

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Doing More with Less
Efficient Approaches to Data Scarcity
A Symposium

The Committee on Applied and Theoretical Statistics
June 12, 2025
1:30 - 5:00 p.m. ET

Register for and watch this virtual symposium:

https://events.nationalacademies.org/45053_06-2025_doing-more-with-less-efficient-approaches-to-data-scarcity

1:30 PM* Welcome and Introduction

Lance Waller, Co-chair of the Committee on Applied and Theoretical Statistics (CATS)

1:40 PM Panel Presentations

Q&A moderated by *Scott Holan*, CATS Member

- Tianxi Cai: 20 minutes (1:40 - 2:00 PM)
- Q&A for Tianxi: 5 minutes (2:00- 2:05 PM)
- Jane Pinelis Presentation: 20 minutes (2:05 - 2:25 PM)
- Q&A for Jane: 5 minutes (2:25 - 2:30 PM)

2:30 - 2:50 PM: BREAK

2:50 PM Panel Presentations (continued)

Q&A moderated by *Scott Holan*

- Lin Ge Presentation: 20 minutes (2:50 - 3:10 PM)
- Q&A for Lin: 5 minutes (3:10 - 3:15 PM)
- Aleksandra (Seša) Slavković Presentation: 20 minutes (3:15 - 3:35 PM)
- Q&A for Seša: 5 minutes (3:35 - 3:40 PM)

3:40 - 4:00 PM: BREAK

4:00 PM Panel Discussion

Moderated by *Scott Holan*

5:00 PM ADJOURN

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

SPEAKER BIOS

Lance Waller, Ph.D. is a professor in the Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University. He is a member of the National Academy of Science Board on Mathematical Sciences and Analytics and has served on National Academies Committees on Applied and Theoretical Statistics, Cancer Near Nuclear Facilities, Geographic Assessments of Exposures to Agent Orange, and Standoff Explosive Technologies. His research involves the development of statistical methods for geographic data including applications in environmental justice, epidemiology, disease surveillance, spatial cluster detection, conservation biology, and disease ecology. His research appears in biostatistical, statistical, environmental health, and ecology journals and in the textbook *Applied Spatial Statistics for Public Health Data* (2004, Wiley). Dr. Waller currently leads planning for the Data Science Initiative in Emory University's Woodruff Health Sciences Center.

Dr. **Scott H. Holan** is a Professor of Statistics and Department Chair at the University of Missouri and serves as a Senior Research Fellow in the Research and Methodology Directorate at the U.S. Census Bureau. His research expertise includes developing statistical and machine learning methodology for dependent data (spatial, spatio-temporal, functional, and multivariate, among others), Bayesian methods, environmental and ecological statistics, official statistics, and survey methodology. He is an elected Fellow of the American Statistical Association (2014), an Elected Member of the International Statistical Institute (2017), and an elected Fellow of the Institute of Mathematical Statistics (2021). Prof. Holan was a previous co-awardee of the Statistical Partnerships Among Academe, Industry, and Government (SPAIG) Award (2017). Prof. Holan has an M.S. in mathematics from University of Illinois at Chicago (1999) and a Ph.D. in Statistics from Texas A&M University (2004).

Tianxi Cai is a major player in developing analytical tools for EHR phenotyping and predictive modeling with biomedical data. She provides statistical leadership on several large-scale projects, including the NIH-funded BD2K PIC-SURE Center of Excellence and N-GRID Center for Excellence in Genomic Science to study neuropsychiatric disease at DBMI. In addition to her collaborative work, Cai's research lab in the Department of Biostatistics at the Harvard T.H. Chan School of Public Health 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.

Jane Pinelis currently serves as the chief AI engineer of the Applied Information Sciences Branch at Johns Hopkins University's Applied Physics Laboratory (JHU/APL). She leads a diverse group of AI scientists and analysts in the development, assurance, and integration of AI capabilities. Her leadership extends to the formulation of cutting-edge assurance-specific products and stringent standards designed to fortify testing of AI-driven systems across the multifaceted landscape of the U.S. Department of Defense

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

(DOD). Dr. Pinelis holds a BS in statistics, economics, and mathematics, an MA in statistics, and a PhD in statistics, all from the University of Michigan, Ann Arbor. Prior to her current role, she served as the inaugural chief of AI assurance at the Chief Digital and Artificial Intelligence Office (CDAO) and the Joint Artificial Intelligence Center (JAIC) at the DOD, where she oversaw the test and evaluation (T&E) and responsible AI (RAI) directorates. Her career has largely focused on operational T&E, both in support of the service operational testing commands and at the Office of the Secretary of Defense level.

Dr. **Lin Ge** is currently a Postdoctoral Research Fellow at the Harvard T.H. Chan School of Public Health and will join Indiana University Bloomington School of Public Health as an Assistant Professor in July 2025. He received his Ph.D. degree in Biostatistics from Rollins School of Public Health at Emory University in 2023. His research focuses on addressing misclassification and measurement errors in disease surveillance, cancer epidemiology and environmental health. By leveraging electronic health record (EHR) and employing innovative sampling strategies, he aims to advance both biostatistical methodology and epidemiological applications to improve public health outcomes.

Aleksandra (Seša) Slavković is a Professor of Statistics and Associate Dean for Graduate Education in Eberly College of Science at Penn State. She received her PhD (2004) and M.S. (2001) in Statistics, and a Master of Human-Computer Interaction (1999) from Carnegie Mellon University. She received her B.A. in Psychology from Duquesne University (1996). Her research interests include methodological developments in the area of data privacy and confidentiality in the context of small and large scale surveys, health, genomic, and network data. Her focus is on the interplay of tools from statistics and computer science that leads to formal privacy protection -such as differential privacy- and broad data access, but also offers guarantees of accurate statistical inference needed to support reliable science and policy. Other past and current research interests include evaluation methods for human performance in virtual environments, statistical data mining, application of statistics to information sciences and social sciences, algebraic statistics, and causal inference.
