June 5, 2023 | 7:00 AM – 8:00 PM

2023 Annual Symposium on Risks and Opportunities of AI in Pharmaceutical Medicine

AI and Causality

Pfizer

The Roux Institute Northeastern University

ASA AMERICAN STATISTICAL ASSOCIATION Promoting the Practice and Profession of Statistics

The Pfizer/Northeastern/ASA Symposium on Risks and Opportunities of AI in Pharmaceutical Medicine is an event jointly sponsored by Pfizer Inc., The Roux Institute at Northeastern University, and the American Statistical Association (ASA).

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.

Keynote Speaker

Guido W. Imbens, PhD
2021 Nobel Memorial Prize Laureate in Economic Sciences
Applied Econometrics Professor and Professor of Economics
Stanford Graduate School of Business

Speakers and Panelists

- Aloka Chakravarty, PhD, US Food & Drug Administration
- Anant Madabhushi, PhD, Emory University
- Bruce Church, PhD, Aitia
- Elizabeth A. Stuart, PhD, Johns Hopkins Bloomberg School of Public Health
- Emre Kıcıman, PhD, Microsoft Research
- George M. Hripcsak, MD, MS, Columbia University
- Haoda Fu, PhD, Eli Lilly
- Kannan Natarajan, PhD, Pfizer Inc.
- Louisa Smith, PhD, Northeastern University
- Mark van der Laan, PhD, University of California, Berkeley
- Michael Katehakis, PhD, Rutgers University
- Michael T. Lu, MD, MPH, Harvard Medical School
- Nigam Shah, PhD, Stanford University
- Sheraz Khan, PhD, Pfizer Inc.
- Tala Fakhouri, PhD, US Food & Drug Administration
- Tianxi Cai, ScD, Harvard University

Registration Fees

In-Person	$125
Virtual	$75
Student/Recent Graduate*	$0

In-person registration is all-inclusive of refreshments; breakfast and lunch will be served.

**The Student/Recent Graduate registration rate is available only to individuals currently enrolled in a graduate, post-graduate, or undergraduate program; or individuals who graduated from such a program in the prior 12 months.*

Thank you to the 2023 AIPM Program Chairs, Steering Committee and Program Committee members for their leadership developing this year's program!

Program Chairs

Demissie Alemayehu, PhD
Pfizer Inc.

David Madigan, PhD
Northeastern University

Sponsors

Steering Committee

- Kannan Natarajan, Pfizer Inc.
- Demissie Alemayehu, Pfizer Inc.
- David Madigan, Northeastern University
- Asieh Golozar, Odysseus Data Services & Northeastern University
- Ronald Wasserstein, American Statistical Association

Program Committee

- Demissie Alemayehu, Pfizer Inc. (Chair)
- Javier Cabrera, Rutgers University
- Margaret Gamalo, Pfizer Inc.
- Shameer Khader, Sanofi
- Kristin Kostka, Northeastern University
- Hana Lee, US Food and Drug Administration (FDA)
- Subha Madhavan, Pfizer Inc.
- Justin Manjourides, Northeastern University
- Suzanne Taranto, Pfizer Inc.

Program Co-Chairs: Demissie Alemayehu, Pfizer Inc. and David Madigan, Northeastern University

7:00 AM – 8:20 AM	Registration and Breakfast
8:20 AM – 8:25 AM	**Welcome – David Madigan, Northeastern University & Ron Wasserstein, American Statistical Association**
8:30 AM – 8:35 AM	**Opening Remarks – Kannan Natarajan, Pfizer Inc.**
8:35 AM – 9:50 AM	**Plenary Session I – Chair, Justin Manjourides, Northeastern University** • *"AI in Drug Development". Tala Fakhouri, US Food and Drug Administration (FDA) (25 min)* • *"Improving Real World Evidence: Reproducibility and Interoperability". Tianxi Cai, Harvard University (25 min)* • *"Predicting Respiratory Illness from Voice with AI/ML Mode". Sheraz Khan, Pfizer Inc. (25 min)*
9:50 AM – 10:00 AM	Break
10:00 AM – 11:15 AM	**Plenary Session II – Chair, Javier Cabrera, Rutgers University** • *"Reinforcement learning for individualized treatments in clinical practice and clinical trials". Michael Katehakis, Rutgers University (25 min)* • *"The best (and worst?) of both worlds? Combining EHR and clinical trial data to understand treatment effect heterogeneity". Elizabeth A. Stuart, Johns Hopkins Bloomberg School of Public Health (25 min)* • *"Causal AI for Learning 'Digital Twins' from Human Multi-Omic Data for Drug Discovery and Clinical Development: a Cardiovasculasr Disease Case Study." Bruce Church, Aitia (25 min)*
11:15 AM – 11:30 AM	Break
11:30 AM – 12:30 PM	**Keynote Address - Moderator, Joseph Cappelleri, Pfizer Inc.** *"Causal Inference with Observational Data". Guido W. Imbens, Stanford University*
12:30 PM – 1:30 PM	Lunch
1:30 PM – 2:45 PM	**Plenary Session III - Chair, Subha Madhavan, Pfizer Inc.** • *"Interpreter of Maladies". Anant Madabhushi, Emory University (25 min)* • *"AI to Predict Risk from Chest X-Rays and CTs". Michael T. Lu, Massachusetts General Hospital and Harvard Medical School (25 min)* • *"Extending a Causal Analysis Suite for Health Analyses: Capturing and Validating Critical Assumptions". Emre Kıcıman, Microsoft Research (25 min)*
2:45 PM – 3:00 PM	Break
3:00 PM – 5:00 PM	**Keynote Panel:** *Causality and AI in Biomedical Research & Development* **Moderator: David Madigan, Northeastern** • Aloka Chakravarty, US Food and Drug Administration • Haoda Fu, Eli Lilly • George M. Hripcsak, Columbia University • Nigam Shah, Stanford University • Mark van der Laan, UC Berkeley
5:00 PM – 5:25 PM	**Closing Remarks – Louisa Smith, The Roux Institute, Northeastern University**
5:25 PM – 5:30 PM	**Acknowledgements – Demissie Alemayehu, Pfizer Inc.**
6:00 PM – 8:00PM	**Evening Reception - Featuring Rai Winslow, The Roux Institute, Northeastern University** *Please join us for an evening reception on "Engineering the Future of Healthcare"* *Across the street at The Roux Institute Harbor Side roof deck*

KEYNOTE ADDRESS

Guido W. Imbens, PhD

Guido Imbens, a laureate of the 2021 Nobel Memorial Prize in Economic Sciences, is Applied Econometrics Professor and Professor of Economics at Stanford Graduate School of Business. After graduating from Brown University, Dr. Imbens taught at Harvard University, UCLA, and the University of California at Berkeley. His research focuses on developing methods for drawing causal inferences in observational studies, using matching, instrumental variables, and regression-discontinuity designs. He holds honorary doctorates from Brown University and the University of St. Gallen. Dr. Imbens is an elected fellow of the Econometric Society, American Academy of Arts and Sciences, and the National Academy of Sciences.

SPEAKER & PANELIST BIOS

Aloka Chakravarty, PhD

Aloka Chakravarty, PhD, is currently the Senior Statistical Advisor and Director of Data Analytics in the Office of the Commissioner, FDA for real-world data and evidence activities related to collaborations on COVID-19 and others. She is also working on select strategic data initiatives including Artificial Intelligence and Machine Learning at FDA with the Chief Data Officer. Prior to that, she was the Deputy Director of the Office of Biostatistics in CDER, FDA. She is an internationally recognized thought leader in multi-regional clinical trials, safety evaluation, real world data and evidence, surrogate markers and biomarkers in drug development. Dr. Chakravarty served as an Adjunct Faculty in Department of Statistics, FAES, NIH and has been on Advisory Board of multiple academic institutions. Dr. Chakravarty has received numerous awards, including the FDA Award of Merit and Dr. Frances O. Kelsey Drug Safety Excellence Award. She received her Ph.D. in Statistics from Temple University, and M. Stat from Indian Statistical Institute. Dr. Chakravarty is a Fellow of the American Statistical Association and an Associate Editor of Statistics in Biomedical Research.

Anant Madabhushi, PhD

Anant Madabhushi, PhD, is a Professor of Biomedical Engineering; and on faculty in the Departments of Pathology, Biomedical Informatics, and Radiology and Imaging Sciences at Emory University. He is also a Research Health Scientist at the Atlanta Veterans Administration Medical Center. Dr. Madabhushi has authored more than 450 peer-reviewed publications and more than 100 patents either issued or pending in the areas of artificial intelligence, radiomics, medical image analysis, computer-aided diagnosis, and computer vision. He is a Fellow of the American Institute of Medical and Biological Engineering (AIMBE), Fellow of the Institute for Electrical and Electronic Engineers (IEEE) and a Fellow of the National Academy of Inventors (NAI). In 2015, he was named by Crain's Cleveland Business as one of "Forty under 40" making positive impact to business in Northeast Ohio. In 2017, he received the IEEE Engineering in Medicine and Biology Society (EMBS) award for technical

achievements in computational imaging and digital pathology. His work on "Smart Imaging Computers for Identifying lung cancer patients who need chemotherapy" was called out by Prevention Magazine as one of the top 10 medical breakthroughs of 2018. In 2019, Nature Magazine hailed him as one of 5 scientists developing "offbeat and innovative approaches for cancer research". In 2019, 2020, 2021 and 2022 Dr. Madabhushi was named to The Pathologist's Power List of 100 inspirational and influential professionals in pathology.

Bruce Church, PhD

Bruce Church, PhD, is Chief Mathematics Officer and EVP, Research and Early Development responsible for developing algorithms for REFS and leading projects for the development of new products and technologies. Bruce is a member of the founding team of Aitia. An expert in statistical physics, machine learning, parallel computing, and causal modeling, Bruce helped develop the original REFS platform. Previously, he spent ten years at Cornell University, developing global optimization methods for computational protein folding, the results of which have been published in several peer-reviewed journals. Bruce has served as the principal investigator on several major grants, including a $2.5 million award from the Department of Energy. He has served on the Board of Directors for U.S. Rugby and coached the under-23 women's Northeast rugby team. Bruce received a BS in applied and engineering physics and a PhD in applied physics from Cornell University.

Elizabeth A. Stuart, PhD

Elizabeth A. Stuart, PhD, is Bloomberg Professor of American Health in the Department of Mental Health at the Johns Hopkins Bloomberg School of Public Health, with joint appointments in the Department of Biostatistics and the Department of Health Policy and Management. She also serves as Executive Vice Dean for Academic Affairs at the School. She received her Ph.D. in statistics in 2004 from Harvard University and is a Fellow of the American Statistical Association (ASA) and the American Association for the Advancement of Science (AAAS). Dr. Stuart has extensive experience in methods for estimating causal effects and dealing with the complications of missing data in experimental and non-experimental studies, particularly as applied to mental health, public policy, and education. Her primary research interests include designs for estimating causal effects in non-experimental settings (such as propensity scores), methods to assess and enhance the generalizability of randomized trials to target populations, and methods for policy evaluation. She has received research funding for her work from the National Science Foundation, the Institute of Education Sciences, the WT Grant Foundation, and the National Institutes of Health and has served on advisory panels for the National Academy of Sciences, the US Department of Education, and the Patient Centered Outcomes Research Institute. She received the mid-career award from the Health Policy Statistics Section of the ASA, the Gertrude Cox Award for applied statistics, Harvard University's Myrto Lefkopoulou Award for excellence in Biostatistics, and the Society for Epidemiologic Research Marshall Joffe Epidemiologic Methods award.

Emre Kiciman, PhD

Emre Kıcıman, PhD, is a Senior Principal Researcher at Microsoft Research, where his research interests span causal inference, machine learning, and AI's implications for people and society. Emre is a co-founder of the DoWhy library for causal machine learning. He received his PhD in Computer Science from Stanford University.

George Hripcsak, MD, MS

George Hripcsak, MD, MS, is Vivian Beaumont Allen Professor at Columbia University's Department of Biomedical Informatics. He is a board-certified internist with degrees in chemistry, medicine, and biostatistics. Dr. Hripcsak's research focus is on the clinical information stored in electronic health records and on the development of next-generation health record systems. Using nonlinear time series analysis, machine learning, knowledge engineering, and natural language processing, he is developing the methods necessary to support clinical research and patient safety. He leads the Observational Health Data Sciences and Informatics (OHDSI) coordinating center; OHDSI is an international network with thousands of collaborators and health records on almost one billion patients. In precision medicine, he serves as a PI on Columbia's eMERGE grant, Columbia's regional recruitment center for the All of Us Research Program, and Columbia's role on the All of Us Data and Research Center. He co-chaired the Meaningful Use Workgroup of U.S. Department of Health and Human Services's Office of the National Coordinator of Health Information Technology. Dr. Hripcsak is a member of the National Academy of Medicine, the American College of Medical Informatics, the International Academy of Health Sciences Informatics, and the New York Academy of Medicine. He has over 500 publications.

Haoda Fu, PhD

Haoda Fu, PhD, is an Associate Vice President and an Enterprise Lead for Machine Learning, Artificial Intelligence, and Digital Connected Care from Eli Lilly and Company. Dr. Fu is a Fellow of ASA (American Statistical Association), and IMS Fellow (Institute of Mathematical Statistics). He is also an adjunct professor of biostatistics department, Univ. of North Carolina Chapel Hill and Indiana university school of medicine. Dr. Fu received his Ph.D. in statistics from University of Wisconsin - Madison in 2007 and joined Lilly after that. Since he joined Lilly, he is very active in statistics methodology research. He has more than 100 publications in the areas, such as Bayesian adaptive design, survival analysis, recurrent event modeling, personalized medicine, indirect and mixed treatment comparison, joint modeling, Bayesian decision making, and rare events analysis. In recent years, his research area focuses on machine learning and artificial intelligence. His research has been published in various top journals including JASA, JRSS, Biometrika, Biometrics, ACM, IEEE, JAMA, and Annals of Internal Medicine. He has been teaching topics of machine learning and AI in large industry conferences including teaching this

topic in FDA workshop. He was board of directors for statistics organizations and program chairs, committee chairs such as ICSA, ENAR, and ASA Biopharm session. He is a COPSS Snedecor Awards committee member from 2022-2026, and will also serve as an associate editor for JASA theory and method from 2023.

Kannan Natarajan, PhD

Kannan Natarajan, PhD, is the 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.

Louisa A. Smith, PhD

Louisa H. Smith, PhD, Assistant Professor, Department of Health Sciences, at Northeastern University. Dr. Smith is an epidemiologist who focuses on developing and applying methods for causal inference in public health research. Dr. Smith approaches her applied research using a target trial framework, through which she has helped to clarify questions about exposures during pregnancy and to improve understanding of the effects of COVID-19 on birth outcomes. Her work in prostate and breast cancer addresses questions about complex treatment strategies over time. Her previous work on sensitivity analysis extended the E-value framework to quantify possible effects of selection bias, alone and jointly with other biases. She has also contributed to the literature on mediation, including sensitivity analysis for unmeasured mediator-outcome confounding. Currently she is researching methods to assess sensitivity to missing data under various assumptions

Mark Johannes van der Laan, PhD

Mark Johannes van der Laan, PhD, is the Jiann-Ping Hsu/Karl E. Peace Professor of Biostatistics and Statistics and Co-Director, Center for Targeted Machine Learning and Causal Inference at the University of California, Berkeley. He has made contributions to survival analysis, semiparametric statistics, multiple testing, and causal inference. He also developed the targeted maximum likelihood methodology. He is a founding editor of the Journal of Causal Inference. He received his Ph.D. from Utrecht University with a dissertation titled “Efficient and Inefficient Estimation in Semiparametric Models”. He received the COPSS Presidents’ Award in 2005, the Mortimer Spiegelman Award in 2004, and the van Dantzig Award in 2005.

Michael Katehakis, PhD

Michael Katehakis, PhD, is a Distinguished Professor of Operations Research in the Department of Management Science and Information Systems at Rutgers University and chair of the Department. Much of his work has been on the interaction between optimization and statistical inference. Specific research interests include Reinforcement Learning, Markovian Decision Processes, Data Analysis, and their application to Biostatistics, Health Care, and Operations Management. Many of these subjects are now known as Data Analytics a rapidly developing field. Professor Katehakis joined the Rutgers University faculty in 1989 after receiving his doctorate in Operations Research at Columbia University under the supervision of Cyrus Derman, and after being a faculty member at the Applied Mathematics & Statistics Department at SUNY Stony Brook and the Technical University of Crete. Also, professor Katehakis was a member of the technical staff at the Operations Research Center of Bell - Laboratories, and a consultant at the Brookhaven National Laboratory. He has held visiting appointments and taught at Columbia University, and Stanford University. He has co-authored many papers, with distinguished leaders in his field including Cy Derman, Herbert E. Robbins, Sheldon M. Ross, Arthur F. Veinott Jr., Jerzy Filar, Uriel Rothblum, and Govindarajulu Z. with whom he won the 1992 Wolfowitz Prize for the paper "Dynamic allocation in survey sampling''. His work has been published in top journals and it has been funded by grants from the NSF and the AFOSR. Many of his Ph.D. students and their academic descendants are listed in The Mathematics Genealogy Project. Dr. Michael N. Katehakis is an Associate Editor for the journals: Annals of Operations Research, Mathematics of Operations Research, Naval Research and Logistics, Operations Research Letters, Probability in the Engineering and Informational Sciences. Michael is a past President of the College of Service Operations, Production and Operations Management Society (POMS). Michael is a Fellow of the Institute for Operations Research and the Management Sciences (INFORMS), an Elected Member of the International Statistical Institute (ISI) and a Senior Member of the Institute of Electrical and Electronics Engineers (IEEE).

Michael T. Lu, MD, MPH

Michael T. Lu, MD, MPH, is Director of AI and Co-Director of the Massachusetts General Hospital (MGH) Cardiovascular Imaging Research Center (CIRC), Director of the MGH Imaging Trials Center (MITC), Associate Chair of Imaging Science for the MGH Department of Radiology, and Assistant Professor of Radiology at Harvard Medical School. He is Associate Editor for the Journal of Cardiovascular Computed Tomography. His research focuses on A) clinical trials of cardiac CT to improve health and B) machine learning to predict health outcomes from multimodal imaging. He is Co-PI of the Data Coordinating Center (DCC) of the REPRIEVE trial, a NHLBI/NIH-sponsored multicenter randomized controlled trial of statins to reduce coronary plaque and prevent cardiovascular events in persons with HIV. Recent work has explored convolutional neural networks to predict long-term mortality, incident lung cancer, postoperative mortality, and longevity from chest radiograph (x-ray) images and to automate coronary artery calcium scoring on chest CT. Dr. Lu

earned his undergraduate, MD, and MPH degrees from Harvard University. He completed his residency in Diagnostic Radiology at the University of California, San Francisco, and fellowships in Thoracic and Cardiac Imaging at MGH.

Nigam Shah, PhD

Nigam Shah, PhD, is Professor of Medicine (Biomedical Informatics) at Stanford University, and serves as the Chief Data Scientist for Stanford Health Care. Dr. Shah's research group analyzes multiple types of health data (EHR, Claims, Wearables, Weblogs, and Patient blogs), to answer clinical questions, generate insights, and build predictive models for the learning health system. In his Chief Data Scientist role, he leads Stanford Healthcare's artificial intelligence and data science efforts in three main areas of impact: advancing the scientific understanding of disease, improving the practice of clinical medicine and orchestrating the delivery of health care. Dr. Shah is an inventor on eight patents and patent applications, has authored over 200 scientific publications and has co-founded three companies. Dr. Shah was elected into the American College of Medical Informatics (ACMI) in 2015 and was inducted into the American Society for Clinical Investigation (ASCI) in 2016. He holds an MBBS from Baroda Medical College, India, a PhD from Penn State University and completed postdoctoral training at Stanford University.

Sheraz Khan, PhD

Sheraz Khan, PhD, is Director of Data Science in Early Clinical Development - Clinical AI/ML and Quantitative Sciences at Pfizer. He obtained his PhD in Computational and Applied Mathematics from the Ecole Polytechnique, France, and has major interests in machine learning, Bayesian statistics and information geometry. He has over two decades of experience in developing signal processing and machine learning algorithms for multivariate time series data. Before joining Pfizer in May 2022, he was an Assistant Professor of Radiology at the Harvard Medical School.

Tala Fakhouri, PhD

Tala Fakhouri, PhD, is the Associate Director for Policy Analysis in the Office of Medical Policy (OMP), Center for Drug Evaluation and Research (CDER), FDA. Dr. Fakhouri's responsibilities are focused on developing policies for drug development and regulatory decision making with emphasis on data science, artificial intelligence (AI) and machine learning (ML), real-world data and real-world evidence (RWD/RWE), and digital health technologies. Prior joining FDA in October of 2020, Dr. Fakhouri served as a Senior Health Scientist and Chief Statistician for the CDC's flagship population survey, the National Health and Nutrition Examination Survey (NHANES). Additionally, she served on the CDC's National Center for Health Statistics Disclosure Review Board, the Cancer Moonshot Data Science Workgroup, and co-led the Federal Committee for Statistical Methodology (FCSM) Nonresponse Bias Subcommittee. Prior to joining NHANES, Dr. Fakhouri served as an Epidemic Intelligence Service Officer with the CDC. She earned a Ph.D. in Oncological Sciences from The Huntsman Cancer Institute at the University of Utah, an MPH in Epidemiologic

and Biostatistical Methods from the Johns Hopkins University School of Public Health, and a postdoctoral fellowship in molecular biology and genetics from Harvard University, and holds a BSc Medical Technology from the Jordan University of Science and Technology.

Tianxi Cai, ScD

Tianxi Cai, ScD, is John Rock Professor of Population and Translational Data Sciences and Professor of Bioinformatics at Harvard T.H. Chan School of Public Health, and Professor of Biomedical Informatics at Harvard Medical School. Dr. Cai is a major player in developing analytical tools for mining EHR data and predictive modeling with biomedical data. She directs the HMS and HSPH translational data science center for a learning health system. Cai's research lab develops novel statistical and machine learning methods for several areas including clinical trials, real world evidence, and personalized medicine using genomic and phenomic data. Cai received her ScD in Biostatistics at Harvard and was an assistant professor at the University of Washington before returning to Harvard as a faculty member in 2002.

KEYNOTE EVENING RECEPTION

Raimond Winslow, PhD

Raimond "Rai" Winslow, PhD, is the Director of Life Science and Medicine Research at the Roux Institute in Portland, Maine. He is also a professor in Northeastern's College of Engineering, and holds appointments in the Khoury College of Computer Sciences and the Bouvé College of Health Sciences' School of Clinical and Rehabilitation Sciences. He joined the Roux Institute from his role as the Raj and Neera Singh Professor of Biomedical Engineering at The Johns Hopkins University School of Medicine, where he was the Founding Director of the Institute for Computational Medicine at The Johns Hopkins University School of Medicine and Whiting School of Engineering. He earned his PhD in biomedical engineering from The Johns Hopkins University School of Medicine and his Bachelor of Science in electrical engineering from Worcester Polytechnic Institute. He is a world-renowned leader in computational medicine, an emerging discipline that applies mathematics, engineering, and computational science to understand human disease.

KEYNOTE ADDRESS

CAUSAL INFERENCE WITH OBSERVATIONAL DATA

Guido W. Imbens, Stanford University

Abstract. In the many disciplines randomized experiments are popular methods for estimating causal effects, partly because their internal validity tends to be high. However, randomized experiments are often small and contain information on only a few variables. At the same time, as part of the big data revolution, large, detailed, and representative administrative data sets have become more widely available. However, the credibility of estimates of causal effects based on such data sets alone can be low. In this paper, we develop statistical methods for systematically combining experimental and observational data to improve the credibility of estimates of the causal effects. We focus on a setting with a binary treatment where we are interested in the effect on a primary outcome that we only observe in the observational sample. Both the observational and experimental samples contain information about the binary treatment, individual characteristics, and a secondary (often short term) outcome. To estimate the effect of a treatment on the primary outcome, while accounting for the potential confounding in the observational sample, we propose a method that makes use of estimates of the relationship between the treatment and the secondary outcome from the experimental sample. We interpret differences in the estimated causal effects on the secondary outcome between the experimental and observational samples as evidence of the presence of unobserved confounders in the observational sample, and develop methods for using those differences to adjust the estimates of the treatment effects on the primary outcome. We illustrate these ideas by revisiting some studies of the effect of small class sizes on educational outcomes. We combine data on class size and third grade test scores from the Project STAR experiment with observational data on class size and both third and eighth grade test scores from the New York school system.

(Based on work with Susan Athey and Raj Chetty)

PLENARY SESSION I

AI IN DRUG DEVELOPMENT

Tala H Fakhouri, Office of Medical Policy, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, Maryland, USA

Abstract. Over the past few decades, the volume of data available to support drug development have increased substantially. These increases in data volume were also accompanied by an expansion in data diversity with data originating from disparate sources including biologic data pharmacometrics data, and clinical data. This growth in data volume and complexity combined with cutting-edge computing power and methodological advancements in artificial intelligence (AI) and machine learning (ML) have the potential to transform how drugs are developed, manufactured and utilized.

Concurrent with these technological advancements, the Food and Drug Administration (FDA) has seen a 2 to 3-fold yearly increase in the number of drug and biologic application submissions using AI and ML components, with over 100 submissions reported in 2021. These submissions traverse the landscape of drug development from drug discovery to post market safety monitoring and cut across a range of therapeutic areas. In general, the application of AI and ML in these submissions aims to improve the efficiency of drug discovery and our understanding of the efficacy and safety of specific treatments. Importantly, the diverse uses of AI in these

submissions highlight the need for a case-by-case regulatory assessment of benefits and risks, and emphasizes the importance of adopting a risk-based management approach that is proportional with measures commensurate with the level of risk posed by the specific context of use for AI and ML.

As with any innovation, AI and ML creates new and unique challenges. To meet these challenges, the FDA has increased its commitment to create a regulatory ecosystem that can facilitate AI innovation and adoption while safeguarding public health. For example, in 2021, the FDA, Health Canada, and the United Kingdom's Medicines and Healthcare products Regulatory Agency (MHRA) jointly published ten guiding principles to inform the development of Good Machine Learning Practices (GMLP) for medical devices that use AI and ML. While these GMLPs were not tailored for drug development specifically, their utility and applicability to drug development is being explored to ensure alignment and consistency whenever possible.

As FDA continues to refine the regulatory ecosystem around the use of AI in decision making, it is important to note that the evidentiary standards needed to support drug approvals remain the same regardless of the technological advances involved. AI and ML will undoubtedly play a critical role in drug development, and the FDA remains committed to robust policy development that both protects and promotes public health.

IMPROVING REAL WORLD EVIDENCE: REPRODUCIBILITY AND INTEROPERABILITY

Tianxi Cai, Harvard University

Abstract. While clinical trials and cohort studies remain critical sources for studying disease progression and treatment response, they have limitations including the generalizability of the study findings to the real world, the limited ability to examine subgroup effects or test broader hypotheses, and the cost in performing these studies. In recent years, due to the increasing adoption of electronic health records (EHR) and the linkage of EHR with specimen bio-repositories and other research registries, integrated large datasets now open opportunities to generate real world evidence (RWE). Generating reliable RWE with EHR studies, however, remain highly challenging due to heterogeneity across healthcare centers in their patient population and health dynamics. In addition, sharing detailed patient level data cross institutions remains infeasible due to privacy constraints. In this talk, I will discuss federated approaches to generating RWE using multi-institutional EHR data.

PREDICTING RESPIRATORY ILLNESS FROM VOICE WITH AI/ML MODELS

Sheraz Khan, Clinical AI/ML and Quantitative Sciences, Pfizer Inc.

Abstract. The Acute Respiratory Illness Surveillance Study (AcRIS) was a fully decentralized, observational clinical trial examining voice changes with respiratory illness (NCT04748445). The study generated real world data of symptoms, voice and biospecimens collected from all 50 states with participants using a mobile app on their personal phones at home. The primary objective was to train and validate a machine learning algorithm for screening SARS-CoV-2-induced COVID-19 illness. Over 9,000 participants were enrolled in the study from April 2021 to April 2022. Participants were 18 years of age or older, and unvaccinated. This talk will present the scientific hypothesis that acoustic features of voice can discriminate between healthy individuals and those with viral respiratory illnesses and the development and implementation of machine learning algorithms developed for screening of respiratory illness. The performance of the algorithms on three different test populations will be discussed, together with lessons learned from this decentralized clinical trial.

PLENARY SESSION II

REINFORCEMENT LEARNING FOR INDIVIDUALIZED TREATMENTS IN CLINICAL PRACTICE AND CLINICAL TRIALS

Michael N. Katehakis, Rutgers University

Abstract. The reinforcement learning (RL) methodology is concerned with the development of efficient algorithms for taking actions, over time, that influence changes in a response-producing environment. The environment generates its responses according to some probability model that depends on the actions taken and some unknown parameters which can be sequentially estimated. Such methods can be employed in a wide range of applications, e.g., to develop AI systems for individualized treatments in clinical practice, clinical trials, marketing, and robotics. In this talk, we discuss the main ideas and theories of the RL field, including the concepts of exploration versus exploitation, regret, and regret rates. We will also discuss the underlying Markov decision processes models, learning from delayed reinforcement, employing empirical models and models with hidden states. We will discuss the difference between good and `optimal' solutions and computational challenges. With the development of personalized medicine and precision medicine reinforcement learning can be use in AI application for treatment regimens and automated medical diagnosis. These ideas can also be used in clinical trials and in personalized treatments. The talk will conclude with a presentation with clinical examples.

THE BEST (AND WORST?) OF BOTH WORLDS? COMBINING EHR AND CLINICAL TRIAL DATA TO UNDERSTAND TREATMENT EFFECT HETEROGENEITY

Elizabeth A. Stuart, Johns Hopkins Bloomberg School of Public Health

Abstract. Estimating treatment effects conditional on observed covariates can improve the ability to tailor treatments to particular individuals. Trials offer unbiased effect estimates but are typically underpowered to detect effect heterogeneity; non-experimental studies such as those using electronic health records can offer large sample sizes but suffer from potential confounding. This talk will discuss methods for combining these data sources, aiming to draw from the strengths of each. A focus will be on recently developed machine learning methods, including causal forests, and their extension to multiple data sources. The methods are motivated by and applied to study of depression treatment; simulation results comparing methods will also be presented.

CAUSAL AI FOR LEARNING "DIGITAL TWINS" FROM HUMAN MULTI-OMIC DATA FOR DRUG DISCOVERY AND CLINICAL DEVELOPMENT: A CARDIOVASCULAR DISEASE CASE STUDY

Bruce Church, Aitia Solutions

Abstract. The amount of clinically curated multi-omic data is growing exponentially and its resolution is increasing by orders of magnitude. While traditional statistics can solve important classes of correlative questions such as risk stratification, the development of Causal AI provides the mathematical foundations for algorithms that learn deep and detailed biological mechanism directly from observational data. The computational complexity of these causal frameworks scales super exponentially with the number of observed variables so the exponential growth in cloud computing power is necessary but not sufficient for high resolution multi-omic Causal AI. Combining Bayesian causal network framework with Metropolis Monte Carlo sampling algorithms from statistical physics allows us to do Causal AI with sufficient resolution and scale to reverse engineer digital twin-mechanistic models of human diseases. These Digital Twins recover both known biology and discover novel interactions that can lead to the discovery of novel therapeutic targets and predictors of response (e.g., biomarkers). Digital Twins have been developed with this approach for neurological disease including Alzheimer's, Parkinson's, and Huntington's diseases and in oncology for blood and solid tumors. Here

we will describe results from a recently published cardiovascular disease Digital Twin for coronary artery disease and unpublished insights in connection to Lp(a).

PLENARY SESSION III

INTERPRETER OF MALADIES: AI FOR PRECISION MEDICINE

Anant Madabhushi, Emory University

Abstract. Traditional biology generally looks at only a few aspects of an organism at a time and attempts to molecularly dissect diseases and study them part by part with the hope that the sum of knowledge of parts would help explain the operation of the whole. Rarely has this been a successful strategy to understand the causes and cures for complex diseases. The motivation for a systems based approach to disease understanding aims to understand how large numbers of interrelated health variables, gene expression profiling, its cellular architecture and microenvironment, as seen in its histological image features, its 3 dimensional tissue architecture and vascularization, as seen in dynamic contrast enhanced (DCE) MRI, and its metabolic features, as seen by Magnetic Resonance Spectroscopy (MRS) or Positron Emission Tomography (PET), result in emergence of definable phenotypes. Within our group has been developing novel computerized knowledge alignment, representation, and fusion tools for integrating and correlating heterogeneous biological data spanning different spatial and temporal scales, modalities, and functionalities. These tools include computerized feature analysis methods for extracting subvisual attributes for characterizing disease appearance and behavior on radiographic (radiomics) and digitized pathology images (pathomics). In this talk I will discuss the development work in our group on new radiomic and pathomic approaches for capturing intra-tumoral heterogeneity and modeling tumor appearance. I will also focus my talk on how these radiomic and pathomic approaches can be applied to predicting disease outcome, recurrence, progression and response to therapy in the context of prostate, brain, rectal, oropharyngeal, and lung cancers. Additionally, I will also discuss some recent work on looking at use of pathomics in the context of racial health disparity and creation of more precise and tailored prognostic and response prediction models.

AI TO PREDICT RISK FROM CHEST X-RAYS AND CTS

Michael T. Lu, Massachusetts General Hospital and Harvard Medical School

Abstract. Chest x-rays and CTs are among the most common medical imaging tests. This talk will discuss advances in artificial intelligence to assess biological age and better assess the risk of future lung cancer and cardiovascular disease from routine chest x-ray and CT images.

EXTENDING A CAUSAL ANALYSIS SUITE FOR HEALTH ANALYSES: CAPTURING AND VALIDATING CRITICAL ASSUMPTIONS

Emre Kıcıman, Microsoft Research

Abstract. The role of a causal software platform is to scaffold the end-to-end causal analysis process to ensure that practitioners follow best practices in modeling and validation. In our experiences, based on usage of our causal tools (DoWhy, EconML, and other tools) across across industrial, agricultural, advertising, health and other domains, we find that data scientists and other practitioners are most challenged by the causal framing of their problems and validation of results. That is, errors in causal analysis are not (just) algorithmic, but in basic assumptions about a particular problem. In extending a causal analysis platform for health analyses, there is an opportunity to embed additional domain knowledge to better bridge the gap between health and causal tasks. In this presentation, I discuss what this can look like and our experiences in causal health analyses.

EVENING KEYNOTE

ENGINEERING THE FUTURE OF HEALTHCARE

Rai Winslow, The Roux Institute at Northeastern University

Abstract. Engineering disciplines apply the following approaches when designing problem solutions: measurements; models; model-based design; design evaluation. Healthcare is undergoing a transformation in which it will become an engineering-based discipline. At The Roux Institute, we are working on two fronts to realize this transformation of healthcare. The first is discovering how to use data (from patients/people) to learn models (of patients/people) that make improved healthcare recommendations (for patients/people). Our goal is for models to become valued and contributing members of the healthcare team. We will illustrate this approach using the ShockAlert and HEART Projects as examples. The second is discovering how models can enable both fundamental understanding of health and disease as well as guide engineering-based design of novel therapies. We will illustrate this approach by presenting results of In Silico studies that predict arrhythmia risk in the setting of genetic mutations, and novel computational approaches for understanding how to modulate risk by drug therapy.