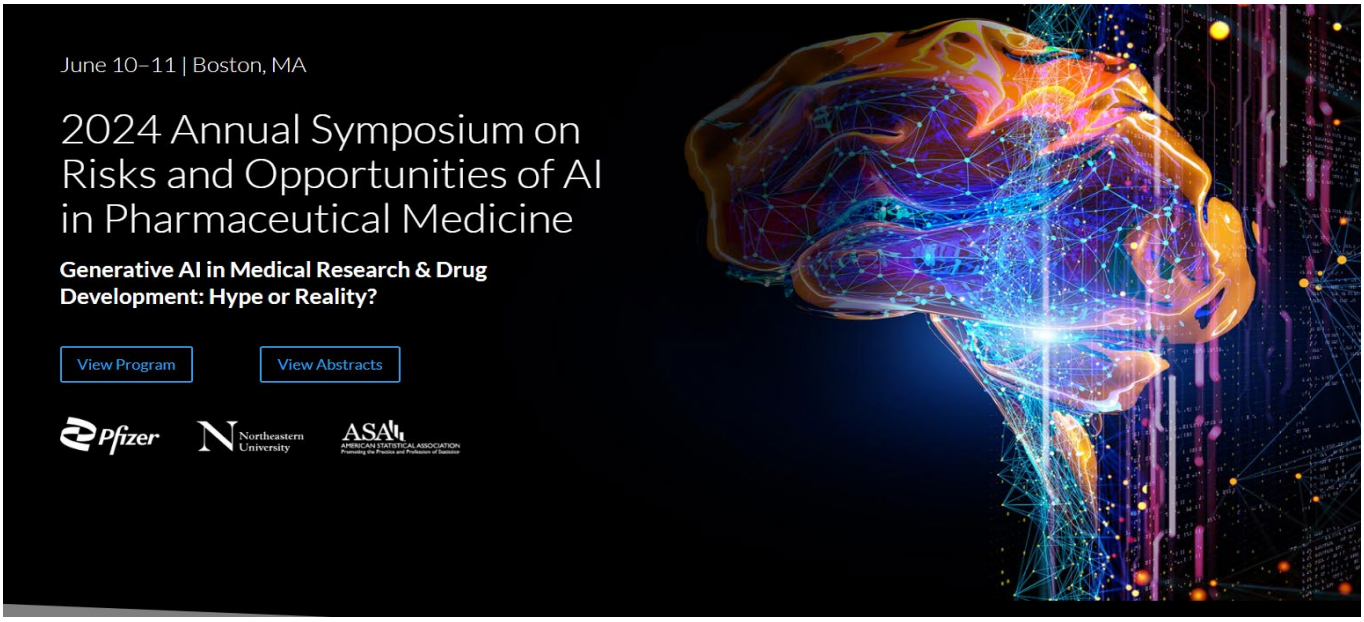


About the Symposium

Where will the AI revolution take Pharmaceutical Science?

The Pfizer/ Northeastern/ASA Symposium on Risks and Opportunities of AI in Pharmaceutical Medicine (AIPM) 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 opportunities, 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.

Co-Chairs

David Madigan, Ph.D. Provost and Senior Vice President for Academic Affairs, Northeastern University

David Madigan, Ph.D., came to Northeastern from Columbia University, where he served as executive vice president for arts and sciences and dean of the Faculty of Arts and Sciences. In those roles, he oversaw five schools with 27 departments and some 50 research centers and institutes. He led successful initiatives to expand lifelong learning programs, and to make Columbia's faculty and student body more diverse. Notably, he spearheaded a program to ensure that Columbia's first-generation students would have the mentoring and support they needed to succeed in a highly demanding academic environment.

At Columbia and in his previous position as Rutgers University's Dean of Physical and Mathematical Sciences, Madigan developed new interdisciplinary learning and research collaborations, and forged industry partnerships to support student and faculty innovation. Under his leadership as department chair, Columbia's statistics faculty ascended to the ranks of the nation's top 10.

A Fellow of the American Association for the Advancement of Science, Madigan has long been a leading researcher at the intersection of Big Data with healthcare innovation. While serving on the statistics faculty at the University of Washington, he was a member of the world-renowned Fred Hutchinson Cancer Research Center.

Madigan also brings a strong global perspective to Northeastern. A native of Ireland, he has led large-scale global collaborations involving academic and pharmaceutical company researchers and government regulators in the U.S. and Europe. Beyond the halls of academia, he has experience as an entrepreneur, innovation leader, researcher, and consultant, both in the U.S. and Ireland. He earned his BA and PhD from Trinity College Dublin.

Demissie Alemayehu, Ph.D., is Vice President of Biostatistics and Head of Statistical Research and Data Sciences (SRDC) in Global Biometrics and Data Management (GBDM) at Pfizer Inc. Demissie has over 25 years of leadership experience in the pharmaceutical industry and has supported projects in almost all therapeutic areas. Demissie has also been influential externally, with decades of research and teaching experience at major institutions. He has co-authored three monographs and published numerous

manuscripts in peer-reviewed journals. In addition, he has held important offices at key professional societies and has served on editorial boards of major journals. He is an elected *Fellow* of the American Statistical Association and holds a Ph.D. degree in Statistics from the University of California at Berkeley.

Steering Committee

Asieh Golozar**, Ph.D.**, **Vice President, Global Head of Data Science, Odysseus Data Services, Inc.**

Asieh Golozar, Ph.D., is a physician-epidemiologist and biostatistician with more than 20 years of experience in health services research, real-world evidence generation, and evaluation of healthcare interventions within government, academia, and industry. She holds a PhD in epidemiology from the Johns Hopkins University School of Public Health and a Master of Health Sciences in biostatistics from JHU, supported by a postdoctoral research fellowship award with the National Cancer Institute's Division of Cancer Epidemiology and Genetics. She earned her medical degree from the Tehran University of Medical Sciences.

After receiving her PhD, Golozar joined the faculty at the JHU Department of Epidemiology, where she focused on cancer and diabetes epidemiology, the application of epidemiologic and statistical methods for robust synthesis of evidence from epidemiologic data, and applying evidence-based findings to strengthen public health infrastructure and policies. In 2017, she moved into industry roles at Bayer AG, AstraZeneca, and Regeneron, where she led lifecycle management and real-world evidence generation activities in oncology, women's health, and other therapeutic areas.

As a professor of the practice, Golozar brings a comprehensive knowledge of healthcare systems to Northeastern. She also has the skills in research methodology and statistics required for addressing safety and effectiveness outcomes and risk assessment, and for generating high quality real-world evidence (RWE) from real-world data. She has played leading roles in large-scale private-public partnerships, including the Innovative Medicines Initiative (IMI) Big Data for Better Outcomes project and the IMI PIONEER Prostate Cancer Study-a-thon.

Since 2018, Golozar has led the OHDSI Oncology Working Group, endeavoring to extend the OMOP common data model to support oncology use cases and advance oncology real-world evidence research. In 2021, Golozar received an OHDSI Titan Award for Clinical Application. In addition to her role at Northeastern, Golozar is an adjunct faculty member at the JHU School of Public Health and vice president of data science at Odysseus Data Services, a leading RWE technology vendor.

Kannan Natarajan, Ph.D., **Senior Vice President, Head of Data Sciences & Analytics, Pfizer Inc.**

Kannan Natarajan, Ph.D., Kannan is Senior Vice President, Head of Data Sciences & Analytics (DSA) and is a leadership team member in the Chief Medical Office at Pfizer Inc. The DSA organization consists of Statistics, Statistical Data Sciences & Analytics and the Advanced Quantitative Data Sciences functions. Each of the functions within DSA support end-to-end development from non-clinical to market access and medical affairs and contributes to the global clinical development and regulatory strategy across all of Pfizer's product portfolio. Prior to this role, Kannan was Senior Vice President, Head of Global Biometrics and Data Management (GBDM).

Kannan is a registered board member of Pfizer Healthcare India Pvt Ltd (PHIPL). PHIPL consists of colleagues in India supporting global Pfizer Research & Development, including product manufacturing. In his role as a PHIPL board member, Kannan represents over 600 Clinical Development colleagues located in the Chennai and Mumbai offices.

He has been in the pharmaceutical industry for over 30 years, working across various therapeutic areas. Prior to joining Pfizer in 2016, he was Senior Vice President and Global Head of Oncology Biometrics and Data Management at Novartis Pharmaceuticals. Kannan holds a Masters in Statistics from the Indian Statistical Institute, Kolkata, India, and a PhD. in Statistics from the University of Florida, Gainesville. Kannan is an elected Fellow of the American Statistical Association.

Ron Wasserstein, Ph.D., **Executive Director, American Statistical Association**

Ronald L. (Ron) Wasserstein, Ph.D., is the executive director of the American Statistical Association (ASA). Wasserstein assumed the ASA's top staff leadership post in August 2007.
In this role, Wasserstein provides executive leadership and management for the association and is responsible for ensuring that the ASA fulfills its mission to promote the practice and profession of statistics. He also is responsible for a staff of 36 at the ASA's headquarters in Alexandria, Va. As executive director, Wasserstein also is an official ASA spokesperson.

Prior to joining the ASA, Wasserstein was a mathematics and statistics department faculty member and administrator at Washburn University in Topeka, Kan., from 1984–2007. During his last seven years at the school, he served as the university's vice president for academic affairs.

Wasserstein is a longtime member of the ASA, having joined the association in 1983, and prior to becoming Executive Director had been active as a volunteer in the ASA for more than 20 years. He was a member of the ASA Board of Directors from 2001–2003.

Wasserstein is a Fellow of the ASA and of the American Association for the Advancement of Science. He was presented the John Ritchie Alumni Award and Muriel Clarke Student Life Award from Washburn University, the Manning Distinguished Service Award from the North American Association of Summer Schools, and the George Mach Distinguished Service Award from Kappa Mu Epsilon National Mathematics Honor Society. For his community service, he received the Champion of Character Award from the Fairfax County Athletic Council and the Administrator of the Year Award from the Virginia Youth Soccer Association.

Ron and his wife, Sherry, live in northern Virginia and enjoy traveling, movies, binge-watching TV series, live theater, audio books, and doting on their children and grandchildren.

Program Committee Members

- Demissie Alemayehu, Pfizer Inc.
- Javier Cabrera, Rutgers University
- Marissa Fiorello, Northeastern University
- Haoda Fu, Lilly
- Meg Gamalo, Pfizer Inc.
- Asieh Golozar, Odysseus
- Nareen Katta, AbbVie
- Emre Kiciman, Microsoft
- Kristin Kostka, Northeastern University
- Hana Lee, FDA
- Subha Madhavan, Pfizer Inc.
- David Madigan, Northeastern University
- Kannan Natarajan, Pfizer Inc.
- Louisa Smith, Northeastern University
- Elizabeth Stuart, Johns Hopkins University
- LeaMarie Suarez-Orozco, Northeastern University
- Suzanne M. Taranto, Pfizer Inc.
- Bingzhi Zhang, Sanofi

Keynote Speaker

Tala H. Fakhouri, Ph.D., MPH
Associate Director,
Policy Analysis at the Food and Drug Administration
CDER, Office of Medical Policy
Tala H. Fakhouri, Ph.D., MPH
Associate Director for Policy Analysis
Food and Drug Administration

Tala H. Fakhouri, Ph.D., MPH, is the Associate Director for Policy Analysis at the Food and Drug Administration. Dr. Fakhouri manages a team tasked with developing, coordinating, and implementing medical policy with a focus on the use of Artificial Intelligence (AI) in drug development. She also contributes to the development of medical policy related to real-world evidence (RWE) for medical product development. In 2023, She was selected by the Office of Management and Budget to serve on the Federal Committee for Statistical Methodology for her expertise in statistical methods. Prior to joining FDA, Dr. Fakhouri served as Chief Statistician for the CDC's flagship population survey, the National Health and Nutrition Examination Survey (NHANES), which is recognized as the premier source of nationally representative data on the health of the nation. Prior to NHANES, she served as an Epidemic Intelligence Service Officer with the CDC, and deputy lead for health surveys at ICF-Macro International. Dr. Fakhouri published over 30 government reports, peer-reviewed papers, and book chapters.

Dr. Fakhouri 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.

Program

Day 1 (June 10, 2024)

East Village, 17th Floor,
Northeastern University
291 St Botolph St, Boston, MA
02115

12:30 PM – 1:00 PM	**Registration**
1:00 PM – 5:00 PM	**Tutorials** (Coordinator, Nareen Katta, AbbVie)
1:00 PM – 3:00 PM	**Tutorial 1:** AI/ML in Pharma. Claire Zhao (Pfizer Inc.) and Haoda Fu (Eli Lilly)
3:00 PM – 3:15 PM	**Break**
3:15 PM – 5:15 PM	**Tutorial 2:** RWE in Drug Development. Hana Lee (FDA) and Susan Gruber (Putnam Data Sciences, LLC)

Day 2 (June 11, 2024)

ISEC, First Floor, Northeastern
University
805 Columbus Ave, Boston,
MA 02120

7:30 AM – 8:30 AM	**Registration and Breakfast**
8:30 AM – 8:35 AM	**Day 2 Welcome** (David Madigan, Northeastern University; Ron Wasserstein, American Statistical Association)
8:35 AM – 8:40 AM	**Opening Remarks and Introduction of Keynote Speaker** (Kannan Natarajan)
8:40 AM – 9:35 AM	**Keynote Address** Tala Fakhouri, Ph.D, MPH, Associate Director for Policy and Analysis, U.S. Food & Drug Administration, CDER, Office of Medical Policy. (Moderator, Kannan Natarajan) (*50 min with Q&A*)
9:35 AM – 10:00 AM	**Break/Posters**
10:00 AM – 11:15 AM	**Day 2 Plenary Session 1** (Chair, Margret Gamalo, Pfizer Inc.) "Empowering Insights in the All of Us Research Program: A Statistical Perspective on the Transformational Role of AI and ML", Qingxia 'Cindy' Chen, Biomedical Informatics, and Ophthalmology & Visual Sciences at Vanderbilt University Medical Center (*25 min*) **"**The Generalist Medical AI Will See You Now", Pranav Rajpurkar, Ph.D., Harvard University (*25 min*) "Generative AI for Case Adjudication in OHDSI", Marc Suchard, UCLA (*25 min*).

11:15 AM – 11:25 AM	**Break/Posters**
11:25 AM – 12:40 PM	**Day 2 Plenary Session 2** (Chair, Bingzhi Zhang, Sanofi) "Generative Mixed-Response State-Space Model for Analyzing Multi-Dimensional Digital Phenotypes", Yuanjia Wang, Columbia University (*25 min*) **"**Machine Learning for Causal Inference", Stefan Wager, Stanford University (*25 min*) "Leveraging the Power of Large Language Models (LLMs) by Statisticians for Pharmaceutical Research & Development: Illustrating with an Example AI App", Junshui Ma, Head of the Biometrics Research Department, Merck Research Lab. (*25 min*)
12:40 PM – 1:30 PM	**Lunch**
1:30 PM – 3:10 PM	**Day 2 Plenary Session 3** (Chair, Emre Kiciman, Microsoft) **"**The Role of Targeted Machine Learning in a Causal Roadmap for Generating High-Quality Real-World Evidence", Lauren Elizabeth Eyler Dang, National Institute of Allergy and Infectious Diseases Biostatistics Research Branch *(25 min)* "The multitude of group affiliations: Algorithmic Fairness, Loss Minimization and Outcome Indistinguishability", Omer Reingold, Stanford University (*25 min*) "Designs for Observational Vaccine Effectiveness Studies", Noam Barda, Ben-Gurion University. *(25 min)* "Drugs in Clinical Trials Out of Generative AI", Petrina Kamya, Insilico Medicine (*25 min*)
3:10 PM – 3:25 PM	**Break**
3:25 PM – 4:50 PM	**Panel Discussion: *Generative AI in Drug Development*** Moderator, David Sontag, MIT Robert Ball, US FDA Brian Caffo, Johns Hopkins University Subha Madhavan, Pfizer Inc. Anthony Philippakis, MIT Hoifung Poon, Microsoft Inc
4:50 PM – 5:00 PM	**Closing Remarks & Acknowledgments, Demissie Alemayehu, Pfizer Inc.**
6:00 PM – 8:00 PM	**Networking Reception** *EXP, 8th Floor, Northeastern University* *815 Columbus Ave, Boston, MA 02120*

Speakers

Our distinguished speakers represent the forefront of AI innovation in pharmaceutical medicine, bringing together leading experts from top research institutions, pharmaceutical companies, and regulatory agencies. From pioneering data scientists to experienced industry practitioners, these thought leaders will drive critical discussions about AI's transformative potential in healthcare.

Robert Ball, M.D., M.P.H., Sc.M

Deputy Director

Center for Drug Evaluation and Research (CDER)

U.S. Food and Drug Administration

Robert Ball M.D., M.P.H., Sc.M., is Deputy Director, Office of Surveillance and Epidemiology (OSE), Center for Drug Evaluation and Research, at the Food and Drug Administration (FDA) where he shares in responsibilities leading OSE staff in premarket and postmarket regulation of drugs and therapeutic biologics through adverse event surveillance, pharmacoepidemiology, risk management, and medication error prevention. From 2008-2013, he served as the Director, Office of Biostatistics and Epidemiology, Center for Biologics Evaluation and Research, FDA where he led statistical and epidemiological evaluation of vaccines, blood, cell, tissue, and gene therapy products. His recent research has included the application of natural language processing and machine learning to improve the evaluation of medical product safety and effectiveness in electronic healthcare data systems.

Noam Barda, M.D., Ph.D.

Head, RWE and Innovation Lab

Chief Epidemiologist

Sheba Pandemic Research Readiness Institute

Noam Barda, M.D., Ph.D. I am a lecturer at the Department of Software and Information Systems Engineering & Department of Epidemiology, Biostatistics and Community Health Sciences at Ben-Gurion University. In parallel, I am the head of the Real-World Evidence Research and Innovation Lab and the Chief Epidemiologist at the Sheba Pandemic Research Readiness Institute (SPRI), both in Sheba Medical Center. I have broad training in medicine (MD, with a specialty in public health and epidemiology), computer science (B.Sc.) and biostatistics (M.Sc.). My post-doctorate was at the department of biomedical informatics at Harvard Medical School. My research focuses on three main fields: causal inference, applied predictive modeling in healthcare, and algorithmic fairness. In each field, I try to use state-of-the-art methods, combining epidemiological and computational methods. Over the last several years, I have published extensively on the topic of Covid-19 vaccines and immunity. These publications have been featured in prestigious medical journals and have been an aid in determining policy in multiple countries.

Brian Caffo, Ph.D.

Professor

Department of Biostatistics

Johns Hopkins University

Brian Caffo, **Ph.D.,** received his doctorate in statistics from the University of Florida in 2001 before joining the faculty at the Johns Hopkins Department of Biostatistics, where he became a full professor in 2013. He has pursued research in statistical computing, generalized linear mixed models, neuroimaging, functional magnetic resonance imaging, image processing and the analysis of big data. He created and led a team that won the ADHD-200 prediction competition and placed twelfth in the large Heritage Health prediction competition. He was the recipient of the Presidential Early Career Award for Scientists and Engineers, the highest award given by the US government for early career researchers in STEM fields. He co-created and co-directs the

SMART (www.smart-stats.org) group focusing on statistical methodology for biological signals. He also co-created and co-directs the Data Science Specialization, a popular MOOC mini degree on data analysis and computing. Dr. Caffo is the former director of the graduate programs in Biostatistics, president of the Bloomberg School of Public Health Faculty Senate, the recipient of the Golden Apple teaching award and AMTRA mentoring awards.

Qingxia (Cindy) Chen, Ph.D.

Professor

Biostats, Bio Med Informatics,

Ophthalmology & Visual Sciences

Vanderbilt University Medical Center

Qingxia Chen, Ph.D., is a Professor of Biostatistics, Biomedical Informatics, and Ophthalmology & Visual Sciences at Vanderbilt University Medical Center. She also serves as the Vice Chair of Education at the Department of Biostatistics and co-leads the statistical efforts at the Data and Research Center for the *All of Us* Research Program.

Lauren Elizabeth

Eyler Dang, M.D., Ph.D., M.P.H.

Mathematical Statistician

National Institute of Allergy and Infectious Diseases (NIAID)

Lauren Eyler Dang, M.D., Ph.D., M.P.H., is a Statistician in the National Institute of Allergy and Infectious Diseases Biostatistics Research Branch. She received her Ph.D., in Biostatistics from University of California, Berkeley and her MD from University of California, San Francisco. Her research focuses on the application, dissemination, and development of

causal inference and targeted learning methods, particularly as applied to global health and infectious disease research.

Margaret Gamalo, Ph.D., FASA

Vice President

Statistics Head, Inflammation & Immunology

Pfizer Inc.

Margaret (Meg) Gamalo, Ph.D., is currently VP, Statistics Head for Inflammation and Immunology in Pfizer Research and Development at Pfizer Inc. She combines expertise in biostatistics, regulatory science and adult and pediatric clinical development. Prior to joining Pfizer, she was Research Advisor, Global Statistical Sciences at Eli Lilly and Company and Mathematical Statistician at the Food and Drug Administration. Meg led the Pediatric Innovation Task Force at the Biotechnology Innovation Organization and spearheaded the push for extrapolation as a default strategy in pediatric drug development to expedite labelling changes in children. She is also a member of the European Forum for Good Clinical Practice – Children's Medicine Working Party that provided guidance on inclusion of adolescents in adult research. Her numerous scientific publications and outreach work in statistics is geared towards efficient methodologies to expedite development of drugs in pediatrics and areas of high unmet medical need. Meg is currently Editor-in-Chief of the *Journal of Biopharmaceutical Statistics* and is passionate about promoting women's role in statistical research. She is involved in many statistical activities in the industry including the recently launched Statistical Perspectives on AI/ML in Pharmaceutical Product Development within the ASA. Recently, she was awarded the distinction of Fellow of the American Statistical Association. She is also actively mentoring statisticians in the pharmaceutical industry and in the building of biostatistics as a research discipline in the Philippines. She received her PhD in Statistics from the University of Pittsburgh and Master of Applied Mathematics – Operations Research from the University of The Philippines

Emre Kiciman, Ph.D.

Senior Principal Researcher

Microsoft Research

Emre Kiciman, Ph.D., is a Senior Principal Researcher at Microsoft Research, where his research interests span causal inference, machine learning, the security of AI systems, and AI's implications for people and society. He received his PhD in Computer Science from Stanford University.

Junshui Ma, Ph.D.

Associate Vice President

Head, Biometrics Research

Merck

Junshui Ma, Ph.D., serves as an Associate Vice President, and the head of the Biometrics Research department at Merck Research Lab. He obtained his PhD in Signal Processing and Machine Learning from Ohio State University in 2001. Following his tenure as a postdoc at Los Alamos National Lab and a faculty member at Ohio State University, he joined Merck in January 2005. Over the past 18 years, he has participated in all stages of drug Research & Development (R&D), encompassing preclinical discovery, clinical development, regulatory filing, and translational medicine. A key area of his research involves the application of AI/ML to drug R&D. He has co-authored more than 30 peer-reviewed journal papers, in addition to numerous conference abstracts and posters.

Subha Madhavan, Ph.D.

Vice President and Head

AI/ML, Quantitative and Digital Sciences, Research and Development

Pfizer Inc.

Subha Madhavan, Ph.D., is a dynamic and results-driven leader with a strong track record of excellence in organizations that operate at the nexus of science, technology and business. She has initiated and successfully directed several productive clinical research and development programs at the Georgetown Lombardi Comprehensive Cancer Center, MedStar hospital network, FDA, NIH and BioPharma industry. She was co-leader of the FDA's Center for Excellence in Regulatory Science and worked with the oncology and vaccine teams. She was an advisory member to the Biden Foundation's Cancer Moonshot Program and advised on pre-competitive data sharing initiatives across pharma, health tech companies and research organizations to drive innovation. She has been recognized for her work through several awards including the Service to America award in the Science and Environment category (2005), Research Acceleration Award by AACR and Pancreatic Cancer Action Network (2015), and Women in Tech Global award (2021). She is currently the Head of Clinical AI/ML & Digital Sciences at Pfizer worldwide R&D where she leads a team focused on advancing precision therapies across multiple treatment areas including Anti-Infectives, Oncology, Immunology & Inflammation among others.

Anthony Philippakis, M.D., Ph.D.

Chief Data Officer

Broad Institute of MIT and Harvard

Co-Director, Eric & Wendy Schmidt Center

Anthony Philippakis, M.D., Ph.D., is the chief data officer of the Broad Institute of MIT and Harvard, and the co-director of the Eric and Wendy Schmidt Center.

He trained as a cardiologist at Brigham and Women's Hospital, with a focus on rare genetic cardiovascular diseases. At the Broad Institute he is the founding director of the Data Sciences Platform, an organization of over 200 software engineers and computational biologists that develops software for analyzing genomic and clinical data. In addition to his roles at the Broad Institute and Brigham and Women's Hospital, Philippakis is a venture partner at GV, focusing on machine learning, distributed computing, and genomics. Philippakis received his M.D. from Harvard Medical School and completed a Ph.D. in biophysics at Harvard. As an undergraduate, he studied mathematics at Yale University, and later completed Part III (equivalent to M.Phil) in mathematics at Cambridge University.

Hoifung Poon, Ph.D.

General Manager

Microsoft Health Futures

Hoifung Poon, Ph.D., is the General Manager at Microsoft Health Futures. His research interest is in developing next-generation AI technology to accelerate progress in access, safety, and preventative care for precision health. At Microsoft, he leads biomedical AI research and incubation, with a particular focus on scaling real-world evidence generation by structuring all medical data. He obtained a B.S. with Distinction in Computer Science from Sun Yat-Sen University, and a Ph.D. in Computer Science and Engineering (my dissertation) from University of Washington. He is an affiliated professor at University of Washington Medical School, and serves as co-PI for various academic projects such as DARPA Big Mechanisms. His past work spans diverse topics in machine learning and NLP, and has been recognized with Best Paper Awards in top conferences such as NAACL, EMNLP, and UAI. For more information, check out his publications and LinkedIn profile.

Pranav Rajpurkar, Ph.D.

Assistant Professor

Harvard University

Pranav Rajpurkar, Ph.D., is an Assistant Professor at Harvard University and a researcher in the field of medical artificial intelligence. With a focus on medical image interpretation, Dr. Rajpurkar's research lab strives to develop AI models that can match the proficiency of top-tier medical doctors. His research group is at the forefront of developing "Generalist Medical AI" systems that can closely resemble doctors in their ability to reason through a wide range of medical tasks, incorporate multiple data modalities, and communicate in natural language. Dr. Rajpurkar has published more than 65 academic papers which have received over 19K citations, including in *Nature*, *NEJM*, and *Nature Medicine*. His work has been covered by media outlets including NPR, *The Washington Post*, and *Wired*. In 2022, Dr. Rajpurkar has been recognized as a recipient of Forbes 30 Under 30 and as a Nature Medicine Early-career Researcher to Watch. Dr. Rajpurkar leads educational initiatives including the Harvard-Stanford Medical AI Bootcamp Program, and CS197: AI Research Experiences at Harvard. Prior to starting at Harvard in 2021, Dr. Rajpurkar earned his B.S., M.S., and Ph.D. degrees, all in Computer Science from Stanford University.

Omer Reingold, Ph.D.

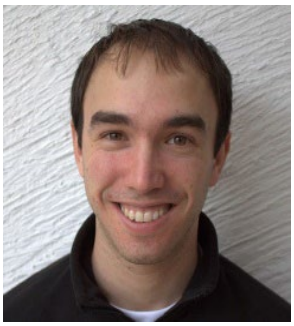
Professor

Computer Science

Stanford University

Omer Reingold, Ph.D., is the Rajeev Motwani professor of computer science at Stanford University and the director of the Simons Collaboration on the Theory of Algorithmic Fairness. Past positions include the Weizmann Institute of Science, Microsoft Research, the Institute for Advanced Study

in Princeton, NJ, AT&T Labs and Samsung Research America. His research is in the foundations of computer science and most notably in computational complexity, cryptography and the societal impact of computation. He is an ACM Fellow and a Simons Investigator. Among his distinctions are the 2005 Grace Murray Hopper Award and the 2009 Gödel Prize.



David Sontag, Ph.D.

Professor

Electrical Engineering and Computer Science

MIT

David Sontag, Ph.D., is a Professor of Electrical Engineering and Computer Science at MIT, part of the Institute for Medical Engineering & Science, the Computer Science and Artificial Intelligence Laboratory, and the J-Clinic for Machine Learning in Health. His research focuses on advancing machine learning and artificial intelligence, and using these to transform health care. Previously, he was an Assistant Professor of Computer Science and Data Science at New York University.

Marc Adam Suchard, Ph.D.

Professor

Biomathematics & Human Genetics

David Geffen School of Medicine

Department of Biostatistics

UCLA Fielding School of Public Health

Marc Adam Suchard, M.D., Ph.D., is Professor in the Departments of Biomathematics and of Human Genetics at the David Geffen School of Medicine at UCLA and in the Department of Biostatistics in the UCLA Fielding School of Public Health at the University of California, Los

Angeles. He was elected as a Fellow of the American Statistical Association in 2012, and he received the COPSS Presidents' Award in 2013.

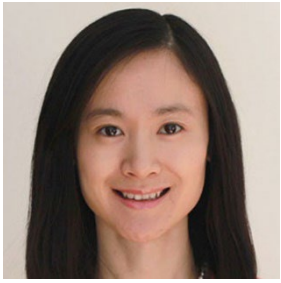
Stefan Wager, Ph.D.

Associate Professor

Operations, Information and Technology

Stanford University

Stefan Wager, PhD., is an associate professor of operations, information and technology at Stanford Graduate School of Business, and an associate professor of statistics (by courtesy). He received his PhD in statistics from Stanford University in 2016, and also holds a BS (2011) degree in mathematics from Stanford. Dr. Wager's research lies at the intersection of causal inference, optimization, and statistical learning. He is particularly interested in developing new solutions to classical problems in statistics, economics and decision making that leverage recent developments in machine learning.



Yuanjia Wang, Ph.D.

Professor

Biostatistics and Psychiatry

Columbia University

Dr. Yuanjia Wang, Ph.D., is a Professor in the Department of Biostatistics and Department of Psychiatry at Columbia University, and a core member of the Division of Biostatistics at New York State Psychiatry Institute. She was elected as a Fellow of the American Statistical Association (ASA) in 2016. Dr. Wang works on developing data-driven approaches to explore relationship between biomarkers, clinical markers, and health outcomes to assist discoveries in disease etiology, and increase diagnostic capabilities of psychiatric and neurological diseases. Her methodological interests include statistical learning, analytics for

personalized medicine, network analysis, and novel design and analysis of clinical trials. Her substantive research area of interest includes psychiatric disorders and neurological disorders.

Petrina Kamya, Ph.D.

Vice President

Global Head of AI Platforms

President of Insilico Medicine Canada

Petrina Kamya, Ph.D., is the VP, and Global Head of AI Platforms and President of Insilico Medicine Canada, overseeing Insilico's end-to-end generative AI-driven drug discovery platform, Pharma.AI. Before joining Insilico, Dr. Kamya's career spanned many key stages of the drug discovery value chain from the early stages to commercialization. She holds a Ph.D. in chemistry and a BS in biochemistry from Concordia University.

Bingzhi Zhang, Ph.D.

Biostatistical Team Leader

Sanofi

Bingzhi Zhang, Ph.D., is the Biostatistical Team Leader at Sanofi. Her work is dedicated to bridging the gaps between project execution and innovative study design and decision-making approaches, making use of emerging sources for evidence generation. Her research interests include study designs utilizing external/historical data, master protocols and AI/ML-powered study design. She received her Ph.D. in Biostatistics from Columbia University.

Tutorial Instructors, Poster Coordinators

Haoda Fu, Ph.D.

Associate Vice President

Enterprise Lead, ML/AI

Eli Lilly and Company

Haoda Fu, Ph.D., 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, University 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 has been 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 an 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.

Javier Cabrera, Ph.D.

Professor

Department of Statistics, Department of Medicine

Rutgers University

Javier Cabrera, Ph.D., is a Professor in the Department of Statistics and the Department of Medicine, Rutgers University, and a member of the Cardiovascular Institute of New Jersey and the Institute of Quantitative Biomedicine. He is a winner of the 2010 SPAIG award of the American Statistical Association, a Fulbright fellow, and a Henry Rutgers fellow. He was Director of the Institute of Biostatistics at Rutgers University and the chief co-editor of the journal, *Computational Statistics and Data Analysis*. Professor Cabrera has numerous publications and books in Statistics and Biostatistics on diverse topics, including Big Data for medical research, functional genomics, analysis of genomic data, statistical computing, graphics, and computer vision. He received his PhD from Princeton University.

Susan Gruber, Ph.D., M.P.H., M.S.

Co-Founder TL Revolution

Founder of Putnam Data Sciences

Dr. Susan Gruber, **Ph.D., M.P.H., M.S.**, is a biostatistician and computer scientist who founded Putnam Data Sciences, a statistical consulting firm specializing in causal inference and predictive modeling and co-founded TL Revolution with Dr. van der Laan. She is the former Director of the Biostatistics Center, Department of Population Medicine at Harvard Pilgrim Health Care and Harvard Medical School, and former Senior Director of the IMEDS Methods program at the Reagan Udall Foundation for the FDA. Dr. Gruber is an expert on targeted learning who developed the first open source R package for TMLE and has an extensive record of publications, presentations, and training sessions on Targeted Learning.

Nareen Katta, M.B.A, M.S.

Head, Data Science & Analytics

AbbVie

Nareen Katta, M.B.A, M.S., works as the Head of Data Science and Analytics at AbbVie. Nareen has over 20 years of experience in the pharmaceutical industry. In his current role, Nareen is responsible for building and executing the advanced analytics strategy, that covers both Scientific and Business Operations, across Clinical Development Continuum, Geostrategy and Study start-up, Centralized and Risk Based Monitoring, Site Engagement, Business Performance, Precision Medicine, Patient Safety and R&D. In addition, Nareen is actively engaged in evaluating the opportunities created by the technology trends like big data, automation, machine learning and AI, digital health etc. and strategically instantiating them at AbbVie to drive organizational transformation. Nareen has an MBA from The University of Chicago Booth School of Business and a MS in Electrical Engineering from University of Texas at Arlington.

Hana Lee, Ph.D., M.S.

Senior Staff Fellow

U.S. Food and Drug Administration

Hana Lee, Ph.D., is a Senior Statistical Reviewer of the Office of Biostatistics (OB) in the CDER, FDA. She leads and oversees various FDA-funded projects intended to support development of the agency's Real-World Evidence (RWE) program. She's been leading the RWE scientific working group of the ASA Biopharmaceutical Section, which is a public private partnership with FDA involving scientists from FDA, academia, and industry to advance the understanding of RWD/E to support regulatory decision-making. Dr. Lee has been

recognized for her outstanding contributions to the field with numerous awards from FDA. Most recently, she received the prestigious FDA Scientific Achievement Award, which honors FDA scientists who have made exceptional advancements in regulatory science and contributed significantly to FDA's mission of protecting public health.

Claire Zhao, Ph.D.

Associate Director

Group Head, Quantitative & Digital Sciences

Pfizer Inc.

Claire Zhao, Ph.D., is currently working for Pfizer as the Group Lead of Clinical AI/ML in AI/ML Quantitative and Digital Sciences (AQDS). She is based in Cambridge, MA. She works on building interpretable and interoperable AI/ML models that leverages multi-modal data to inform clinical development. In her previous work at Philips, she led the development of AI/ML solutions to improve clinical and operational workflow for healthcare workers and the hospital. She obtained her Ph.D. from the Department of Biomedical Engineering at Johns Hopkins University, with a thesis in Computational Medicine

Louisa H. Smith, Ph.D.

Assistant Professor

Department of Health Sciences

Northeastern University

Ge Cheng

Department of Statistics

Rutgers University

Mahan Dastgiri

Department of Statistics

Rutgers University

Yajie Duan

Department of Statistics

Rutgers University

Chloe R. Bennett, MPA

PhD Candidate

Population Health

Bouvé College of Health Sciences

Northeastern University

Victor M. Castro, MS, FAMIA

PhD Student

Bouvé College of Health Sciences

Northeastern University

Research Data Scientist, Mass General Brigham

Natalia Coriano-Díaz, MPH

Ph.D. Student

Bouvé College of Health Sciences

Northeastern University

Darien Mather, MPH

PhD Candidate

Population Health

Institute of Health Equity and Social Justice Research

Bouvé College of Health Sciences

Northeastern University

Victoria M. Nielsen, MPH

PhD Candidate

Population Health

Bouvé College of Health Sciences,

Northeastern University

Senior Surveillance Epidemiologist Massachusetts Department of Public Health

Mohana Saha

Master's Student

Regulatory Affairs

College of Professional Studies

Northeastern University

Abstracts

2024 Annual Symposium on Risks and Opportunities of AI in Pharmaceutical Medicine

Day 2 Plenary Session 1

Title:

"Empowering Insights in the *All of Us* Research Program: A Statistical Perspective on the Transformational Role of AI and ML", Qingxia 'Cindy' Chen, Biomedical Informatics, and Ophthalmology & Visual Sciences at Vanderbilt University Medical Center

Abstract:

The *All of Us* Research Program is building a diverse cohort of a million or more Americans engaged longitudinally in data sharing including biospecimens, electronic health records, surveys, and digital health technology to advance precision medicine research and fuel new insights into human health. This data is at the fingertips of researchers through the Researcher Workbench—a cloud-based analytical platform equipped with graphical interface tools and bolstered by Jupyter notebooks. Underpinned by the FAIR principles, these resources render data and discoveries Findable, Accessible, Interoperable, and Reproducible in the domain of biomedical research. The *All of Us* Research Program leverages cutting-edge AI and ML methods. These endeavors include devising models to enhance missing data with external resources and various data types, forecasting participant engagement, and refining polygenetic risk scores for underrepresented populations within biomedicine. At its core, the *All of Us* Researcher Workbench stands as a dynamic hub, fostering methodological innovation and catalyzing outcome research.

Title:

"The Generalist Medical AI Will See You Now", Pranav Rajpurkar, Ph.D., Harvard University

Abstract:

Accurate interpretation of medical images is crucial for disease diagnosis and treatment, and AI has the potential to minimize errors, reduce delays, and improve accessibility. The focal point of this presentation lies in a grand ambition: the development of 'Generalist Medical AI' systems that can closely resemble doctors in their ability to reason through a wide range of medical tasks, incorporate multiple data modalities, and communicate in natural language. Starting with pioneering algorithms that have already demonstrated their potential in diagnosing diseases from chest X-rays or electrocardiograms, matching the proficiency of expert radiologists and cardiologists, I will delve into the core challenges and advancements in the field. The discussion will navigate towards the topic of label-efficient AI models: with a scarcity of meticulously annotated data in healthcare, the development of AI systems capable of learning effectively from limited labels has become a key concern. In this vein, I'll delve into how the innovative use of self-supervision and pre-training methods has led to algorithmic advancements that can perform high-level diagnostic tasks using significantly less annotated data. Additionally, I will talk about initiatives in data curation, human-AI collaboration, and the creation of open benchmarks to evaluate the generalizability of medical AI algorithms. In summary, this talk aims to deliver a comprehensive picture of the state of 'Generalist Medical AI,' the advancements made, the challenges faced, and the prospects lying ahead.

Title:

"Generative AI for Case Adjudication in OHDSI", Marc Suchard, UCLA

Abstract:

Observational Health Data Science and Informatics (OHDSI) aims to improve health by empowering an open-science community to collaboratively generate evidence that promotes better health decisions and better care. OHDSI is actively exploring ways in which large language models (LLMs) can help deliver this aim through the Assessment of Pre-trained Observational Large Longitudinal models in OHDSI (APOLLO) project. This talk highlights two APOLLO directions. First, medical case validation stands as a necessary element of regulatory-grade evidence, but currently requires human adjudication that is time and resource-intensive. APOLLO provides open-source development of LLMs to improve the reliability and scalability of case validation. Across 10 diseases in two different data sources, we find substantial heterogeneity in agreement between human reviewers and that LLMs agree with humans as much as humans agree with each other. Second, APOLLO explores the use of LLMs pre-trained to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) data sources to perform patient-level prediction, where a pre-trained model may prove more accurate with less training data than current non-pre-trained models, missing value

imputation, almost identical to a bidirectional pre-training task, and counterfactual prediction. In this latter task, the LLM provides a prediction of what happens to a patient in the future given both treatment options. We suspect further applications will continue to emerge across the OHDSI community.

Day 2 Plenary Session 2

Title:

"Generative Mixed-Response State-Space Model for Analyzing Multi-Dimensional Digital Phenotypes", Yuanjia Wang, Columbia University

Abstract:

Digital technologies (e.g., mobile phones) can be used to obtain objective, frequent, and real-world digital phenotypes from individuals. However, modeling these data poses substantial challenges since observational data is subject to confounding and various sources of variabilities. For example, signals on patients' underlying health status and treatment effects are mixed with variation due to the living environment and measurement noises. The digital phenotype data thus shows extensive variabilities between and within patients as well as across different health domains (e.g., motor, cognitive, and speaking). Motivated by a mobile health study of Parkinson's disease (PD), we develop a mixed-response state-space (MRSS) model to jointly capture multi-dimensional, multi-modal digital phenotypes and their measurement processes by a finite number of latent state time series. These latent states reflect the dynamic health status and personalized time-varying treatment effects and can be used to adjust for informative measurements. We conduct comprehensive simulation studies and demonstrate the advantage of MRSS in modeling a mobile health study that remotely collects real-time digital phenotypes from PD patients. We discuss extensions to deep latent state-space models for generating digital phenotype time-series data to learn optimal treatment strategies.

Title:

"Machine Learning for Causal Inference", Stefan Wager, Stanford University

Abstract:

Given advances in machine learning over the past decades, it is now possible to accurately solve difficult non-parametric prediction problems in a way that is routine and reproducible. In this talk, I'll discuss how machine learning tools can be rigorously integrated into observational study analyses, and how they interact with classical statistical ideas around randomization, semiparametric modeling, double robustness, etc. I'll also survey some recent advances in methods for treatment heterogeneity. When deployed carefully, machine learning enables us to

develop causal estimators that reflect an observational study design more closely than basic linear regression-based methods.

Title:

“Leveraging the Power of Large Language Models (LLMs) by Statisticians for Pharmaceutical Research & Development: Illustrating with an Example AI App”, Junshui Ma, Head of the Biometrics Research Department, Merck Research Lab.

Abstract:

This presentation aims to explore the transformative potential of ChatGPT-like large language models (LLMs) in the realm of pharmaceutical research and development (R&D), with a specific focus on its application by statisticians. The intricate nature of pharmaceutical R&D necessitates that statisticians in this industry consistently work within multidisciplinary teams. They are expected to rapidly acquire a comprehensive understanding of the scientific and medical domains they engage with. LLMs have showcased a broad spectrum of skills, such as drafting and refining messages, translating across various languages, elucidating and summarizing documents, programming in multiple computer languages, strategizing, and reasoning for problem-solving. They also possess knowledge in numerous fields, including statistics, science, and clinical medicine, among others. Statisticians who can harness this technology have demonstrated a significant increase in efficiency in fulfilling their roles within the team. Real-world cases showcasing the successful implementation of this technology in various settings will be presented. The presentation will conclude with a discussion on strategies to address general issues related to the use of LLMs.

Day 2 Plenary Session 3

Title:

“The Role of Targeted Machine Learning in a Causal Roadmap for Generating High-Quality Real-World Evidence”, Lauren Elizabeth Eyler Dang, National Institute of Allergy and Infectious Diseases Biostatistics Research Branch

Abstract:

A broad spectrum of studies use real-world data (RWD) to produce real-world evidence (RWE), ranging from randomized controlled trials with outcomes assessed using RWD to fully observational studies. The statistical target parameter that answers the research question in such studies is usually not equivalent to a coefficient in a potentially misspecified parametric model; machine learning algorithms can help to avoid parametric modeling assumptions but must be implemented in a way that provides reliable inference. The Targeted Machine Learning approach is aimed at efficiently and consistently estimating target parameters that directly

answer research questions through a rigorous, pre-specified process that minimizes statistical assumptions. This approach involves estimating components of the observed data likelihood that are relevant for the selected parameter using data-adaptive ensemble machine learning (Super Learning). Targeted maximum likelihood estimation (TMLE) is then used to target initial estimates to optimize the bias-variance tradeoff for the target parameter of interest while providing accurate inference. Petersen and van der Laan's Roadmap for Causal Inference provides a unifying framework for translating causal questions and true knowledge into a statistical parameter that may be estimated using a targeted approach. This talk will use real and simulated data examples to introduce fundamental concepts in Targeted Learning, including the Causal Roadmap, Super Learning, and TMLE applied to estimation of a causal average treatment effect. Examples of more complex target parameters and a brief overview of software for user-friendly implementation of these methods through the *tlverse* software ecosystem will be given.

Title:

"The multitude of group affiliations: Algorithmic Fairness, Loss Minimization and Outcome Indistinguishability", Omer Reingold, Stanford University

Abstract:

We will discuss a rather recent and very fruitful line of research in algorithmic fairness, coined multi-group fairness. We will focus on risk prediction, where a machine learning algorithm tries to learn a predictor to answer questions of the form "what is the probability that patient x will have a particular adverse reaction to a drug?" Training a risk predictor to minimize a loss function fixed in advance is the dominant paradigm in machine learning. However, global loss minimization may create predictions that are mis-calibrated on sub-populations, causing harm to individuals of these populations. Multi-group fairness tries to prevent forms of discrimination to a rich (possibly exponential) collection of arbitrarily intersecting groups. In a sense, it provides a computational perspective on the meaning of individual risks and the classical tension between clinical prediction, which uses individual-level traits, and actuarial prediction, which uses group-level traits. Surprisingly, it also provides unique robustness properties to the learned risk predictors, such as robustness to distributional shifts and to changing objectives. Multi-group fairness has found practical usages in health care, which we will discuss as well.

Based on a sequence of works joint with (subsets of) Cynthia Dwork, Shafi Goldwasser, Parikshit Gopalan, Úrsula Hébert-Johnson, Lunjia Hu, Adam Kalai, Christoph Kern, Michael P. Kim, Frauke Kreuter, Guy N. Rothblum, Vatsal Sharan, Udi Wieder, Gal Yona and others.

Modeling Covid-19 Immunological Reactions and Clinical Susceptibility in the Context of Long-Term Follow-up of a Prospective Cohort of Healthcare Workers, Noam Barda, Ben-Gurion University. *(25 min)*

Title:

“Designs for Observational Vaccine Effectiveness Studies”, Noam Barda, Ben-Gurion University

Abstract:

The Sheba healthcare worker cohort is a prospective cohort of ~15,000 healthcare workers that intensively monitors Covid-19 infections, symptoms, and vaccinations. The cohort’s most unique characteristic is, however, its frequent tracking of participants’ immune response with monthly serological tests of antibody levels, neutralizing antibody response, and cell-mediated immunity markers. Many studies based on the cohort’s data have been published in prestigious journals, and have helped direct public health policy. Previous studies based on the cohort’s data have shown that neutralizing antibodies are strong correlates for protection, and that immune response differs substantially between individuals. We now hypothesize, that with over two years of clinical and immunological data, spanning multiple SARS-CoV-2 variants and vaccine types, data from the cohort can be used to predict an individual’s immune response to vaccination and infection, and to recommend optimal times for additional vaccination.

Title:

“Drugs in clinical trials out of generative AI”, Petrina Kamya, Insilico Medicine.

Abstract:

Generative AI is not new. GPT3 was published in mid-2020. In biomedicine, the generative AI revolution started around 2017 with generative adversarial networks and in 2018 with transformers. Around 2019, the technology matured enough to achieve molecular precision and companies started the first discovery and development programs. Some of these therapeutics have now reached human clinical trials and some are in Phase II.