'translation' (what do we predict will be observed in man, and how confident in those predictions should we be?) On GSK's Research governance boards he strives for statistical rigour in portfolio decision-making – and was joint winner of the 2019 Royal Statistical Society's Pharmaceutical Excellence award for his work on Quantitative Decision-Making. His 1990s PhD, from the University of Glasgow, concerned Bayesian methods for image analysis – early 'machine learning'? – and as a consequence he is a champion for synergy between the statistical and AI/ML communities. Graeme spent a sabbatical year in 2016 as a speech-writer for a British cabinet minister.

Jesse Berlin, Johnson & Johnson

Dr. Berlin received his doctorate in biostatistics from the Harvard School of Public Health in 1988. After spending 15 years as a faculty member at the University of Pennsylvania, in the Center for Clinical Epidemiology and Biostatistics, Jesse left Penn to join Janssen Research & Development in Biostatistics. He now serves as Vice President and Global Head of Epidemiology across Johnson & Johnson, responsible for pharmaceuticals, devices and consumer products. He has coauthored over 270 peer-reviewed publications in a variety of clinical and methodological areas, including papers on the study of meta-analytic methods for both randomized trials and epidemiology. He served on an Institute of Medicine Committee that developed recommendations for the use of systematic reviews in clinical effectiveness research, and served as Chair of the Scientific Advisory Committee to IMEDS (Innovation in Medical Evidence Development and Surveillance), part of the Reagan-Udall Foundation. Dr. Berlin co-chairs the Scientific Oversight Committee (with Greg Pappas from FDA) for MDEpiNet, a public-private partnership which is working toward developing

methods and data sources for the evaluation of medical devices. Dr. Berlin served as an editorial team member of working group X for CIOMS (The Council for International Organizations of Medical Sciences), which has published guidelines for meta-analysis of drug safety data in the regulatory context. He was elected as a fellow of the American Statistical Association in 2004. In 2013, Dr. Berlin received the Lagakos Distinguished Alumni Award from the Department of Biostatistics at the Harvard School of Public Health.

Lawrence Carin, Duke University

Professor Lawrence Carin is the James L. Meriam Distinguished Professor of Electrical and Computer Engineering at Duke, where he also serves as the Vice Provost for Research. He is one of the most prolific authors in the world in the fields of machine learning and artificial intelligence, regularly publishing in the most competitive forums. He has taught at Duke for over 23 years, and has published nearly 400 peer-reviewed papers. He is also the Chief Scientist at a machine learning startup, Infinia ML.

Noémie Elhadad, Columbia University

Professor Noémie Elhadad is
is an Associate Professor of Biomedical Informatics, affiliated with Computer Science at Columbia University. Her research is at the intersection of machine learning, technology, and medicine. She investigates ways to support clinicians, patients, and health researchers in their information workflow by automatically extracting and making accessible information from unstructured, large, often messy data.

Igor Jurisica, Krembil Research Institute and University of Toronto

Dr. Igor Jurisica, PhD, DrSc is a Senior Scientist at Krembil Research Institute, Professor at U Toronto and Visiting Scientist at IBM CAS. He is an Adjunct Professor at Pathology and Molecular Medicine at Queen's U, an adjunct scientist at the Institute of Neuroimmunology, Slovak Academy of Sciences, and an Honorary Professor at Shanghai Jiao Tong University. Since 2015, he has also served as Chief Scientist at the Creative Destruction Lab, Rotman School of Management. He has been included in Thomson Reuters 2014, 2015 & 2016 lists of Highly Cited Researchers (http://highlycited.com), and The World's Most Influential Scientific Minds: 2015 & 2014 Reports. In 2019, he has been included in the Top 100 AI Leaders in Drug Discovery and Advanced Healthcare list (Deep Knowledge Analytics, http://analytics.dkv.global). He has also won numerous awards, including Tier II and Tier I Canada Research Chairs.

Faisal Khan, AstraZeneca

Dr. Faisal M. Khan is the Executive Director of Advanced Analytics and Artificial Intelligence at AstraZeneca. His work focuses on the intersections of data science, biostatistics, bioimaging, personalized medicine, and healthcare delivery. His career has encompassed all aspects of healthcare and biomedical analytics, including diagnostics, devices, clinical trials/therapeutics, and payers/insurance. Dr. Khan has worked or consulted across academia and industry with both startups and Fortune-50 companies. He has over 90 published papers, abstracts, and patents.

Kannan Natarajan, Pfizer Inc.

Dr. Kannan Natarajan is Head of Global Biometrics and Data Management and is part of Global Product Development Leadership Team at Pfizer Inc. The GBDM organization supports the global clinical development strategy and data sciences across all of Pfizer product portfolio. He is also the Chief Statistical Officer of Pfizer, managing statistical functional excellence across all Pfizer business units. Prior to joining Pfizer, Kannan was Senior Vice President and Global Head of Oncology Biometrics and Data Management at Novartis Pharmaceuticals. Kannan has been in the pharmaceutical industry for over 20 years working across various therapeutic areas. Kannan holds a PhD. Degree in Statistics from the University of Florida.

Amol Navathe, University of Pennsylvania

Dr. Amol Navathe is an Assistant Professor of Health Policy and Medicine and a Senior Fellow at the Leonard Davis Institute for Health Economics at the University of Pennsylvania. Dr. Navathe is a practicing physician, health economist, and engineer. He has extensive expertise in policy analysis and design, physician and hospital economic behavior, and application of informatics and predictive analytics to health care data. His work on advanced health data analytics and technology to improve healthcare delivery has been implemented at numerous large health systems. He has applied his skills to payment and delivery transformation, federal policy for health care evidence development and data infrastructure, and the study of physician and hospital economic behavior. Dr. Navathe completed his medical training at the University of Pennsylvania School of Medicine and his post-graduate medical training at the Brigham and Women's Hospital at Harvard Medical School. He obtained his PhD in Health Care Management and Economics from The Wharton School at the University of Pennsylvania.

Adler Perotte, Columbia University

Dr. Adler Perotte is an Assistant Professor in the Department of Biomedical Informatics as well as the co-founder and faculty advisor to the Health Tech Assembly. Dr. Perotte's primary research areas are Bayesian inference and prediction based on electronic health record data and metabolomics. Prior to joining the department in this capacity, Dr. Perotte was an NLM Postdoctoral Fellow in Biomedical Informatics at Columbia University. Before that, Dr. Perotte worked as a Research Specialist between Princeton's Computational Memory and Artificial Intelligence laboratories studying probabilistic models of memory. While earning his MD from Columbia University, Dr. Perotte interned with Columbia's technology transfer office to evaluate promising university technology for commercial potential.

Patrick Ryan, Janssen Research and Development

Patrick Ryan, PhD is Vice President, Observational Health Data Analytics at Janssen Research and Development, where he is leading efforts to develop and apply analysis methods to better understand the real-world effects of medical products. He is an original collaborator in Observational Health Data Sciences and Informatics (OHDSI), a multi-stakeholder, interdisciplinary collaborative to create open-source solutions that bring out the value of observational health data through large-scale analytics. He served as a principal investigator of the Observational Medical Outcomes Partnership (OMOP), a public-private partnership chaired by the Food and Drug Administration, where he led methodological research to assess the appropriate use of observational health care data to identify and evaluate drug safety issues. Patrick has worked in various positions within the pharmaceutical industry at Pfizer and GlaxoSmithKline, and also in academia at the University of Arizona Arthritis Center.

Mihaela van der Schaar, University of Cambridge

Professor van der Schaar is John Humphrey Plummer Professor of Machine Learning, Artificial Intelligence, and Medicine at the University of Cambridge and Turing Faculty Fellow at The Alan Turing Institute in London. She was elected IEEE Fellow (2009) and has received numerous awards, including the Oon Prize on Preventative Medicine from the University of Cambridge (2018). She holds 35 granted USA patents. In 2019, she was identified by National Endowment for Science, Technology and the Arts as the female researcher based in the UK with the most publications in the field of AI.

Mary Ann Slack, US Food and Drugs Administration (FDA)

Ms. Slack has 30+ years of extensive leadership and management experience in both the public and private sectors, developing informatics strategy and implementing business solutions. She currently serves as Director of FDA's CDER Office of Strategic Programs (OSP). OSP plays a lead role in many of the Center's and FDA's strategic initiatives, including decision support, data standards, program analysis, informatics and governance. Since joining CDER in 2003, Ms. Slack has led numerous large, complex initiatives with broad stakeholder impact, including establishing CDER's data standards program and leading a team of experts in defining FDA's operational implementation of the FDA-EU mutual recognition of inspections. Ms. Slack serves on several Boards and Committees where she supports FDA's needs and perspectives.

Simon Tavaré, FRS, FMedSci, Columbia University

Professor Simon Tavaré is Professor of Statistics and of Biological Sciences and the Herbert and Florence Irving Director of the Irving Institute for Cancer Dynamics at Columbia University. Prior to joining Columbia, he was Director of the Cancer Research UK Cambridge Institute, Professor of Cancer Research at the Department of Oncology and Professor in the Department of Applied Mathematics and Theoretical Physics at the University of Cambridge. Simon is a fellow of the Academy of Medical Sciences and the Royal Society, and a member of EMBO. He was president of the London Mathematical Society from 2015 to 2017, and was elected a Fellow of the American Mathematical Society and a Foreign Associate of the National Academy of Sciences in 2018.

Li Wang, AbbVie

Dr. Li Wang is currently Senior Director and the Head of Statistical Innovation group in AbbVie. Prior to this senior leadership role, he led Immunology and Solid Tumor statistical design and strategy discussions and multiple innovation projects in AbbVie Development Design Center from 2017 to 2019. He joined AbbVie in 2013 as the lead statistician for RINVOQ RA phase 2 and multiple other compounds in early immunology in RA, OA and GI. Prior to that, Li worked in Bristol-Myers Squibb as Lead Statistician to support multiple NDAs, SNDAs and ROW submissions for blockbusters Eliquis and Onglyza in Cardiovascular and Metabolic Therapeutic Areas. He is enthusiastic in teaching statistical courses to non-statisticians and promoting novel statistical methodologies. He is also active in statistical communities as chair of ICSA Midwest Chapter and executive committee for DIA VJC

ABSTRACTS

Opening Plenary Session

Title. *A Perspective on the Promise and Pitfalls of Artificial Intelligence in Clinical Drug Development*. Kannan Natarajan, Pfizer Inc.

Abstract. The advance in high-speed computing and accessibility to large volumes of diverse and rich data, coupled with the availability of smart algorithms, have presented considerable challenges and opportunities to pharmaceutical companies and regulatory agencies in the efforts to develop much needed medicines with speed, quality and efficiency. In this talk, we highlight the potential role of artificial intelligence (AI) at every stage of drug development, from drug discovery to the design and execution of clinical studies intended for registration, while stressing the need to be vigilant about unintended consequences that might inauspiciously impact patient safety and privacy. Illustrations of AI use will be provided with reference to best practices routinely implemented at Pfizer.

Keynote Address

Title. *Machine Learning: Changing the future of healthcare - from next generation clinical trials to individualized treatment effects.* Mihaela van der Schaar, University of Cambridge

Abstract. TBD

Plenary Session I

Title*. ML and AI in Clinical Development and Statistical Innovation*. Li Wang, AbbVie

Abstract. ML and AI are widely used in technology industry now and are finding their ways into drug development in pharmaceutical industry especially in manufacturing and discovery. In clinical development, how and where to appropriately apply ML/AI and what value it can bring to the business remain big questions without clear answers to researchers. For traditional clinical statisticians, it is not only a big challenge but also an exciting opportunity. AbbVie's statistical innovation group is revamped with the new vision and focus on Digital, RWD, ML and Bayesian methodologies. We are closely collaborating with different Development functions to leverage multiple data sources to help make smarter decisions. Experiences and some use cases will be shared.

Title. *Opportunities for AI in Healthcare***.** Lawrence Carin, Duke University

Abstract: Early progress is being demonstrated on how machine learning (ML) and artificial intelligence (AI) can impact analysis of data connected to health, with both basic and clinical applications. Areas of particular focus have been in deep learning for image analysis, natural language processing, and analysis of tabular data. There have also been advances in analysis of time-dependent data. In this talk we will review progress and directions in these areas.

Plenary Session II

Title. *Data-Powered Patient-Centered Care*. Noémie Elhadad

Abstract. Despite the pervasive deployment of electronic health record systems, their promise to streamline entering and retrieving clinical data, and the recent advances in artificial intelligence in healthcare, getting meaningful and actionable information at the point of care is still a formidable

challenge for clinicians. In this talk, I will present our approach to designing, building, and deploying tools to support clinicians in their decision-making workflow, as well as facilitating the patient-provider partnership in shared-decision making.

Title*. Artificial Intelligence in Pharmaceutical R&D at AstraZeneca*. Faisal Khan, AstraZeneca

Abstract. Data Science and artificial intelligence are employed at AstraZeneca across the full R&D spectrum, impacting research from pre-clinical and discovery, through Phase 3 trials and beyond. This presentation will cover some of those applications and the lessons for good data science which have been learned.

Plenary Session III

Title. *Bias from Treatment Contamination in Clinical Decision Rules*. Amol Navathe, University of Pennsylvania

Abstract. Artificial intelligence approaches are increasingly being applied to clinical decisions. In this talk I examine challenges in producing clinically meaningful, unbiased predictions to support human clinical decision-making based on routinely collected electronic health record data. I use the case of sepsis to illustrate. Wide variation in practice patterns and outcomes for sepsis spawned national campaigns to standardize care. The quick Sepsis Related Organ Failure Assessment [qSOFA] score was included in the 2016 Sepsis-3 criteria. However, decision rules like qSOFA suffer from treatment contamination bias, because outcomes in both treatment and control groups used in their development were measured after the administration of sepsis treatment (e.g., antibiotics). Later definitions of clinical sepsis (Rhee et al, 2017) suffer from the same bias. In this talk I examine the effects of treatment contamination bias in sepsis decision rules and propose alternative approaches.

Title. *Finding Natural Experiments in Observational Data.* Adler Perotte, Columbia University

Abstract. Evidence on clinical treatments based on real world data has the great benefit of representing effectiveness in a broader population than is typically studied in prospective experiments. However, confounders that influence both the choice of treatment and the outcomes of interest can disguise the truth about a given treatment's effectiveness. Methods such as propensity score matching and weighting offer approaches to minimize the effect of these confounders, but leave open the question of which populations of people should be compared. In this work, we explore a computational approach to identifying the subgroup from all treatment arms of an observational study that are comparable, effectively finding the natural experiment for the target treatments. This work leverages a deep generative model based on generative adversarial networks and results in the identification of a comparison distribution that minimizes the variance of estimates such as the average treatment effect.