

IDWSDS 2025



October 14, 2025

International Day of Women in Statistics and Data Science

Thriving in Your Environment

2025 INTERNATIONAL DAY OF WOMEN IN STATISTICS AND DATA SCIENCE

CONFERENCE PROGRAM AND ABSTRACTS

OCTOBER 14TH, 2025

A VIRTUAL 24 HOUR CONFERENCE

[HTTPS://IDWSDS.ORG](https://idwsds.org)

ORGANIZED BY

CAUCUS FOR WOMEN IN STATISTICS AND DATA SCIENCE

HOSTED BY

CAUCUS FOR WOMEN IN STATISTICS AND DATA SCIENCE VIA ZOOM

© 2025

CAUCUS FOR WOMEN IN STATISTICS AND DATA SCIENCE

[HTTPS://WWW.CWSTAT.ORG](https://www.cwstat.org)



Welcome



Welcome to the IDWSDS Conference!

We are delighted to welcome you to this year's International Workshop on Statistical Data Science (IDWSDS), where our theme is "**Thriving in Your Environment**." This theme reflects our commitment to fostering innovation, adaptability, and collaboration in the ever-evolving landscape of data science and statistics. We are presenting talks that showcase the many ways we can thrive when our individual work, workplace or academic atmosphere, or larger world can make it difficult to collaborate or advance.

Whether you're joining us for the first time or returning, we're excited to offer an enriching experience. This year's program features a wide range of sessions designed to inspire, challenge, and connect our community of researchers, practitioners, and thought leaders.

What's New This Year?

• Conference App – Sched

We're going digital! All session information, speaker bios, and scheduling tools are now available through our conference app, **Sched**. Be sure to download it to customize your agenda, receive real-time updates, and connect with fellow attendees. We heard your request for networking and **Sched** is the way to meet others.

• Speed Sessions

New this year, our Speed Sessions offer fast-paced, focused presentations that highlight emerging ideas and innovative approaches. These are a great way to sample a variety of topics in a short time. Support these presenters by attending our speed sessions, then contact them to learn more about their work.

Acknowledging Our Sponsors

We are proud to recognize our generous sponsors:

Platinum Level

- CANSSI

Silver Level

- The International Society for Bayesian Analysis (ISBA)
- Medical Research Council Biostatistics Unit - University of Cambridge
- Stata

Their support helps make this conference possible. Be sure to check out their **sponsored sessions** throughout the program for exciting insights and opportunities to engage.

Stay Connected

We encourage you to take full advantage of the networking opportunities, interactive sessions, and collaborative spirit that define IDWSDS. Whether you're presenting, attending, or simply exploring, we hope you leave feeling energized and equipped to thrive in your own environment.

Welcome and enjoy the conference!

Cynthia Bland and Jessica Kohlschmidt

Co-chairs, 2025 International Day of Women in Statistics and Data Science Organizing Committee



Organizing Committee

CO-CHAIR



CYTHIA BLAND

Caucus for Women in Statistics and Data Science

Cynthia Bland is the Past President of the Caucus for Women in Statistics and Data Science. This year she has chaired the nominating committee and an award committee. Professionally she is a survey statistician and directs the Center for Official Statistics at RTI International, a non-profit research firm specializing in government contracting. In her two decades at RTI, she has designed and analyzed national surveys, with an emphasis on patient experience of care surveys. Cynthia leads a center of more than 70 statisticians as they innovate for clients and grow in their own careers. Cynthia serves as the Deputy Editor in Chief of RTI Press, an open-access, peer-reviewed journal, has been active with the NC Chapter of the American Statistical Association, and is a lecturer in statistics for the Kenan-Flagler School of Business at the University of North Carolina at Chapel Hill. Outside of the statistical world, she serves on the board of a local credit union and enjoys ballet.

<https://www.researchgate.net/profile/Ligia-Rodrigues-2>


LÍGIA HENRIQUES-RODRIGUES

University of Évora and CIMA

Lígia Henriques-Rodrigues (LHR) is an Assistant Professor in the Department of Mathematics at the School of Science and Technology, University of Évora. LHR completed a postdoctoral fellowship in Statistics of Extremes at the Faculty of Sciences, University of Lisbon, and earned her Ph.D. in Probability and Statistics from the same university. Her primary research focuses on the statistics of extremes and computational statistics, particularly in developing new classes of semi-parametric reduced-bias estimators for extreme event parameters. The applications of her research span environmental and financial sciences. Beyond her academic pursuits, LHR has been a board member of the Portuguese Statistical Society (SPE) since January 2024 and is a member of the executive committee for the Master's Degree in Statistical Modeling and Data Analysis at the University of Évora.

CO-CHAIR



JESSICA KOHLSCHMIDT

Caucus for Women in Statistics and Data Science

Dr. Jessica K. Kohlschmidt is a distinguished biostatistician, educator, and leader in the statistical community. A first-generation college student, her early passion for math and teaching evolved into a dedicated career in statistics after discovering its real-world applications.

She has built a multifaceted career that combines teaching, high-impact clinical research, and significant leadership roles in professional organizations. For 12 years, she applied her expertise to leukemia outcomes research at The Ohio State University, and she has been a college educator since 1999.

Dr. Kohlschmidt is deeply committed to mentorship and outreach, particularly encouraging young women to pursue statistics, and holds key leadership positions in several statistical organizations, including the American Statistical Association (ASA).

<https://fisher.osu.edu/people/kohlschmidt.1>


SHILI LIN

The Ohio State University

Shili Lin is a Professor of Statistics at the Ohio State University. Her research interests lie in the development and application of statistical methods to multi-omics data from cell lines, populations, and family samples. Shili is an active member of and serves the statistical profession in various capacities, including editorial service for various journals (e.g. current Associate Editor (AE) of Biometrics and former AE of the Journal of the American Statistical Association), serving as panel members in NIH Study Sections (e.g. former standing member of the Biostatistical Methods and Research Design), and serving professional organizations such as the American Statistical Association (ASA), the Institute of Mathematical Statistics (IMS), the International Biometrics Society (IBS), and the International Statistical Institute (ISI). She also served as the 2018 Caucus for Women in Statistics (CWS) President and is currently serving in the Board of Directors of the Canadian Statistical Sciences Institute (CANSSI). Shili is a co-founder and the president of the Florence Nightingale Day for Statistics and Data Science. She is a Fellow of the ASA, a Fellow of the IMS, a Fellow of the American Association for the Advancement of Science (AAAS), and an elected member of the ISI.

<https://www.bioss.ac.uk/people/altea>


ALTEA LORENZO-ARRIBAS

Biomathematics and Statistics Scotland (BioSS)

Altea is a socio-economic statistician at Biomathematics and Statistics Scotland (BioSS) working in collaboration with researchers at the Scottish Environment, Food and Agriculture Research Institutions. She is an elected council member of the Royal Statistical Society, member of the Society's AI Task Force, chair of the Celebrating Diversity Special Interest Group and the secretary of the History of Statistics Section, as well as a member of the Women Committee of the Spanish Society of Statistics and Operations Research (SEIO), and the Spanish Biostatistics Network (Biostatnet).



Organizing Committee

continued...



SARAH LOTSPEICH

Wake Forest University

Sarah Lotspeich is an Assistant Professor in Statistical Sciences at Wake Forest University, with a secondary appointment in Biostatistics and Data Science at the Wake Forest University School of Medicine. She is enthusiastic about mentoring student research and co-leads the Spatial and Environmental Statistics in Health (SESH) Lab at Wake Forest and the Missing and INcomplete Data (MIND) Lab at the University of North Carolina (UNC) at Chapel Hill. She co-organizes Florence Nightingale Day at Wake Forest annually, engaging local students in statistics and data science, and holds elected positions in the Caucus for Women in Statistics and Data Science and other organizations. Sarah completed a postdoctoral fellowship in Biostatistics at UNC Chapel Hill and earned her Ph.D. in Biostatistics from Vanderbilt University. Her research tackles challenges in analyzing error-prone observational data, focusing on international HIV cohorts, electronic health records, and health disparities. She also develops methods for statistical modeling with censored covariates, applicable to Huntington's disease. When she's not writing code, you can find Sarah cross-stitching, adventuring in new places, or rewatching her favorite TV shows.



VANDA M. LOURENÇO

NOVA University of Lisbon

Vanda M. Lourenço (VML) is an Assistant Professor in the Department of Mathematics at the NOVA School of Science and Technology, part of the NOVA University of Lisbon. She earned her PhD in Statistics and Stochastic Processes from Instituto Superior Técnico at the University of Lisbon. Her primary research focuses on robust, non-parametric, and computational statistics, particularly in genetic and genomic association studies and the prediction of quantitative traits. While her expertise has been predominantly applied to challenges in plant breeding, she is also keen on extending these methodologies to animal and human studies. In addition to her academic work, VML actively contributes to professional and statistical communities. She served as President of the Supervisory Board of the Portuguese Statistical Society (SPE) from 2021 to 2024 and currently represents Portugal in the Caucus for Women in Statistics (CWS), where she is also a member of the Nominations Committee. Additionally, she is an associate editor for the Journal of Data Science, Statistics, and Visualization (JDSSV), a publication of the ISI International Association for Statistical Computing. Beyond her formal roles, she is involved in various other initiatives within and outside statistical societies.

<https://www.linkedin.com/in/hunna-watson-52030352/>



HUNNA WATSON

University of North Carolina at Chapel Hill

Dr. Hunna Watson is a Research Associate Professor in the School of Medicine at the University of North Carolina at Chapel Hill, where she serves as a biostatistician in the Center of Excellence for Eating Disorders. She also holds Adjunct Research Fellow appointments in the School of Medicine at The University of Western Australia and in the Discipline of Psychology at Curtin University in Australia. Specializing in the field of eating disorders, Dr. Watson has authored over 120 peer-reviewed scientific articles and book chapters, focusing on critical areas such as diagnostic nosology, prevention and treatment, molecular genetics, and randomized clinical trials. Her research has gained global recognition through lectures, papers, and workshops presented worldwide. Dr. Watson serves on the editorial board of the International Journal of Eating Disorders and is a member of the Alumni Group of the National Eating Disorders Collaboration Steering Committee in Australia. Dr. Watson completed her PhD in Clinical Psychology at Curtin University and holds Master's degrees in Biostatistics from the University of Melbourne and Clinical Psychology from Curtin University.



Keynote Sessions



Session Info

K1

KEYNOTE SESSION

October 14th, 00:30 - 1:30 UTC

The Three Rs: Reliable, Rigorous and Robust

USE SCHED TO ACCESS THE ZOOM ROOM

Thriving in any environment necessitates the application of reliable, rigorous, and robust scientific principles. As statisticians and data scientists, we play a crucial role in ensuring that evidence-based decision-making is deliberate and supported by principles that yield objective evidence. Our expertise uniquely positions us to contribute across various domains, particularly in health policy.

In this talk, I will share my experiences as a statistician working within multi-disciplinary teams to analyze the impact of environmental factors on mental and physical health. I will provide concrete examples from health policy research that demonstrate the significant contributions of statistical science. By employing robust study designs, we can effectively evaluate the outcomes of environmental interventions on health. Through innovative modeling techniques for observational data, we can identify improvements in health and well-being. Moreover, by leveraging comprehensive primary data collection and community- engagement, we can explore how socioeconomic factors influence individuals' capacity to thrive in their environments.

Drawing from these successes, I will also highlight future opportunities for statisticians to make impactful contributions in critical areas where the stakes are high. Involving statistical expertise will be essential for developing effective policy solutions that promote thriving communities.

ORGANIZER: Jessica Kohlschmidt, Caucus for Women in Statistics and Data Science

CHAIR: Jessica Kohlschmidt, Caucus for Women in Statistics and Data Science



Speaker Bios

K1



Madhumita (Bonnie) Ghosh-Dastidar

Senior statistician, RAND; Head of RAND Statistics Group; President of the American Statistical Association

Bonnie Ghosh-Dastidar is a senior statistician at RAND working in public policy; she also leads the RAND Statistics Group. Ghosh-Dastidar's experience includes being lead statistician for large-scale surveys on sexual assault and harassment in the U.S. military, as well as for the first Singapore mental health survey to assess national prevalence. Her current collaborations center on multiple National Institutes of Health studies leveraging quasi-experimental design to evaluate the effect of neighborhood-level change on residents' diet, sleep, and cardiovascular health. She is a fellow of the ASA and has been recognized for her collaboration, leadership and mentoring. she was recently honored by the Penn State Eberly College of Science with the 2025 Outstanding Science Alumni award. Ghosh-Dastidar is the 119th president of the American Statistical Association.



Session Info

K2

KEYNOTE SESSION

October 14th, 07:00 - 8:00 UTC

Building Power with Code and Data: Tools for Thriving in Any Environment

USE SCHED TO ACCESS THE ZOOM ROOM

Thriving in today's world often means navigating spaces where power has not traditionally been shared equally. Data and code can help shift that balance. They provide tools to ask sharper questions, monitor conditions, and build evidence that drives change. In this talk, I will share examples from my experiences as a journal editor and R package developer to show how coding and data literacy are more than technical skills—they are forms of agency and empowerment. Used well, they become tools for creating opportunities, amplifying voices, and shaping decisions in any environment.

ORGANIZER: Hunna Watson, University of North Carolina at Chapel Hill

CHAIR: Altea Lorenzo-Arribas, Biomathematics and Statistics Scotland (BioSS)



Speaker Bios

K2



Dr. Dianne Cook

Monash University

Dianne Cook is a Professor of Statistics in the Department of Econometrics and Business Statistics at Monash University in Melbourne, Australia. She holds a PhD in Statistics from Rutgers University. Her research focuses on statistical graphics, especially interactive visualisation of high-dimensional data and statistical inference for visualisation, and she incorporates these topics into her teaching. Di is a strong advocate for open source software, ethical data practices, and gender equity. Many of the R packages she contributes to are named after Australian wildlife—highlighting both a connection to place and a celebration of the country's unique and threatened species.

Session Info



KEYNOTE SESSION

October 14th, 08:00 - 9:00 UTC

Why I Write Biographies

USE SCHED TO ACCESS THE ZOOM ROOM

In my talk, I will explain why and how I have written the two existing book-length biographies of twentieth century women statisticians. I will introduce you to the subjects of these biographies, Elizabeth Scott and F. N. David, and why they are celebrated in our profession. I will show you how each did what was expected of them in order to succeed, but then took a road less traveled for statisticians. My hope is that, by the end of my talk, you will begin to ask yourself, "What will be my own road less travelled?"

K3



Speaker Bios

K3



Amanda L. Golbeck

Professor of Biostatistics and Associate Dean for Academic Affairs in the Fay W. Boozman College of Public Health at the University of Arkansas for Medical Sciences; and Professor Emerita in the Department of Mathematics and Statistics at San Diego State University

Amanda L. Golbeck is Professor of Biostatistics and Associate Dean for Academic Affairs in the Fay W. Boozman College of Public Health at the University of Arkansas for Medical Sciences; and Professor Emerita in the Department of Mathematics and Statistics at San Diego State University. The Committee of Presidents of Statistical Societies gave her the Elizabeth L. Scott Award for her outstanding efforts in enhancing the status of women and minorities, fostering new leadership opportunities for women and men, promoting diversity at all levels, and advocating for a more inclusive, open and supportive atmosphere in statistical sciences. The American Statistical Association gave her a Fellow Award for her highly influential leadership, especially for her significant organizational and system changes in several academic institutions; for her exemplary mentorship of junior faculty and students; for her pioneering work in health numeracy; and for important contributions to research in public health and medicine. Grinnell College gave her an Alumni Award for energizing her colleagues to initiative creative academic programming and exercising strategic planning skills, fact-finding, probing and analysis to aid in the development of academic innovations. She also received a Fulbright Specialist Award to the University of Latvia at Riga. Golbeck has over 140 publications, is a member of the editorial board of Significance Magazine, and has four books: (1) Leadership and Women in Statistics (2016, Chapman and Hall/CRC, edited with I. Olkin and Y. Gel), (2) Equivalence: Elizabeth L. Scott at Berkeley (2017, Chapman and Hall/CRC), (3) Leadership in Statistics and Data Science: Planning for Inclusive Excellence (2021, Springer, edited), and (4) Florence Nightingale David: A Passionate Probabilist, Statistician, Historian, and Leader (2025, Springer). The edited books are the only published books on leadership in statistics, and the authored books are the only published book-length biographies of 20th Century women in statistics. Golbeck earned her PhD in Biostatistics from the University of California at Berkeley after earning master's degrees there in anthropology and in statistics.



Session Info

K4

KEYNOTE SESSION

October 14th, 17:00 - 18:00 UTC

Evidence-Based Leadership: Using Statistical Thinking to Transform Higher Education

USE SCHED TO ACCESS THE ZOOM ROOM

How can statistical thinking become a transformative leadership tool in higher education? This presentation demonstrates how embracing data-driven decision-making has revolutionized institutional effectiveness at the University of Belize, leading to successful, comprehensive policy reform and sustainable quality improvement systems.

Drawing from ten years of leadership experience, Dr. Enriquez-Savery will share how initially viewing leadership as separate from technical expertise evolved into recognizing statistical thinking as the foundation of transformational change. Through concrete examples, including accreditation preparation, development of evidence-based workload allocation models, and implementation of university-wide institutional effectiveness frameworks, attendees will learn practical strategies for leveraging data to drive institutional excellence. Stop asking if you're qualified for leadership and start asking how organizations can make good decisions without statistical thinking. Your analytical superpowers aren't a career limitation—they're a competitive advantage the world desperately needs.

Learning Objectives

By the end of this session, participants will be able to:

1. Apply Statistical Thinking to Strategic Planning: Understand how to use quantitative analysis to monitor strategic plan implementation and calculate meaningful completion rates
2. Design Evidence-Based Policy Frameworks: Learn methodologies for developing data-driven policies that ensure equitable resource allocation and transparent decision-making
3. Implement Institutional Effectiveness Systems: Gain practical knowledge for establishing systematic data collection and analysis processes that inform continuous improvement
4. Lead Accreditation Through Data: Understand how statistical evidence strengthens accreditation narratives and demonstrates institutional compliance with quality standards

ORGANIZER: IDWSDS Program Committee

CHAIR: Cynthia Bland, Caucus for Women in Statistics and Data Science





Sherlene Enriquez-Savery

University of Belize, VP Academics

Dr. Sherlene Enriquez-Savery serves as Vice President for Academics at the University of Belize, a position she has held since August 2021. Her distinguished career at the university spans over two decades, beginning in August 2003, during which she has demonstrated exceptional leadership across multiple academic and administrative roles.

Dr. Enriquez-Savery earned her Bachelor's degree in Mathematics from the University of Belize, followed by a Master of Science in Mathematics from Clemson University in South Carolina. She further advanced her academic pursuits with a Master of Arts degree in Statistics from the University of South Florida and culminated her distinguished educational journey with a Doctor of Philosophy in Mathematics and Statistics, also from the University of South Florida.

Her career trajectory demonstrates exemplary progressive leadership development:

Mathematics/Statistics Lecturer (2003-2008, 2012-2016), Chair of Mathematics, Physics and Information Technology (2015-2018), Dean of Faculty of Science and Technology (2020-2021), and Vice President (2021-present). She brings invaluable international experience from her distinguished research positions at the University of South Florida, where she served as a data analyst for educational programs and specialized in advanced statistical modelling.

Dr. Enriquez-Savery's distinguished leadership philosophy centers on evidence-based decision making, comprehensive stakeholder engagement, and continuous institutional improvement. Her strategic approach has successfully transformed the University of Belize into a forward-thinking institution exceptionally prepared for successful accreditation and enhanced educational excellence.

Renowned as a champion of academic integrity and institutional effectiveness, Dr. Enriquez-Savery continues to foster an enriched educational environment that serves both students and faculty while strategically advancing the university's mission of providing quality higher education that meets national and regional needs. Her comprehensive leadership has positioned the University of Belize for sustained success in the evolving landscape of higher education.



Session Info

K5

KEYNOTE SESSION

October 14th, 23:00 - 00:00 UTC

Rooted and Resilient

USE SCHED TO ACCESS THE ZOOM ROOM

Working in data science and analytics can sometimes feel like planting seeds in soil you cannot fully control. Priorities change, reports break unexpectedly, and the people responsible for making sense of it all can easily feel overwhelmed. This talk is about how we can build analytics work and teams that stay strong and flexible, no matter what shifts around us. I will share lessons and stories from my own experience as an analytics engineer, mentor, and leader. We will explore how to create solid foundations so your work remains reliable even when deadlines are tight. I will talk about simple habits to catch mistakes early before they become big problems. More importantly, I will discuss how to build a team culture where everyone feels safe to speak up, ask questions, and learn together. Whether you are supporting a teammate or influencing decision-makers, I will share how to lead with clarity and care during uncertain times. By the end of the talk, I hope everyone, whether just starting their career or already leading a team, walks away with fresh ideas for building an analytics practice and a team that can grow and succeed through any challenge

ORGANIZER: Cynthia Bland, Caucus for Women in Statistics and Data Science

CHAIR: Cynthia Bland, Caucus for Women in Statistics and Data Science



Speaker Bios

K5



Dr. Ché Smith

Senior Analytics Engineer, Netflix

Ché Smith is a data enthusiast whose career spans multiple industries, from public health to tech, all connected by a passion for uncovering insights amid uncertainty. She is currently a Senior Analytics Engineer at Netflix, where she helps shape content strategy by refining how performance metrics are understood and applied.

Ché holds a BS in Mathematics from Spelman College and a PhD in Biostatistics from the University of North Carolina at Chapel Hill. She began her career working with local and state health departments to analyze and communicate public health data in innovative ways. She later served as a mathematical statistician at the U.S. Food and Drug Administration, reviewing oncology and hematology drug applications. Before joining Netflix, she taught data science and statistics at Davidson College and directed a consulting, analytics, and tutoring lab supporting students and local entrepreneurs.

Inspired by the power of mentorship and community, Ché is passionate about helping women in data science and statistics grow their confidence, share their stories, and support each other as they navigate their careers and make an impact around the world.



Invited Sessions and Panels



Session Info

S101

OPENING SESSION

October 14th, 00:00 - 00:30 UTC

Opening

USE SCHED TO ACCESS THE ZOOM ROOM

This session will welcome attendees, discuss the history of the IDWSDS Conference, and you'll hear from the representatives from the Caucus for Women in Statistics & Data Science (CWS), American Statistical Association, and the Portuguese Statistical Society (SPE). They will share how their organizations help their members thrive in their environment. We will also share technical information about how to access key features of the Sched App and our Zoom rooms. We will conclude with a speed session, which is a new feature of IDWSDS this year. That speaker is Mariza Veruari. Her talk is titled "Thriving Against the Odds: My Journey into AI Engineering".



UMUT OZBEK
CWS President



BONNIE GHOSH-DASTIDAR
ASA President



LGIA HENRIQUES-RODRIGUES
SPE President



CYNTHIA BLAND
Program Co-Chair



JESSICA KOHLSCHMIDT
Program Co-Chair



Speaker Bios

S101

**MISS MARIZA VERUARI**

RLadies Global

I'm Mariza Veruari, a software developer, future AI engineer, and community leader with PyLadies Tirana and R-Ladies Tirana. Some people know me as "The Singing Programmer" because I combine music and coding to make learning more fun. My interest in technology began with a love for math and robotics. I was curious about how machines could be built, programmed, and given simple decision-making abilities through sensors and basic algorithms. That curiosity grew into a lasting interest in AI. I have over five years of experience in web development, and along the way I've learned how much statistics and data analysis contribute to building intelligent systems and solving practical problems. Most of what I know has been learned on my own, through persistence and curiosity. I enjoy finding solutions to challenges, mixing creative ideas with technical skills, and encouraging others to keep learning.





Thriving Against the Odds: My Journey into AI Engineering

Mariza Veruari, R-Ladies Global

For me, “thriving in your environment” means finding ways to grow, even when resources and opportunities are limited. In this session, I will share my path from web development to AI engineering, a goal I have pursued for many years. My interest in technology began with a love for math and robotics, where I was fascinated by how machines could be designed, programmed, and guided by data to perform meaningful tasks. Over time, I built skills in programming, statistics, and data analysis to move closer to working in AI. Most of this learning happened independently, using accessible resources and guidance from online communities. I will discuss how skills such as data cleaning, analysis, and interpreting statistical patterns played a central role in my transition, and how I stayed motivated by combining technical work with creative projects. This talk offers a personal perspective on making steady progress in less-than-perfect conditions and using the skills you already have to open new opportunities, wherever you are starting from.



Session Info

S102

CAREER OR PROFESSIONAL DEVELOPMENT SESSION

October 14th, 1:30 - 2:30 UTC

More Than Code: The People, Purpose, and Power of R-Ladies+

USE SCHED TO ACCESS THE ZOOM ROOM

Curious about R-Ladies+? Join us for a welcoming and inspiring panel introducing R-Ladies+, a global organisation dedicated to promoting gender diversity and inclusion among users of R, whether you're a researcher, statistician, educator, student, or professional working in any field. This session is ideal for those new to R-Ladies+ or considering getting involved, whether by attending events, joining a local chapter, or starting one in your local area. Hear from panellists across career stages and regions as they share their personal journeys with R-Ladies+: how they got involved, what the community has meant to them, and how R-Ladies+ has supported their academic and professional growth. You'll learn about the mission and structure of R-Ladies+, the kinds of local and global activities members participate in, and the many opportunities the network offers~from mentorship and leadership to skill-building and inclusive learning spaces. Whether you're just beginning your journey with R or have years of experience, this session will offer practical tips, honest reflections, and meaningful insights into how R-Ladies+ fosters a sense of belonging and empowers its members to thrive. Come discover how you can connect with this vibrant, international community, and maybe even start a new chapter of your own!

ORGANIZER: Hunna Watson, University of North Carolina at Chapel Hill

CHAIR: Hunna Watson, University of North Carolina at Chapel Hill



Speaker Bios

S102



MS. DIONNE ARGYROPOULOS

Walter and Eliza Hall Institute of Medical Research

Dionne is a molecular epidemiologist and population geneticist currently positioned at Walter and Eliza Hall Institute. Dionne's avid interest in genetics, evolution and coding led her to apply these approaches to studies of infectious diseases, such as *Plasmodium falciparum* malaria in her Masters of Science and PhD, to *P. vivax* malaria and various microbes in her current first post-doc position. The research topics of interest include drug or antimicrobial resistance, spatiotemporal trends, epidemiological statistics, statistical modelling and machine learning techniques, data wrangling and visualisation. Her work has involved international collaborations in West Africa and North America, as well as supervision and support of students, with various opportunities for knowledge exchange and capacity-building in coding and genetic epidemiology. Dionne is an active member of the R-Ladies Melbourne community, and the current President.



MS. FELICIA BONGIOVANNI

Walter and Eliza Hall Institute of Medical Research

Felicia is a PhD student at the Walter and Eliza Hall Institute in Parkville under the supervision of Professor Ivo Mueller, Professor Jodie McVernon, and Dr Eamon Conway. Her research focuses on developing diagnostic algorithms to better understand changes in the multi-antigen antibody profiles of those infected with *Plasmodium vivax* and conducting scenario-modelling to facilitate the adoption of serological testing and treatment interventions to support the elimination of malaria in the Asia-Pacific and Americas.



DR. GIORGIA MORI

Australian BioCommons

Giorgia's journey began in academia with a PhD in Molecular Microbiology (IUSS, Italy), where she characterised molecular targets of antitubercular drugs. She later focused on the human microbiome in cancer and tuberculosis, applying multi-omics approaches to understand host-microbe interactions in disease. She subsequently joined the battle against cancer at the European Institute of Oncology, where she quantitatively validated clinical genomics workflows to perform various bioinformatics tasks. Giorgia discovered her passion for education while supervising and training students in computational skills for efficient and reproducible research and became an Instructor and Trainer with The Carpentries community. Now at Australian BioCommons, Giorgia combines her technical expertise with her love for training. She lives in Sydney with her husband and two kids and volunteers in diversity initiatives to support the representation of gender minorities in the programming community.



DR. KRISTY ROBLEDO

University of Sydney

Dr Kristy Robledo is an early career academic biostatistician (within five years of PhD, relative to opportunity), with over 17 years of experience in clinical trials research. She began her career as a professional statistician and now holds an NHMRC Emerging Leadership Fellowship, and over \$43 million in competitive grant funding. Kristy provides statistical leadership across multiple clinical trial networks—including oncology, neonatology, and cardiovascular medicine—supporting the design, conduct, analysis, and reporting of high-impact trials. Her methodological research focuses on the analysis of biomarker data in clinical trials and the development of risk prediction models. A passionate advocate for reproducible research, Kristy actively promotes efficient coding practices in statistical software. She co-leads the R-Ladies+ Sydney chapter and facilitates monthly Coding Conundrums sessions at the University of Sydney.



Session Info

S103

CAREER OR PROFESSIONAL DEVELOPMENT SESSION

October 14th, 2:30 - 3:00 UTC

From the Dental Chair to the Data Cloud: Thriving in an Interdisciplinary Environment

USE SCHED TO ACCESS THE ZOOM ROOM

In the context of a professional trajectory bridging the clinical rigor of dentistry with the quantitative precision of data science, thriving transcends mere success; it represents a state of sustained professional vitality and flourishing. This state is fundamentally rooted in the ability to effectively leverage domain-specific expertise to navigate and contribute within an inherently interdisciplinary ecosystem. The session will define thriving as the process of operationalizing clinical heuristics into data-driven hypotheses. It is about a continuous, iterative cycle of learning and adaptability, whereby foundational skills in biostatistics and computational methods are integrated with nuanced insights from patient care and public health. This involves mastering the end-to-end data lifecycle, from data acquisition in Electronic Health Records (EHRs) to the deployment of predictive models and the application of Natural Language Processing (NLP) for unstructured clinical notes. The session will illustrate that thriving is the conscious act of transforming a unique background from a perceived vulnerability into a strategic asset, ultimately driving innovative, evidence-based solutions for complex oral and systemic health challenges. This is not simply about professional survival, but about leading the future of interdisciplinary collaboration.

ORGANIZER: Nisha Daniel, University of Edinburgh, Scotland, UK**CHAIR:** Nisha Daniel, University of Edinburgh, Scotland, UK

Speaker Bios

S103



DR. NISHA DANIEL

University of Edinburgh, Scotland, UK

<https://drnishadanielbds.wixsite.com/profile>

Nisha Daniel is a dentist and a health data scientist with a unique, interdisciplinary expertise, she is at the intersection of data and public health through her fully funded postgraduate studies being an Edinburgh Global Scholar'2023. Her work focuses on leveraging data science methodologies, such as machine learning and public health informatics, to address pressing oral health challenges. Nisha has applied her skills to projects ranging from community oral health programs and oral cancer screening to digital health initiatives, showcasing her commitment to using data to drive equitable and impactful health outcomes. She is a recognized student leader and a passionate advocate for interdisciplinary collaboration in healthcare. Being a dental healthpreneur, she also has valuable publications, ranging from a scoping review of oral cancer to data ethics to implementation assessment of public health policies, while pursuing research in data science in dentistry.



Session Info

S104

TECHNICAL STATISTICAL SESSION

October 14th, 3:00 - 4:00 UTC

Advancing Real-World Data Research in Asia-Pacific: Practices and Opportunities

USE SCHED TO ACCESS THE ZOOM ROOM

This panel brings together leading experts from Australia, China, Hong Kong, and Japan, from both academia and industry, to examine the dynamic landscape of real-world data (RWD) and real-world evidence (RWE) across the Asia-Pacific region. Panelists will explore the most commonly used data sources—such as claims databases, electronic medical records (EMRs), and disease registries—and present case studies showcasing successful applications of RWE in healthcare decision-making. The discussion will also highlight key challenges and identify opportunities for regional collaboration. By comparing experiences across countries, this session aims to foster cross-border dialogue and share best practices for maximizing the impact of RWD research. Attendees will gain practical insights into navigating the complexities of real-world studies and advancing the use of RWE in a rapidly evolving data environment.

ORGANIZER: Zhihong Cai, Eli Lilly Japan**CHAIR:** Swathi Pathadka, Daiichi Sankyo Co Ltd

Speaker Bios

S104

**JINNAN LI**

Eli Lilly China

Jinnan Li is a Principal Statistician at Eli Lilly China, specializing in Real-World Evidence (RWE) research, post-marketing safety studies, and medical affairs publications. She earned her MPH from University of California, Los Angeles (UCLA) and worked on a range of observational studies at UCLA and Palo Alto Medical Foundation Research Institute before joining Eli Lilly China.

**XUE LI**

University of Hong Kong

Xue Li is an Assistant Professor at the University of Hong Kong, jointly appointed by the Department of Medicine and the Department of Pharmacology and Pharmacy. Her research focuses on health technology assessment (HTA), health economics, and real-world evidence (HEOR), using decision models and electronic medical records. She specializes in healthcare data analytics, cost-effectiveness modeling, and health policy evaluation, and has led numerous HEOR projects in collaboration with academic institutions, government, industry, NGOs, and key opinion leaders.

**KATE MILLER**

Population Health Research Network (Australia)

Kate Miller is the Senior Specialist for Science and Partnerships with the Population Health Research Network (PHRN). With a background in epidemiology, infectious disease, and type 1 diabetes, she now drives strategic projects to strengthen national data linkage capacity in Australia and develops training programs to support researchers and ethics committees.

**SWATHI PATHADKA**

Daiichi Sankyo Co., Ltd. (Japan)

Swathi Pathadka is a Lead Associate at Daiichi Sankyo Co., Ltd., specializing in pharmacoepidemiology research. She earned her PharmD from Manipal University, India. Swathi has experience working with Asian healthcare databases, including Japanese claims databases (JMDC, MDV, IQVIA), EHR-based databases (JAMDAS), and Hong Kong's CDARS database.



Speaker Bios

S105

**MENGQI YIN**

Washington State University

Maggie is a fifth-year Ph.D. student in Statistics at Washington State University, with a strong dedication to both research and teaching. Her studies focus on statistical scrutiny in AI classification, particularly in managing ambiguity through neutral zones. Through her work, she strives to enhance the interpretability, robustness, and practical deployment of AI-driven decision systems, ensuring that statistical rigor remains at the core of modern machine learning applications.

**WEIWEI XIE**

Washington State University

Weiwei Xie is a doctoral candidate in the statistics program at Washington State University's Department of Mathematics and Statistics. Her research centers on the investigation on random walk null models for time series entropy analysis. Working under the esteemed guidance of Dr. Daryl DeFord, Weiwei aims to advance the understanding of randomness and complexity within time-dependent data. Prior to doctoral studies, she has earned a master's degree in Statistics from Washington State University in 2021.

**DAVID RICE**

Washington State University

David Rice is a 5th year PhD student at Washington State University studying statistical science. In his free time, he enjoys making homemade pizzas and going on runs.



**Abstract 1**

Mengqi Yin, Department of Mathematics & Statistics, Washington State University

As Artificial Intelligence (AI) becomes more ubiquitous, it makes sense to think about what exactly we mean by artificial intelligence. As Statistics is generally considered the science of data, it makes sense that we think of statistical thinking in AI. However, the conundrum is that there are not as many papers looking at statistical thinking in AI, more in the deployment of AI to data problems. A seminal paper by Yu and Kumbier (2017) introduced the P-Q-R-S framework-Population, Question, Representation, and Scrutiny-as essential components in deploying AI systems. This presentation focuses on the S component, exploring neutral zones as a method for managing ambiguity in classifications. Specifically, we discuss the challenges of constructing neutral zones for non-model-based methods, such as ANN, MLP, and CNN, where output is limited to class probabilities. Through this exploration, we aim to highlight the importance of statistical scrutiny in enhancing the reliability and interpretability of AI system.

multilayer perceptions will be discussed, and conclude with future areas of research.

Abstract 2

Weiwei Xie, Department of Mathematics & Statistics, Washington State University

Permutation entropy has emerged as a widely used statistical measure for assessing the complexity of time series, with applications spanning diverse fields such as biology, economics, and physics. Recent work has applied divergence measures, including Kullback-Leibler divergence and Jensen-Shannon divergence, to quantify the similarity between time series based on ordinal patterns, and to measure the deviation from generative models. This approach also allows for creating meaningful embeddings of the data. While early approaches assumed a uniform null model for the underlying data generation process, many real-world applications involve more complex distributions and random walk models. In this study, we present an analysis of several null models for generating time series data, from both theoretical and empirical perspectives. We successfully derive theorems describing the behavior of random walks with uniformly and normally distributed steps respectively, as well as introducing a novel random walk null model based on transition matrices inferred from real-world data using Markov chains. This data-driven approach allows for a more accurate representation of empirical distributions. We demonstrate the practical utility of these methods by applying them to real world datasets from economics and other domains, highlighting the effectiveness of divergence measures in capturing the complexity of empirical time series data.

Abstract 3

David Rice, Department of Mathematics & Statistics, Washington State University

We will briefly introduce Kolmogorov-Arnold Networks, a relatively new class of neural networks which uses basis functions as activation for each neuron. The optimization on basis coefficients allow activations to change during training, and can offer improved interpretability. We outline differences from



Session Info

S106

TECHNICAL STATISTICAL SESSION

October 14th, 5:00 - 6:00 UTC

Monash EBS PhD Contest

USE SCHED TO ACCESS THE ZOOM ROOM

This session features presentations from three PhD Students in Statistics and Data Science.

The speakers will be:

Jayani Lakshika, Monash University

Title: Visualise your fitted non-linear dimension reduction model in the high-dimensional data space

Meg Tully, The University of Melbourne

Title: Evaluation of a pharmacokinetic-pharmacodynamic model for predicting parasitological outcomes in Phase 2 studies of new antimalarial drugs

Sylviane Miharisoa

Title: Spatiotemporal mapping of multi-locus resistance in Anopheles mosquitoes

The best presenter in this session will be awarded a \$AU500 cash prize, as judged by the expert panel consisting of:

A/Prof Freya Shearer, The University of Melbourne

Dr Joanna Batstone

Tyler Reysenbach, Australian Department of Foreign Affairs and Trade

ORGANIZER: Michael Lydeamore, Monash Business School

CHAIR: Michael Lydeamore, Monash Business School



Speaker Bios

S106



JAYANI LAKSHIKA

Monash University

Jayani Lakshika is a final-year PhD student in Statistics at Monash University. Her research focuses on developing interactive visual tools and statistical methodologies for selecting and evaluating non-linear dimension reduction layouts of high-dimensional data. She is the author of several R packages. Jayani is a member of R-Ladies Melbourne and is passionate about teaching, with extensive experience tutoring statistics and data science courses, and mentoring students in both Australia and Sri Lanka. In addition to her academic work, she brings industry experience as a data analyst.



MEG TULLY

The University of Melbourne

Meg Tully has a Masters of Biostatistics from the University of Melbourne and is currently undertaking a PhD specialising in novel antimalarial combination therapies to combat the ongoing threat of drug resistance. She is based in the Centre for Epidemiology and Biostatistics within the Melbourne School of Population and Global Health and supervised by Prof Julie Simpson, Dr David Price and Dr Rob Commons. Her research is currently focused on pharmacokinetic-pharmacodynamic modelling of the within-host dynamics of malaria infection and treatment.



SYLVIANE MIHARISOA

Nelson Mandela African Institution of Science and Technology

Sylviane Miharisoa is a PhD candidate from Madagascar based at the Nelson Mandela African Institution of Science and Technology. With a Master's in entomology, she has experience in field surveillance, molecular screening and bioecology of vector populations. Her doctoral research develops spatiotemporal models of genetic resistance in Anopheles mosquitoes to predict resistance spread and itself and vector control on the abundance of Anopheles. Her goal is to translate these model outputs into practical recommendations for national malaria control programmes.



Session Info

S107

CAREER OR PROFESSIONAL DEVELOPMENT SESSION

October 14th, 6:00 - 7:00 UTC

Thriving in Statistics in New Zealand

USE SCHED TO ACCESS THE ZOOM ROOM

This session will be presented by four statisticians working in New Zealand who are passionate about thriving in our environment. Three of the presenters are academic biostatisticians working in the University sector and one is a consultant statistician working in the Corporate Sector. We are all at different stages of our careers, ranging from early/mid career to senior management. We have had experience in networking, contributing to our work communities and considering what makes our working lives enjoyable. This includes developing academic networks, support networks for mentoring, and managing statisticians with a focus on thriving. Our session will comprise four presentations with room for discussion at the end. These will cover (1) the development and maintenance of a biostatistical network at the University of Otago, (2) the establishment of a mentoring programme through the New Zealand Statistical Association (that has been running for several years), (3) how these activities support early career statisticians and (4) how we can be valued within a large organisation through clear policy/marketing/strategy supporting a culture where biostatisticians succeed.

ORGANIZER: Claire Cameron, Biostatistics Centre, University of Otago

CHAIR: Claire Cameron, Biostatistics Centre, University of Otago



Speaker Bios

S107



DR. CLAIRE CAMERON

Biostatistics Centre, University of Otago

<https://www.otago.ac.nz/dsm-psm/research/biostatistics/biostatistics-people/profile?id=1088>

Claire Cameron is the Director of the Biostatistics Centre and has been a biostatistician since 2011. The Centre improves the quality of health related research through biostatistical collaboration and advice across the Division of Health Sciences (primarily) at the University. Her involvement in research is varied but, usually, she contributes methodological leadership and expertise to projects in a wide range of application areas. Her position is Research Associate Professor. She has been a member of the New Zealand Statistical Association since 1990 and a member of the Caucus of Women in Statistics and Data Science since 2020. Over the years she has contributed to several networks for different academic groups including biostatisticians. She has also been involved in mentoring (formally and informally) for many years. She sees supporting colleagues at all levels and being part of these networks as an essential part of having a healthy working life.



MRS. LISA THOMSEN

Fonterra

<https://www.linkedin.com/in/lisa-thomsen-984990b1/>

Lisa Thomsen has been working as a consultant industry statistician at the Fonterra Research & Development Centre in Palmerston North, New Zealand, for over 9 years. In this role she has supported teams with their stats and data needs for projects to launch new dairy products and tweak existing ones. This has included modelling the performance of milk powders, analysing cheesecake performance, and investigating consumer preferences of yoghurts. Lisa has a particular interest in encouraging robust data management and experimental design. Lisa is an active member of the New Zealand Statistical Association, having been the Student & Early Career rep for 5 years before becoming the Mentoring Program Co-ordinator. The Mentoring Program aims to increase connection and networking opportunities amongst NZSA members.



DR. ALICE HYUN MIN KIM

University of Otago Wellington

<https://www.otago.ac.nz/wellington/departments/central-services/biostatisticservices/staff/alice-kim>

Dr. Alice Hyun Min Kim is a Biostatistician/Senior Research Fellow based in Wellington, New Zealand. Her research interests focus on the applications of Epidemiology and Statistics in Medicine, Psychology and Public Health. She studied Economics at the University of Auckland and Harvard University, and has an MSc in Statistics from the University of Auckland and a PhD in Health Sciences from the University of Canterbury. Her teaching experiences include Epidemiology (first year) and Health Issues in the Community (postgrad) courses. In her current role at the University of Otago Wellington, she provides biostatistical and methodological input to various health research projects working with a diverse group of researchers across disciplines and institutions. She is a member of the WHO Thematic Platform for Health



PROF. ROBIN TURNER

Biostatistics Centre, University of Otago

<https://www.otago.ac.nz/dsm-psm/people/profile?id=2842>

Robin is a Professor in the Biostatistics Centre and Head of Department of Preventive and Social Medicine. She has worked at the University of Otago since 2017, where she was the inaugural Director of the Biostatistics Centre. In this role she focused on supporting and developing the Biostatisticians in the Centre to thrive in a complex political environment. This included development of a marketing strategy to improve how people engage with the Centre, clarifying policies to protect and empower biostatisticians and supporting biostatisticians to get promoted in recognition of their experience. She also focused on upskilling academics and students in biostatistics which has led to easier communication in research settings. She works across a variety of health-related areas and is passionate about supporting biostatisticians and creating work cultures that support biostatisticians to work to their best.



Session Info

S108

TECHNICAL STATISTICAL SESSION

October 14th, 9:00 - 10:00 UTC

Challenges in Modern Statistical Inference

USE SCHED TO ACCESS THE ZOOM ROOM

The WISK (Women in Statistics in Korea) session focuses on the pressing challenges in modern statistical inference.

Contemporary statistical inference must address issues arising from massive and complex data, computational constraints, limited interpretability, and evolving understanding of core statistical concepts. There is a growing demand for statistical methods capable of handling high-dimensional and complex data structures while still delivering reliable, interpretable, and actionable results.

ORGANIZER: Jaehee Kim, Duksung Women's University

CHAIR: Yang-Jin Kim, Sookmyung Women's University



Speaker Bios

S108

**PROF. BYOL KIM**

Sookmyung Women's University

She got Master's and Ph.D in statistics from Chicago Univeristy. Her research interest is high-dimensional data analysis and distribution-free statistical analysis

**PROF. YEI EUN SHIN**

Seoul National University

She got Master of Statistics from Seoul National University, and got PhD in Statistics, Texas A&M University. Her research interest is advanced survival analysis.

**PROF. JOOYOUNG LEE**

Chung-Ang University

She got Master of Mathematics in Actuarial Science, University of Waterloo, and Doctor of Philosophy of Mathematics in Biostatistics, University of Waterloo, ON, Canada.



Black-Box Tests for Algorithmic Stability*Byol Kim, Sookmyung Women's University*

Algorithmic stability is a concept from learning theory that expresses the degree to which changes to the input data (e.g. removal of a single data point) may affect the outputs of a regression algorithm. Knowing an algorithm's stability properties is often useful for many downstream applications—for example, stability is known to lead to desirable generalization properties and predictive inference guarantees. However, many modern algorithms currently used in practice are too complex for a theoretical analysis of their stability properties, and thus we can only attempt to establish these properties through an empirical exploration of the algorithm's behavior on various datasets. In this work, we lay out a formal statistical framework for this kind of black-box testing without any assumptions on the algorithm or the data distribution, and establish fundamental bounds on the ability of any black-box test to identify algorithmic stability. This is a joint work with Rina Foygel Barber.

Survival Analysis in Subsampled Cohorts: Making the Most of Limited Data*Yei Eun Shin, Seoul National University*

In large cohort studies, subsampling designs such as case-cohort and nested case-control sampling are often used to reduce cost and data collection burden. While practical, these designs reduce statistical efficiency and raise methodological challenges for survival analysis, particularly when estimating absolute risks, accounting for competing events, or incorporating time-varying covariates. This talk introduces recent methods that aim to improve the performance of survival analysis under such designs. The first part presents approaches for improving the estimation and validation of risk prediction models including absolute risks, proportional and additive hazards models when risk factors are not fully available in a cohort. For competing risk analysis with event-specific subsamples, a proportional risk model provides a simple and statistically efficient way in a joint framework. Additional topics include estimating transition probabilities in multi-state models and applying landmarking methods when only limited subsampled data are available. The talk also introduces ongoing work on multiple imputation techniques for case-cohort and nested case-control designs, and on counter-matching strategies to improve efficiency in estimation. Together, these methods illustrate how thoughtful use of subsampled data can support efficient and flexible survival analysis, even when full cohort data are not available. The talk aims to provide practical strategies and methodological insights.

Immortal Time Bias in Failure Time Analysis*Jooyoung Lee, Chung-Ang University*

In survival analyses of observational studies evaluating drug effectiveness, immortal time bias may arise when eligibility criteria and treatment assignment are misaligned. Specifically, immortal time arises when post-eligibility information is used to assign treatment, or when post-assignment information is used to redefine eligibility. In real-world settings, treatment initiation may depend on biomarkers, which in turn determine the length of the immortal time period. Several methods have been proposed to address immortal time bias, including the time-dependent Cox

model, prescription time-distribution matching, landmark analysis, the sequential trial approach, and the cloning approach. In this article, we derive the limiting values of estimators, such as hazard ratios and survival probabilities, when immortal time and biomarker values are not properly addressed under various settings. We investigate the impact of survival time distribution, biomarker effects, and treatment effects on the asymptotic and finite-sample properties of estimators derived from each method. We also evaluate the impact of intermittently measured biomarkers on estimators of interest. An application to a study of the risk of psoriatic arthritis in patients with psoriasis is given for illustration.



Session Info

S109

TECHNICAL STATISTICAL SESSION

October 14th, 10:00 - 11:00 UTC

Methodological Innovation and Conceptual Advancement in Statistical Design

USE SCHED TO ACCESS THE ZOOM ROOM

This session spans methodological innovation and conceptual advancement in statistical design. One study develops explicit formulas for optimal degradation test plans under heterogeneous Wiener process models, improving both computational efficiency and interpretability compared to traditional numerical methods. Complementing this, a two-part presentation underscores the critical role of experimental design in shaping statistical inference, even prior to data collection, and introduces the emerging field of order-of-addition experiments. This novel design framework challenges conventional modeling assumptions and offers fresh perspectives on predictor encoding, with the aim of broadening engagement and fostering deeper reflection within the statistical community.

ORGANIZER: Yi-Ting Hwang, National Taipei University**CHAIR:** Yi-Ting Hwang, National Taipei University

Speaker Bios

S109



PROF. I-CHEN LEE

National Cheng Kung University

<https://researchoutput.ncku.edu.tw/zh/persons/i-chen-lee>

I-Chen Lee received her Ph.D. in Statistics from National Tsing Hua University in 2016. She is currently an Associate Professor in the Department of Statistics at National Cheng Kung University. Her research interests focus on reliability and engineering statistics, particularly on the application of experimental design in the field of reliability.



MISS JING-WEN HUANG

Academia Sinica

<https://sites.google.com/view/jingwenhuang>

Jing-Wen Huang completed her PhD at the Institute of Statistics and Data Science, National Tsing Hua University, Taiwan. Her research focuses on mathematical structures underlying modern experimental designs, such as network or order-of-addition experiments. She has developed solid background knowledge of Mathematics, Medical Sciences from undergraduate studies and Statistics from graduate studies. She is now working as a fresh postdoctoral researcher at the Institute of Statistical Science, Academia Sinica, Taiwan. Her current career goal is to obtain a stable academic position in Applied Mathematics or Statistics in or near Taiwan. Although she is introverted in daily life, she can get excited easily when engaging in academic discussions. Please kindly let her know if you find she is potentially suitable for some opening positions.



MISS YA-SHAN CHENG

National Central University

Ya-Shan Cheng earned her B.A. in Business Administration and M.S. in Statistics from National Central University. She gained nearly five years of industry experience at Powerchip Technology and ASUS, followed by research work at Academia Sinica's Institute of Statistical Science. In 2021, she began her Ph.D. at National Tsing Hua University, focusing on reliability analysis and optimal degradation testing. From June 2024 to May 2025, she visited the University of Maryland under the Thousand Talents Program, collaborating with Professor Mei-Ling Ting Lee on longitudinal threshold regression and expanding into public health research. After completing her doctoral exam in May 2025, she briefly served as a postdoctoral researcher at Academia Sinica. In August, she joined National Central University's Institute of Statistics as an assistant professor.



Optimal Robust Strategies for Fatigue Testing of Polymer Composite Materials

I-chen Lee, National Cheng Kung University

Polymer composite materials are widely used in transportation and renewable energy for their lightweight, high strength, and durability. Ensuring their long-term reliability under fatigue is essential, requiring efficient accelerated life testing (ALT). ALT seeks accurate lifetime predictions at minimal cost, but conventional methods often depend on locally optimal designs based on precise parameter estimates—an unrealistic assumption due to inherent uncertainty. This study proposes a standardized minimax optimal design for fatigue testing of polymer composites, accounting for parameter uncertainty by considering a range of plausible values and safeguarding against worst-case scenarios. The design employs a hybrid optimization strategy, integrating particle swarm optimization (PSO) with supplementary techniques to tackle a non-differentiable criterion and nested optimization challenges. Numerical results show that the proposed minimax designs outperform traditional locally optimal and Bayesian designs in both efficiency and reliability. This framework offers a robust, practical approach for evaluating the long-term performance of polymer composites in real-world conditions.

affected by model parameters and experimental costs. Some numerical examples are presented to support the efficiency and applicability of the optimal test plans. This work is a collaboration with Dr. Chien-Yu Peng from Academia Sinica.

Experimental Design Introduction and Related Challenges in Order-of-addition Experiments

Jing-Wen, Academia Sinica

This talk will be divided into two main parts. The first part aims to introduce the importance and foundation of Experimental Designs. Sometimes, people think of statistician only when they have collected their data. However, decisions made before data collection can greatly influence the theoretical quality of estimators, models and inferences. I hope this talk will help the audiences have impression that Statistics also plays a crucial role before any data is available, even in the simple linear regression case. In the second part, I will briefly introduce the development and challenges of a growing field: the design and analysis for order-of-addition experiments. This kind of experiments encourages deeper reflection on what statistical modeling entails and offers new insights into what should we code for the predictors in linear models. I have been working on this topic recently and would like to share this interesting area to broader audiences.

Optimal Test Planning for Heterogeneous Wiener Processes

Ya-Shan, National Central University

Degradation models based on heterogeneous Wiener processes are commonly used to assess information on the lifetime of highly reliable products. An optimal test plan given limited resources is generally obtained using numerical methods for heterogeneous Wiener processes. However, numerical searches for optimal test plans have the disadvantage of being time-consuming and may provide unclear explanations for the findings. To overcome these difficulties, we derive an explicit expression for decision variables (such as the termination time, number of measurements, and sample size) of D- and V-optimal test plans with cost constraints. The theoretical results not only ensure that the optimal test plan is found, but also provide clear insights into the decision variables



Session Info

S110

TECHNICAL STATISTICAL SESSION

October 14th, 11:00 - 11:30 UTC

Embracing Complexity: Advanced Statistical and Computational Methods in Modern Data Analysis

USE SCHED TO ACCESS THE ZOOM ROOM

This session showcases cutting-edge statistical and machine learning research that tackles complex data challenges using sophisticated computational and analysis techniques. While the first three presentations span innovative Bayesian methodologies, including nonparametric hypothesis testing for multiple comparisons, adaptive LASSO quantile regression for handling missing data, and dynamic spatiotemporal modeling for real-time forecasting, the last one introduces a state-of-the-art application of deep learning and natural language processing in e-commerce fake review detection. Together, these talks highlight the ongoing evolution of modern statistics and data analysis practices, emphasizing the development and utilization of these advanced methods to extract meaningful insights from increasingly intricate datasets across diverse domains.

ORGANIZER: Mai Dao, Wichita State University**CHAIR:** Mai Dao, Wichita State University

Speaker Bios

S110



DR. ZHUANZHUAN MA

The University of Texas Rio Grande Valley

<https://www.linkedin.com/in/zhuanzhuan/>

Dr. Zhuanzhuan Ma is an assistant professor of statistics in the School of Mathematical and Statistical Sciences at the University of Texas Rio Grande Valley. She obtained her Ph.D. in Statistics from Texas Tech University in August 2022 under the supervision of Drs. Min Wang (Univ. of Texas at San Antonio) and Dimitri Volchenkov (Texas Tech Univ.). Her primary research interests include Bayesian statistics, statistical machine learning, and high-dimensional data analysis.



DR. MAI DAO

Wichita State University

https://www.wichita.edu/profiles/academics/fairmount_college_of_liberal_arts_and_sciences/Math/Faculty/Dao-Mai.php

Dr. Mai Dao is an Assistant Professor of Statistics in the Department of Mathematics, Statistics, and Physics. She graduated with her Ph.D. from Texas Tech University under the supervision of Professors Min Wang (The University of Texas at San Antonio) and Souparno Ghosh (University of Nebraska-Lincoln). Her primary research interests focus on Bayesian statistics, high-dimensional inference, and statistical machine learning.



DR. NADEESHA JAYAWEERA

The University of Akron

<https://www.linkedin.com/in/nadeesha-jayaweera>

Dr. Nadeesha Jayaweera is an Assistant Professor of Statistics at the University of Akron. She earned her Ph.D. in Mathematics (Statistics concentration) from Texas Tech University, focusing on model selection, and completed a postdoc at Worcester Polytechnic Institute specializing in Bayesian statistics, time series, and applied modeling. Her research centers on time series analysis, discrete choice modeling, hierarchical Bayesian methods, and applications in health, marketing, and social sciences. She is passionate about data-driven decision-making and developing accessible statistical tools for interdisciplinary work. Dr. Jayaweera mentors undergraduate research, promotes inclusive teaching, and presents her work at national and international conferences.



DR. THILINI JAYASINGHE

University of Dayton

<https://udayton.edu/directory/artsscience/mathematics/jayasinghe-thilini.php>

Dr. Jayasinghe completed her Ph.D. in Mathematics, concentration in Statistics, at Texas Tech University in the summer of 2022, and M.S. in Statistics from the same institution in 2019. Previously, she received a B.S. in Statistics from the University of Kelaniya, Sri Lanka, in 2015. After she completed her Ph.D., she joined Wittenberg University as an adjunct instructor. Dr. Jayasinghe joined the faculty of the Department of Mathematics at the University of Dayton in Fall 2023. She likes playing badminton, traveling, and watching movies outside of teaching, research, and advising students.



Bayesian Nonparametric Hypothesis Testing Methods on Multiple Comparisons*Zhuanzhuan Ma, The University of Texas Rio Grande Valley*

In this talk, we introduce Bayesian testing procedures based on the Bayes factor to compare the means across multiple populations in classical nonparametric contexts. The proposed Bayesian methods are designed to maximize the probability of rejecting the null hypothesis when the Bayes factor exceeds a specified evidence threshold. It is shown that these procedures have straightforward closed-form expressions based on classical nonparametric test statistics and their corresponding critical values, allowing for easy computation. We also demonstrate that they effectively control Type I error and enable researchers to make consistent decisions aligned with both frequentist and Bayesian approaches, provided that the evidence threshold for the Bayesian methods is set according to the significance level of the frequentist tests. Importantly, the proposed approaches allow for the quantification of evidence from empirical data in favor of the null hypothesis, an advantage that frequentist methods lack, as they cannot quantify support for the null when the null hypothesis is not rejected. We also present simulation studies and real-world applications to illustrate the performance of the proposed testing procedures.

Bayesian Adaptive LASSO Quantile Regression with Non-Ignorable Missing Responses*Mai Dao, Wichita State University*

In this paper, we develop a fully Bayesian adaptive LASSO quantile regression model to analyze data with non-ignorable missing responses, which frequently occur in various fields of study. Specifically, we employ a logistic regression model to deal with missing data of non-ignorable mechanism. By using the asymmetric Laplace working likelihood for the data and specifying Laplace priors for the regression coefficients, our proposed method extends the Bayesian LASSO framework by imposing specific penalization parameters on each regression coefficient, enhancing our estimation and variable selection capability. Furthermore, we embrace the normal-exponential mixture representation of the asymmetric Laplace distribution and the Student-t approximation of the logistic regression model to develop a simple and efficient Gibbs sampling algorithm for generating posterior samples and making statistical inferences. The finite-sample performance of the proposed algorithm is investigated through various simulation studies and a real-data example.

Dynamic Bayesian Spatiotemporal Modeling for Real-Time Disease Forecasting with Likelihood-Based Weighting*Nadeesha Jayaweera, The University of Akron*

Forecasting infectious disease spread remains a complex task due to the evolving nature of outbreaks and the influence of contextual factors such as population density. Here, we present a new Bayesian spatiotemporal modeling strategy that incorporates likelihood-based weighting within the Integrated Nested Laplace Approximation (INLA) framework. Designed to adapt to shifting transmission patterns, the method applies a dynamic discounting mechanism to historical observations, assigning higher relevance to recent data in real-time prediction. By explicitly incorporating

population density and temporal variation, the model enhances responsiveness to non-stationary trends in disease progression. We validate the approach using daily COVID-19 case data from Massachusetts counties, demonstrating improved forecasting accuracy over conventional methods. This likelihood-weighted INLA approach provides a flexible and computationally efficient tool for real-time disease monitoring, with important implications for public health surveillance and resource planning.

Detecting Fake Reviews on Amazon Using Deep Learning and NLP Techniques*Thilini Jayasinghe, University of Dayton*

In the age of e-commerce, the surge in fake reviews poses a serious threat to the credibility of online consumer feedback systems. This study addresses the issue of Amazon, a leading e-commerce platform in the United States, where fraudulent reviews have become increasingly prevalent. Due to the scarcity of authentic, labeled datasets, we introduce a novel methodology for distinguishing between genuine and deceptive reviews across both verified and non-verified purchases. Our approach leverages the bootstrap distribution of cosine similarity scores to establish a statistically grounded framework for review classification. We integrate Convolutional Neural Networks (CNNs) with word embedding techniques and emotion analysis through Natural Language Processing (NLP). The proposed model achieves an impressive accuracy of over 96% in detecting fake reviews. By enhancing the reliability of online feedback, this research contributes to fostering consumer trust and safeguarding the integrity of e-commerce platforms.



Session Info

S111

TECHNICAL STATISTICAL SESSION

October 14th, 11:30 - 12:00 UTC

Gender Inequality in Brazilian Civil Engineering Graduate Programs: An Analysis of Productivity Grant Allocation

USE SCHED TO ACCESS THE ZOOM ROOM

Gender inequality in academia, particularly in STEM fields, remains a persistent issue marked by structural barriers and gender bias that limit women's advancement. This study focuses on Brazilian graduate programs in Civil Engineering and investigates disparities in the distribution of productivity grants awarded by CNPq (National Council for Scientific and Technological Development). A case analysis was conducted on the scientific output of 307 grant recipients (PQs) across categories 1A, 1B, 1C, 1D, and 2. Gender was identified using GenderAPI and ChatGPT, given the absence of such data in official records, with a 95% agreement between tools. Descriptive statistics, hypothesis testing (t-test and non-parametric median test), and multinomial logistic regression were used to analyze differences in academic production and identify possible gender bias in grant classification. Results indicate that men and women with similar performance indicators (publications, supervision, years since PhD) are evaluated differently: while each undergraduate supervision increases men's odds of attaining higher PQ levels, it decreases women's odds. Moreover, A1 journal publications raise women's odds of reaching top categories more than they do for men, suggesting women must meet higher standards to achieve similar recognition. In light of this scenario, the present study aims to contribute to more equitable and inclusive environments in Brazilian science.

ORGANIZER: Gabriela Paschoal , Federal University of Juiz de Fora

CHAIR: Gabriela Paschoal , Federal University of Juiz de Fora



Speaker Bios

S111



MS. GABRIELA PASCHOAL

Federal University of Juiz de Fora

<https://www.linkedin.com/in/gabriela-paschoal-da-silva-503b19189/>

I hold a Bachelor's degree in Mathematical Sciences from UFJF (Federal University of Juiz de Fora) and I am currently in the final semester of my undergraduate studies in Statistics, also at UFJF. Additionally, I am an intern at the Center for Public Policies and Education Evaluation (CAEd/UFJF), working with the Contextual Instruments and Indicators team.



Session Info

S112

TECHNICAL STATISTICAL SESSION

October 14th, 12:00 - 12:30 UTC

Advanced Methods in Health Research by the PostDocs at MRC Biostatistics Unit, University of Cambridge

USE SCHED TO ACCESS THE ZOOM ROOM

This session highlights the breadth of biostatistics through two distinct but complementary talks by the postdocs on their ongoing work at the MRC Biostatistics Unit. One focuses on enhancing statistical tests in clinical trials while the other applies advanced methods in genetic epidemiology. The first presentation introduces alternative formulations of the Allocation Probability test statistic to enhance statistical power in response-adaptive clinical trials. The second presentation talks about sex and menopause specific variability in total cholesterol and its potential link to coronary artery disease. Both of these talks show the diversity of applications within biostatistics.

ORGANIZER: Rajenki Das, MRC Biostatistics Unit, University of Cambridge

CHAIR: Rajenki Das, MRC Biostatistics Unit, University of Cambridge



Speaker Bios

S112

**DR. STINA ZETTERSTROM**

MRC Biostatistics Unit, University of Cambridge

Stina is currently doing her postdoc at MRC Biostatistics Unit, University of Cambridge. She is part of the Efficient Study Design group and is working with Dr Sofia Villar and Dr David Robertson on the development of statistical inferential methods to use in response-adaptive clinical trials and software for implementing these methods in practice. She did her PhD at the Department of Statistics, Uppsala University, Sweden, under the supervision of Ingeborg Waernbaum where she focused on selection bias in observational studies.

**DR. JANNE POTT**

MRC Biostatistics Unit, University of Cambridge

Janne is a Research Associate working under the supervision of Dr Stephen Burgess at the MRC Biostatistics Unit. The goal of her research is to contribute to precision medicine via genetic epidemiology, and improve human health and well-being, especially regarding women's health. This includes analysing the molecular differences in multiple OMICS levels on disease risk. The differences are either between the sexes or between age groups, e.g. before and after menopause. She uses genetic summary statistics from large-scale studies, genetic consortia and electronic health data in causal models of Mendelian Randomization. Prior to this, her PhD was in Genetical Statistics at Leipzig University, Germany, where she worked on genetics of steroid hormones to identify sex-dependent regulatory mechanisms.





Alternative Formulations of the Allocation Probability Test Statistic to Enhance Power in Response-Adaptive Clinical Trials

Stina Zetterstrom, MRC Biostatistics Unit, University of Cambridge

In response-adaptive clinical trials, treatment allocation probabilities are updated throughout the study based on the outcome of previous participants in the study. One reason for this strategy is to increase the probability that participants receive the most effective treatment compared to equal randomisation. However, imbalances in allocation can reduce statistical power when testing for treatment differences. Recent research has introduced a testing approach based on allocation probabilities (AP) rather than direct outcomes, which can improve power. We present alternative AP test formulations by modifying the functional form of the test statistic. Using simulation studies with the Bayesian Response Adaptive Randomization (BRAR) algorithm for binary outcomes in a two-arm trial, we assess the performance of these variants. The results indicate that changing the functional form of the AP test statistic can substantially increase the power compared to the original formulation. Furthermore, the AP test can offer greater power than traditional methods while maintaining type I error control. Although our evaluation focuses on BRAR, the AP framework is applicable to any response-adaptive randomisation, and its performance can be investigated in those contexts.

Longitudinal Genome Analysis of Total Cholesterol Stratified by Menopausal Status in Women Using Multivariable Mendelian Randomization

Janne Pott, MRC Biostatistics Unit, University of Cambridge

Total cholesterol (TC) is an easy accessible lipid biomarker that displays differences between the sexes regarding both its mean levels and its variability over life. TC is known for its role in the development of coronary artery disease (CAD), and CAD risk is also sex-differential, with a significant increase for women after menopause. We aimed at exploring the variability of TC levels in the longitudinal EHR data of the UKB and in the INTERVAL study in a stratified manner to detect differences in women due to menopausal status, with young and old men as control groups. We performed longitudinal genome-wide association studies (GWAS) in the UKB per sex and age group (pre- and post-menopausal, young and old men) to select SNPs as instrumental variables, and validated their association with TC in INTERVAL. Using an multivariable Mendelian Randomization (MVMR) approach, we test for independent effects of TC variability and between age groups. CAD outcome data was obtained from publicly available GWAS summary statistics. Although there were signals specific for male and female TC variability, we did not observe a significant effect of the variability on CAD. A possible reason for the lack of effect is the cross-sectional CAD data, which cannot determine whether events occurred before or after menopause.



Session Info

S113

TECHNICAL STATISTICAL SESSION

October 14th, 12:30 - 13:00 UTC

Innovations and Trends in Oncology Clinical Trials

USE SCHED TO ACCESS THE ZOOM ROOM

Oncology clinical trials are rapidly evolving through innovations such as precision medicine with biomarker-driven patient selection, adaptive trial designs, and the growing focus on immunotherapy and combination treatments. The integration of real-world data, digital technologies, and decentralized trial models is enhancing patient accessibility and data collection. Additionally, artificial intelligence is improving patient selection and data analysis, while regulatory agencies are adopting more flexible frameworks to expedite drug development. Together, these trends are making oncology trials more personalized, efficient, and patient-centric, ultimately accelerating the development of effective cancer therapies. This session will give an overview of early, late and healthy participant clinical trial designs and analysis.

ORGANIZER: Arinjita Bhattacharyya, Merck**CHAIR:** Arinjita Bhattacharyya, Merck

Speaker Bios

S113

**DR. ANANYA ROY**

Merck

Dr. Ananya Roy is a Senior Director of Biostatistics at Merck, where they lead statistical strategies to support late-stage clinical development and regulatory submissions. Dr. Roy has over 10 years of experience in designing and analyzing complex clinical trials across multiple indications in Oncology. Their expertise bridges advanced statistical methodologies and practical applications to drive data-driven decision-making in pharmaceutical research. Dr. Roy holds a PhD in Statistics from University of Florida and has previously worked as a faculty member at University of Nebraska, Lincoln.

**DR. TIAN HE**

Merck

Tian He is a senior scientist, biostatistics from Biostatistics and Research Decision Sciences, Early Oncology, at Merck. She obtained a PhD in biostatistics from Indiana University Indianapolis in 2022. Her research interest includes master protocol, dose finding and Bayesian analysis.

**DR. ARINJITA BHATTACHARYYA**

Merck

Dr. Arinjita Bhattacharyya is an Associate Principal Scientist at the Biostatistics and Research Decision Sciences department at Merck. She has been with Merck for the past 4 years. She graduated with a PhD in Biostatistics from the University of Louisville, Kentucky, in 2020. Dr. Bhattacharyya has multiple first authored publications with several awards. She has served as a journal editor and reviewer for numerous noteworthy journals such as Contemporary Clinical Trials, Statistics in Medicine and have served in several committees such as the American Statistical Association, ENAR. She has gained hands on experience on drug development process and is involved with several research projects. She has delivered several invited talks and organized sessions at several conferences. Her research interests lie in Bayesian shrinkage priors, clinical trials, and real-world data among others.



Design Considerations in Late Stage Oncology Trials*Ananya Roy, Merck*

Late stage trials, specifically randomized Phase 3 trials, aim to confirm the efficacy and safety of a drug in a larger population and form the basis for regulatory approvals. In this presentation, we will discuss the key statistical aspects of the design of Phase 3 Oncology trials. We will present examples of Phase 3 trials and discuss the considerations in choice of endpoints, multiplicity control methodology, and planned analysis methods which are used to analyze and interpret the trial data. We will also discuss assumptions, e.g., proportional hazard assumptions for survival analysis, which guide the choice of the analysis methods and sensitivity analyses which are often proposed in case of violation of the assumptions.

Backfilling Strategies for BOIN Design in Early Oncology Trials with Late-Onset Outcomes and Dose Optimization*Tian He, Merck*

There is growing interest in backfilling patients during early phase oncology trials, where patients are concurrently assigned to safe doses within a dose-finding study. Backfilling streamlines drug development by collecting additional safety and activity data. Meanwhile, the US Food and Drug Administration (FDA) launched Project Optimus, which emphasizes the need for dose optimization in early phase trials. In response, we extend the widely used Bayesian optimal interval (BOIN) design to incorporate practical backfilling strategies. Our proposed design consists of two seamless parts: a backfilling dose-finding trial, followed by an extended backfilling for dose optimization. In part 1, we assign backfilling patients by jointly considering efficacy and toxicity, and adaptively backfill them to the dose that optimizes the toxicity-efficacy trade-off. To further accelerate the trial, we adopt late-onset imputation for pending outcomes to allow for real-time dose assignment decisions for new patients. In part 2, after identifying the maximum tolerated dose, we implement an extended backfilling to select the optimal dose. Extensive simulation studies show that the proposed design achieves the dual goals of dose-finding and dose optimization in a seamless trial framework, while providing superior operating characteristics such as more efficient patient enrollment and improved dose selection.

Early Phase Clinical Trials in Oncology in Healthy Participants*Arinjit Bhattacharyya, Merck*

While most early phase oncology trials enroll cancer patients, selected early phase studies may involve healthy volunteers to evaluate investigational agents related to cancer treatment or supportive care. These trials generally focus on:

- Pharmacokinetics (PK) and Pharmacodynamics (PD): Assessing how the drug is absorbed, distributed, metabolized, and excreted, as well as its biological effects, in a controlled setting without the confounding effects of cancer or prior treatments.
- Safety and Tolerability: Determining initial safety profiles and identifying adverse effects at low doses before testing in cancer patients.
- Supportive Care Agents: Evaluating drugs aimed at managing side effects of cancer therapies (e.g., anti-nausea medications, growth factors) where testing in healthy volunteers is ethically

appropriate.

Imaging and Diagnostic Agents: Testing novel imaging tracers or diagnostic compounds that aid cancer detection or monitoring. However, cytotoxic or highly targeted anticancer agents with significant toxicity are rarely tested in healthy participants due to ethical concerns. Instead, these drugs proceed directly to patient-based trials. This talk will summarize the clinical trial designs and analysis for these studies in healthy participants.



Session Info

S114

TECHNICAL STATISTICAL SESSION

October 14th, 13:00 - 13:30 UTC

Quantitative Approaches to Safeguarding Child Welfare

USE SCHED TO ACCESS THE ZOOM ROOM

Safeguarding child welfare requires collaboration across government, public health, and community organizations. To understand how we can best serve the needs of youth across the globe, policy-makers often rely on insight coming from data. Thus, data analysts and (bio)statisticians are essential contributors to child welfare advocacy, as they can provide quantitative evidence to guide policy development.

In this session, sponsored by the Caucus for Women in Statistics, the speakers will discuss their work with linked data and big data in the child welfare policy space. First, Dr. Rameela Raman will describe examples of real-world policy-relevant work with a US child welfare system that highlights the advantages and challenges of using linked data. Dr. Svetlana Yampolskaya will then discuss the usage of big data and sub-group analysis to explore trajectories of substance use among youth in the child welfare system. Both of these speakers utilize data to drive positive changes in child welfare policy and ensure that all children have the support needed to thrive in their environment.

ORGANIZER: Ashley Mullan, Vanderbilt University

CHAIR: Catherine Chalikian, Johns Hopkins Applied Physics Lab



Speaker Bios

S114

**DR. RAMEELA RAMAN**

Vanderbilt University Medical Center

Rameela Raman, Ph.D., is an Associate Professor of Biostatistics at Vanderbilt University. She has published over 100 peer-reviewed articles and has been collaborating for over a decade as the lead statistician with the Vanderbilt Center of Excellence for Children in State Custody, an academic partner to the Tennessee Department of Children's Services. She works with child psychologists, epidemiologists, biostatisticians and substantive experts to formulate relevant questions, place the results in context and communicate the results to policymakers. Much of her work involves critically analyzing current practices and evaluating intervention effectiveness to guide research, policy, and advocacy for children in or at-risk of coming into state custody. Her research interests include leveraging multiple sources of data to incorporate social determinants to study outcomes of interest. Dr. Raman graduated with a MA and PhD in Biostatistics from the State University of New York at Buffalo.

**DR. SVETLANA YAMPOLSKAYA**

University of South Florida

Dr. Svetlana Yampolskaya is a Research Associate Professor in the Department of Child & Family Studies, University of South Florida. She is a sociologist by training but her primary expertise and research interests lie in improvement of well-being of at-risk youth. Within this area of research, Dr. Yampolskaya conducted numerous studies examining issues related to child maltreatment, fatal child abuse, mental health, and involvement with juvenile justice among child welfare involved youth. Another focus of her work has been the application of advanced statistical methods in behavioral health and child welfare services research, with specific emphasis on the utilization of large administrative data sets or data that have been previously collected. Dr. Yampolskaya's current research interests include prevention of adverse outcomes for youth involved in multiple systems as well as innovative methodological approaches.





Informing policy and implementation in the child welfare system using integrated administrative data

Rameela Raman, Vanderbilt University School of Medicine

Administrative data sources provide a wealth of information and are a valuable resource for research. However, they are limited in the amount of information they capture and may not provide a comprehensive picture. Recently, integrating data from other sources (e.g., neighborhood-level socioeconomic characteristics from the United States Census) and linking them to child welfare administrative data has been shown to be an approach to address the limitations. In this talk, I will present examples of real-world policy-relevant research projects from my work with a US child welfare system to motivate a discussion on the advantages and challenges of using linked data. I will also touch on the types of research questions that can be addressed through linked data and demonstrate interactive ways of presenting and communicating analysis results.

Big Data and Sub-Group Analysis: How Child Welfare Involvement Impacts Trajectories of Substance Use among Child Welfare Involved Youth

Svetlana Yampolskaya, University of South Florida

In recent years, the role that “big data” play in evaluation has attracted a great deal of attention. Besides cost-effectiveness, “big data,” or data that has been already collected, can be useful in providing answers to a variety of research and evaluation questions, allowing for examination of heterogeneous effects, as well as comparison of these effects across different groups of individuals. This presentation will focus on the use of a growth curve modeling approach with data collected for the longitudinal National Survey of Child & Adolescent Well-Being-II to examine the effect of child welfare involvement on substance use trajectories among youth with child maltreatment histories. This approach will allow for better understanding of youth needs and will illustrate the importance of tailoring programs and services to youth with different profiles in the system. Implications for targeted interventions and utilization of secondary data analysis will be discussed.



Session Info

S115

CAREER OR PROFESSIONAL DEVELOPMENT SESSION

October 14th, 13:30 - 14:00 UTC

From Ecology to Epidemiology: 2025 CWS Woodroffe Awardee's Journey

USE SCHED TO ACCESS THE ZOOM ROOM

This session features the 2025 Michael Woodroffe Award recipient, Dr. Staci Hepler, Associate Professor of Statistical Sciences at Wake Forest University and the Ollen R. Nally Faculty Fellow.

The Michael Woodroffe Award, presented by the Caucus for Women in Statistics and Data Science (CWS), honors a mid-career woman statistician or data scientist for outstanding contributions to real-world problem solving through the innovative application of statistical theory and methods.

In her presentation, Dr. Hepler will share her professional journey across disciplines, beginning with ecological and environmental statistics and evolving into public health applications. Her research highlights the power of Bayesian modeling and spatio-temporal analysis in addressing complex, data-rich challenges.

The talk will showcase key projects from both domains, including her recent work on the opioid epidemic, and will offer reflections on how a statistician's foundational skills can lead to meaningful impact across scientific frontiers.

This session will appeal to a broad audience interested in applied statistics, career development, and interdisciplinary research with real-world consequences.

ORGANIZER: Dong-Yun Kim, NHLBI/NIH**CHAIR:** Dong-Yun Kim, NHLBI/NIH

Speaker Bios

S115

<https://sites.google.com/a/wfu.edu/hepler>

PROF. STACI HEPLER

Wake Forest University

Dr. Staci Hepler is the Ollen R. Nally Faculty Fellow and Associate Professor of Statistical Sciences at Wake Forest University, where she also serves as Associate Chair of the department. Her research on spatial statistics and Bayesian modeling has been funded by the NIH and NSF, and her work focuses on applications in environmental statistics and statistical epidemiology.



Session Info

S116

TECHNICAL STATISTICAL SESSION

October 14th, 14:00 - 14:30 UTC

Bridging Science and Strategy: The Role of Statisticians in Drug Development

USE SCHED TO ACCESS THE ZOOM ROOM

Drug development is a complex, multi-phased process that aims to identify, evaluate, and ultimately bring safe and effective therapies for patients. Statisticians are integral to every stage of drug development from initial discovery through pre-clinical studies, clinical trials, chemical manufacturing and control, and post-market surveillance within multidisciplinary teams. In this panel, we will explore the unique contributions and challenges of statisticians working in the drug development space, highlighting how their responsibilities differ from those of statisticians in academic or other applied settings. Panelists, comprising experienced drug developer statisticians from industry, will share insights on their involvement across the development lifecycle, from study design and regulatory strategy to real-world evidence generation and benefit-risk assessment. We will also address the mindset shift required to become a drug developer who applies statistical thinking not only to data but to decision-making, clinical strategy, and cross-functional collaboration. Attendees will gain a deeper understanding of what it means to be a statistician and a drug developer, and how statisticians can shape the future of pharmaceutical innovation.

ORGANIZER: Umut Ozbek, Eli Lilly and Company**CHAIR:** Umut Ozbek, Eli Lilly and Company

Speaker Bios

S116

**W. SCOTT CLARK**

Retired

W. Scott Clark earned his B.S. in mathematics from the University of the South. He completed both his M.S. and Ph.D. at Emory University in the field of Biostatistics. From 1990 to 1998, he was an Assistant, then Associate Professor of Biostatistics at the Rollins School of Public Health at Emory University. He also served as the Director of the Biostatistics Consulting Center at Emory. In 1998, he joined Eli Lilly and Company as Senior Statistician. Since that time until his retirement, he served in numerous leadership roles across a variety of therapeutic areas and stages of drug development. He was Associate Vice President, Global Statistical Sciences, at Eli Lilly and Company, leading the statisticians supporting discovery, development, manufacturing, and a group in bioinformatics before his retirement. He is an ASA Fellow. His research interests include Markov modeling, clinical trials, and all aspects of statistical consulting.

**RICHARD HIGGS**

Eli Lilly and Company

Rick received master's degrees in electrical engineering (machine learning) and statistics at Purdue University. He has worked in discovery research at Lilly for over 30 years focusing primarily on high throughput technologies like mass spectrometry-based proteomics and RNAseq with application to target identification, biomarker identification, and translational science. He is currently leading a group of statisticians and genomics scientists supporting pre-clinical research at Lilly. He has made key contributions to multiple projects that have resulted in drug candidates that are in clinical development and the market. Rick is a co-inventor on several patents and a co-author on over 75 peer-reviewed manuscripts.



Session Info

S117

CAREER OR PROFESSIONAL DEVELOPMENT SESSION

October 14th, 14:30 - 15:30 UTC

Thriving in an Academic Statistical Consulting Center

USE SCHED TO ACCESS THE ZOOM ROOM

Academic statistical consulting centers collaborate and provide statistical support to faculty and graduate students for their research projects. A collaborative and respectful environment is of utmost importance for a thriving SCC. Inclusive leadership helps the SCC evolve and it builds trust and ownership among the statistical consultants. Graduate students research assistants build networks and improve communication skills and broaden their repertoire of statistical methods. A well design project database helps with managing competing demands efficiently and empowers the consultants. Everyone benefits if strategies for communication and cognitive accessibility are implemented. It benefits everyone, statistical consultants and their research collaborators.

ORGANIZER: Marianne Huebner, Michigan State University**CHAIR:** Marianne Huebner, Michigan State University

Speaker Bios

S117

**DR. SICHAO WANG**

Michigan State University

Sichao Wang is a statistician at the Center for Statistical Training and Consulting (CSTAT) at Michigan State University and holds a joint appointment as a biostatistician with the College of Veterinary Medicine's Clinical Trials Program (CLIP). She also serves as CSTAT's thematic research team leader for veterinary medicine studies. Her research focuses on applying rigorous and appropriate statistical methods to address clinically meaningful questions.

**MS. GUANQI LU**

Michigan State University

Guanqi Lu is a Biostatistics PhD candidate in the Department of Epidemiology and Biostatistics and a graduate student consultant in the Center of Statistical Training and Consulting (CSTAT) at Michigan State University. Her research interests focus on quantitative genetics and multi-omics data analysis.

**DR. WENJUAN MA**

Michigan State University

Wenjuan Ma is a statistician at the Center for Statistical Training and Consulting (CSTAT) at Michigan State University. She works on projects in a variety of disciplines including communication, media, information, social sciences, and education. Her research focus is on structural equation modeling, agent-based modeling, and mixed-effect modeling.

**DR. MARIANNE HUEBNER**

Michigan State University

Marianne Huebner is the Director of the Center of Statistical Training and Consulting (CSTAT) and a professor of the Department of Statistics and Probability at Michigan State University, USA. She is also a co-chair of the international STRATOS Initiative which is a large collaboration of statistics experts whose aim is to develop accurate and accessible guidance on design and analysis of observational studies. Her research encompasses statistical methodology in initial data analysis and applied statistics research in sports and aging and in health outcomes.



Inclusive Leadership and Leadership Teams: Fostering Trust, Collaboration, and Ownership

Sichao Wang, Michigan State University

Academic statistical consulting centers are collaborative environments—both within the center and in partnerships with clients. Building leadership teams with shared responsibilities for developing and nurturing stakeholder relationships builds support networks among the consultants and fosters a shared sense of ownership for the mission and vision of the statistical consulting center. It empowers consultants and utilizes talent and complementary expertise to develop strategic goals. It creates opportunities for professional growth. Joint appointments with partners can strengthen the connection and trust. However, partnerships also need to have higher level leadership support to navigate expectations and memorandum of understandings.

Thriving in Statistical Consulting as a Graduate Student: Skills Beyond Statistical Expertise

Guanqi Lu, Michigan State University

As a graduate student consultant in an academic statistical consulting center, thriving requires more than statistical knowledge. Working with clients from various academic backgrounds involves expressing research objectives statistically, developing clear analysis plans, and delivering interpretable results that address their research questions. At CSTAT, graduate student consultants learn and grow in a supportive environment. We have weekly staff meetings to discuss projects with other colleagues and receive regular guidance from senior consultants. These experiences help with not only technical growth but also empathy, adaptability, and the ability to adapt to different communication styles. It is also important for me to manage my time to balance personal research with consulting responsibilities. This talk will share strategies, challenges, and lessons learned that may benefit other graduate students and early-career consultants seeking to succeed in academic consulting environments.

The Critical Role of a Well-Designed Project Database in Empowering Statistical Consultants

Wenjuan Ma, Michigan State University

Statistical consulting is a unique professional practice. It is characterized by frequent intellectual context-switching across diverse disciplines. This feature demands rapid adaptation, precise communication, and meticulous documentation from a statistical consultant. This talk explores the critical role of a tailored operational project database developed within our unit to support effective consulting workflows. I illustrate how the project database streamlines bid generation, time tracking, note-taking, and documentation, enabling consultants to manage competing demands efficiently. Through practical examples, I highlight how this tool fosters continuity, reduces cognitive load, and enhances transparency and productivity. Ultimately, the project database exemplifies how purpose-built tools can empower statistical consultants to thrive in their complex and dynamic environment.

Facilitating Accessibility in Academic Statistical Consulting Centers

Marianne Huebner, Michigan State University

Academic statistical consulting centers collaborate and provide statistical support to faculty and graduate students for their research projects. Clients seeking statistical support may face communication and cognitive accessibility barriers that consultants may not be aware of. Many researchers do not disclose their personal circumstances, for example neurodivergent behaviors. Statistical consulting centers can facilitate accessibility by building a welcoming and inclusive environment and by training and mentoring in accessibility topics. Universal design for learning is a framework that aims to support individuals regardless of their abilities or learning styles, by providing multiple means of engagement, presentation, and action and expression. We offer strategies applicable to statistical consulting environments to improve access and empower clients and consultants. Benefits of adapting these principles for statistical consulting and taking time to assess specific needs of individuals improves experiences of clients and consultants and leads to high quality statistical support and successful completion of research projects.



Session Info

S118

OTHER

October 14th, 15:30 - 16:30 UTC

The Global Role of Women in Pharma Analytics: Scalable Impact Models & Measurement

USE SCHED TO ACCESS THE ZOOM ROOM

This session will explore the development of scalable impact models by the Business Insights & Analytics team at Eli Lilly. Our team builds the full end-to-end pipeline for measuring the effectiveness of pharma marketing analytics thereby enabling data-driven decision making for marketing spend. The robustness of these analyses and recommendations are becoming increasingly important for global companies to make decisions across a large portfolio of disease states, countries, and customers. Join us to learn about:

1. Data Engineering: Integrating diverse data sources into useable data products and tools
2. Modeling: Machine learning models to evaluate marketing impact and optimize budget allocation, ensuring data-driven decisions that maximize return on investment
3. Operational Execution: Scalable pipelines and technology to enable fast and reliable results delivery
4. Influencing Decisions: Sharing results, partnering with teams, and influencing outcomes.

ORGANIZER: Ayushi Singh, Eli Lilly & Co**CHAIR:** Ayushi Singh, Eli Lilly & Co

Speaker Bios

S118



MS. SUMEGHA SETIA

Eli Lilly & Co

Sumegha Setia is a Consultant in Data Engineering at Eli Lilly and Company, with nine years of experience across e-commerce, telephony, conversational media platforms, and pharmaceuticals. Her diverse professional background has provided her with deep exposure to a wide array of modern tools and technologies, enriching both her technical skill set and domain understanding. She began her career as a full-stack web developer before transitioning into the data space, where she discovered a strong passion for building scalable data systems and driving analytics solutions. Sumegha holds a Bachelor of Technology in Computer Science Engineering and an Executive MBA in International Business from the Indian Institute of Foreign Trade (IIFT). She is currently focused on advancing her expertise in data engineering and analytics, with a keen interest in how these fields can transform decision-making and operational efficiency across industries.



MS. ANANYA KARMAKAR

Eli Lilly & Co

Ananya Karmakar is seasoned analytics professional with 7 years of experience across consultancy, technology, and pharmaceutical domains. Currently a Sr. Data Scientist II at Eli Lilly, she has spent 4 years applying advanced analytics to drive impactful decisions. With honors and master's degrees in economics, specializing in econometrics, Ananya thrives on building statistical models, coding, debugging, and solving complex problems. Her passion lies in applying her academic foundation to real-world challenges—from simple regressions to intricate market mix models.



MS. AYUSHI SINGH

Eli Lilly & Co

Ayushi Singh is a dynamic analytics leader at Eli Lilly & Co. with over a decade of experience, where she serves as Senior Manager in the Business Insights & Analytics (BI&A) team. With a focus on operational scalability and innovation, Ayushi leads initiatives that streamline the end-to-end analytics pipeline—from data engineering and model deployment to decision influence and impact measurement. Ayushi was tech lead at IBM India where she led watson.ai innovation team. Ayushi holds a master's in computer application degree from GBTU Lucknow UP. Ayushi is also actively engaged in leadership development and talent coaching, contributing to Lilly's career growth frameworks and inclusive analytics culture.



MRS. RACHEL SILVESTRINI

Eli Lilly & Co

Rachel Silvestrini is the Vice President of Business Insights and Analytics (BI&A) at Eli Lilly and Company. She leads the high performing BI&A team to deliver analytical insights and cutting-edge AI solutions that drive business decisions and impact in the commercial space. Her team consists of data engineers, data analyst, data scientists, ML engineers, and business integrators. Prior to joining Eli Lilly and Company, Rachel was an Associate Professor at the Rochester Institute of Technology in Rochester, NY. She holds a PhD and MS in Industrial Engineering from Arizona State University and a BS in Industrial Engineering from Northwestern University.



Data Engineering: Integrating Diverse Data Sources Into Useable Data Products and Tools*Sumegha Setia, Eli Lilly & Co*

Data Engineering: Integrating Diverse Data Sources into Useable Data Products and Tools
Presented by Sumegha Setia, Consultant – Data Engineering, Eli Lilly and Company

In an era where data is both abundant and fragmented, the ability to transform disparate sources into coherent, analytics-ready products is a cornerstone of modern enterprise success. This session explores the critical role of data engineering in bridging the gap between raw data and actionable insights.

Drawing on her nine years of experience across e-commerce, telephony, conversational media platforms, and pharmaceuticals, Sumegha Setia will share practical strategies for building scalable data systems that support decision-making across diverse business functions. From her early career as a full-stack developer to her current role as a data engineering consultant at Eli Lilly, Sumegha has cultivated a deep understanding of how to harmonise data pipelines with business needs.

The talk will cover:

Techniques for integrating structured and unstructured data from multiple domains.
Best practices for designing reusable ingestion components and scalable processing frameworks.
Real-world examples of transforming legacy systems and siloed datasets into unified data products that power analytics and machine learning.
Attendees will gain insights into how modern data engineering enables operational efficiency, supports cross-functional collaboration, and drives measurable business impact.

AI Model Development*Ananya Karmakar, Eli Lilly & Co*

Modeling: Machine Learning Models to Evaluate Marketing Impact and Optimize Budget Allocation, Ensuring Data-Driven Decisions That Maximise Return on Investment
Presented by Ananya Karmakar, Data Scientist – Business Insights & Analytics, Eli Lilly and Company
In today's competitive pharmaceutical landscape, marketing decisions must be both agile and analytically grounded. This session explores how machine learning (ML) models are transforming the way organisations evaluate marketing impact and optimise budget allocation to maximise return on investment.

Drawing from her work on the BI&A Impact Models at Eli Lilly, Ananya Karmakar will showcase how advanced ML techniques—ranging from ensemble learning to simulation-based optimisation—are used to quantify the effectiveness of marketing channels and guide strategic spend decisions. Her talk will highlight the integration of econometric principles with scalable ML pipelines to deliver actionable insights across therapeutic areas and geographies.

Key takeaways include:

1. How ML models are used to simulate marketing scenarios and predict outcomes with high accuracy.
2. The role of optimisation solvers in identifying budget allocations that yield the highest impact.
3. Real-world examples of model deployment and validation within Lilly's dynamic targeting framework.

Attendees will gain a deeper understanding of how data science can elevate marketing strategy, improve cross-functional alignment, and drive measurable business value.

Operational Execution: Scalable Pipelines and Technology to Enable Fast and Reliable Results Delivery*Ayushi Singh, Eli Lilly & Co*

Operational Execution: Scalable Pipelines and Technology to Enable Fast and Reliable Results Delivery
Presented by Ayushi Singh, Senior Manager – Business Insights & Analytics, Eli Lilly and Company

In the fast-paced world of pharmaceutical analytics, operational execution is the bridge between innovation and impact. This session explores how scalable pipelines and modern MLOps technologies are transforming the delivery of analytics solutions—making them faster, more reliable, and enterprise-ready.

Drawing from her experience leading global MLOps initiatives at Eli Lilly, Ayushi Singh will share how her team standardised and automated the project lifecycle—from development to deployment and operations—to accelerate innovation and reduce turnaround times. Her talk will highlight the integration of tools like MLflow, Prefect, and CI/CD pipelines to enable self-service workflows and proactive performance management.

Key themes include:

1. Building reusable, versioned feature stores and template repositories for consistent development.
2. Enabling cross-functional collaboration through knowledge transfer sessions and SME development.
3. Scaling AI delivery by removing technical and organisational barriers and aligning with global impact metrics.

Attendees will gain insights into how operational excellence in data science and engineering can unlock agility, foster experimentation, and ensure that analytics solutions are not only built—but delivered—with speed and confidence.

Influencing Decisions: Sharing Results, Partnering With Teams, and Influencing Outcomes*Rachel Silvestrini, Eli Lilly & Co*

Influencing Decisions: Sharing Results, Partnering with Teams, and Influencing Outcomes
Presented by Rachel Silvestrini, VP – Business Insights & Analytics, Eli Lilly and Company

In a world of increasingly complex data ecosystems and fragmented media landscapes, the ability to influence decisions through analytics is more critical than ever. This session explores how strategic communication, cross-functional collaboration, and data storytelling can elevate analytics from insight generation to outcome ownership.

Drawing on her experience leading global BI&A teams at Eli Lilly, Rachel Silvestrini will share how her organisation builds trust and alignment across diverse stakeholders—from marketing and medical affairs to executive leadership—by translating analytical outputs into compelling narratives that drive action. Her talk will highlight the importance of owning the measurement process, partnering early in the decision cycle, and tailoring insights to the needs of each audience.

Key themes include:

1. Building influence through transparency, consistency, and shared accountability.
2. Creating feedback loops that improve both model performance and stakeholder engagement.
3. Leveraging data as intellectual property to shape brand strategy and corporate reputation.

Attendees will gain practical strategies for becoming trusted advisors in their organisations, ensuring that analytics not only inform—but influence—the decisions that matter most.



Session Info

S119

OTHER

October 14th, 16:30 - 17:00 UTC

Pop Stats: Where Fandom Meets Functions

USE SCHED TO ACCESS THE ZOOM ROOM

Statistics isn't just for clinical trials and economic forecasts—it can also illuminate the worlds we escape to in books, movies, music, and television. In this session, three women statisticians bring rigorous methods to unexpected, entertaining data sources: survival analysis in *The Hunger Games*, interrupted time series for Taylor Swift's impact on football podcast popularity, and classification and text mining using data from the series *Severance*. Together, these talks show how statistical tools can uncover surprising patterns in the stories and cultural moments that capture our attention, proving that data science has a place in both the lab and the living room.

ORGANIZER: Sarah Lotspeich, Wake Forest University**CHAIR:** Sarah Lotspeich, Wake Forest University

Speaker Bios

S119



MS. TRISTAN BULLOCK

Smith College

Tristan Bullock recently graduated with honors from Wake Forest University with a B.S. in Mathematics and a minor in Biology. She is currently attending Smith College as a Post Baccalaureate student in mathematics and is applying to PhD programs in both mathematics and biostatistics. Her current research interests include mathematical modeling, statistical inference, and topology. For her undergraduate thesis, she explored how the Ising model, a tool from statistical physics used to describe molecular spin behavior, relates to knot theory. In particular, she studied how planar graphs with signed edges can represent knot diagrams, and how the complex number produced by the Ising model might help distinguish between different knots and links. Tristan is especially interested in using mathematical tools to understand real-world patterns in science and health.

<https://ashleymullan.github.io/>



MS. ASHLEY MULLAN

Vanderbilt University

Ashley Mullan is a second year Ph.D. student in biostatistics at Vanderbilt University. She earned her master's degree in Statistics and graduate certificate in Data Science from Wake Forest University, and she earned her bachelor's degrees in Applied Mathematics and Philosophy from the University of Scranton. Ashley's main (bio)statistical interests include developing and applying methods for mismeasured or missing data, collaborating with child welfare policy makers to enhance the quality of services provided to at-risk children, and supporting the next generation of (bio)statistics superstars. In her spare time, she collects and analyzes data about her own pop culture consumption habits. She's an avid fan of (sit/rom)com, (data) comm, and (oxford) commas!

<https://www.lucymcgowan.com/>



DR. LUCY D. MCGOWAN

Wake Forest University

Lucy D'Agostino McGowan is an associate professor in the Department of Statistical Sciences at Wake Forest University. She received her PhD in Biostatistics from Vanderbilt University and completed her postdoctoral training at Johns Hopkins University Bloomberg School of Public Health. Her research focuses on causal inference, statistical communication, analytic design theory, and data science pedagogy. Dr. D'Agostino McGowan was the 2023 chair of the American Statistical Association's Section on Statistical Graphics and can be found blogging at livefreeordichotomize.com, on Twitter @LucyStats, and podcasting on the American Journal of Epidemiology partner podcast, Casual Inference.





Blood, Stats, and Tears: Survival in the Arena

Tristan Bullock, Smith College

This project applies survival analysis to character outcomes in *The Hunger Games* (2012), examining how traits such as sex, age, and rank influence survival time. Using data from 24 characters with a high event rate (22 deaths), equal gender distribution, and an average age of 15.79 years, we modeled survival using Kaplan-Meier estimates and compared distributions by sex using the log-rank test. A Cox proportional hazards model was fitted with sex and rank as categorical predictors, and age as a continuous covariate in a secondary analysis. Confidence intervals for survival curves were calculated using Greenwood's formula, and Wald's test assessed uncertainty in the Cox model. Results will be presented with Kaplan-Meier survival plots, forest plots for hazard ratios, and summary statistics. This project highlights how survival analysis can offer meaningful insights into narrative structure and demonstrates the broader applicability of statistical tools in analyzing popular media.

Talking Over a Football Game: The Taylor Swift Effect on Football Podcast Popularity

Ashley Mullan, Vanderbilt University

Since the release of her first album in 2006, Taylor Swift has transformed the modern music industry. She often writes songs about the ups and downs of her dating life, infusing traits of the men she loves into her lyrics. Swift's current partner, Kansas City Chiefs tight end Travis Kelce, has appeared in some versions of her song "Karma" and also as a guest performer on her record-breaking Eras Tour. Kelce also helped her promote her new album *Life of a Showgirl* by inviting her as a guest on *New Heights*, the podcast he shares with his brother, fellow (American) football star Jason Kelce. The audience of *New Heights* is primarily composed of football fans, as the brothers use their platform to discuss life in the National Football League and give their perspective on sports headlines. However, Swift's appearance on the podcast sparked a deviation from their regularly scheduled programming. In this talk, we will use interrupted time series to explore how Swift's appearance on *New Heights* affected its viewership, measured in terms of search volume.

The Art of Data Refinement: Severance Analyses

Lucy D'Agostino McGowan, Wake Forest University

This talk demonstrates data extraction from multiple sources using the popular television series *Severance* as an example. For example, we collected and analyzed elevator sounds predict narrative events, voice recordings underwent cepstral analysis to estimate fundamental frequencies and characterize speaker-specific distributions, with k-nearest neighbors used for classification, and text mining was performed on episode scripts to quantify dialogue patterns. These analyses illustrate how statistical methods can be applied to unconventional data sources from entertainment media.



Session Info

S120

TECHNICAL STATISTICAL SESSION

October 14th, 18:00 - 18:30 UTC

Prompting Progress: Integrating LLMs into Statistical Research

USE SCHED TO ACCESS THE ZOOM ROOM

This session highlights how statisticians are incorporating large language models (LLMs) into their research to enhance and automate complex tasks. Speakers will share work using LLMs to generate causal diagrams, suggest data validation and augmentation algorithms, and classify hate speech at scale, illustrating both methodological innovation and real-world impact. Together, the talks explore the power and pitfalls of using LLMs in statistical workflows.

ORGANIZER: Sarah Lotspeich, Wake Forest University**CHAIR:** Sarah Lotspeich, Wake Forest University

Speaker Bios

S120

**DR. LUCY D. MCGOWAN**

Wake Forest University

<https://www.lucymcgowan.com/>

Lucy D'Agostino McGowan is an associate professor in the Department of Statistical Sciences at Wake Forest University. She received her PhD in Biostatistics from Vanderbilt University and completed her postdoctoral training at Johns Hopkins University Bloomberg School of Public Health. Her research focuses on causal inference, statistical communication, analytic design theory, and data science pedagogy. Dr. D'Agostino McGowan was the 2023 chair of the American Statistical Association's Section on Statistical Graphics and can be found blogging at livefreeordichotomize.com, on Twitter @LucyStats, and podcasting on the American Journal of Epidemiology partner podcast, Casual Inference.

**DR. DAPHNA HAREL**

New York University

<https://steinhardt.nyu.edu/people/daphna-harel>

Dr. Daphna Harel is an Associate Professor of Applied Statistics at New York University. Her Queering and Elevating Empirical Research (QUEER) data lab works on projects at the intersection on LGBTQIA+ related issues and statistics.

**DR. SARAH LOTSPEICH**

Wake Forest University

<https://www.sarahlotspeich.com/>

Sarah Lotspeich is an Assistant Professor of Statistics at Wake Forest University. She enthusiastically mentors student research and co-leads collaborative labs at Wake Forest and the University of North Carolina (UNC) at Chapel Hill. She co-organizes Florence Nightingale Day at Wake Forest annually, engaging local students in statistics and data science, and holds elected positions in the Caucus for Women in Statistics and Data Science and other organizations. Sarah completed a postdoctoral fellowship at UNC Chapel Hill and earned her Ph.D. in Biostatistics from Vanderbilt University. Her research tackles challenges in analyzing error-prone data, focusing on international HIV cohorts, electronic health records, and health disparities. She also develops methods for statistical modeling with censored covariates, applicable to Huntington's disease. When she's not writing code, you can find Sarah cross-stitching, adventuring in new places, or rewatching her favorite TV shows.



Exploring the Potential of Large Language Models in Generating Saturated DAGs for Causal Inference

Lucy D'Agostino McGowan, Wake Forest University

This talk investigates whether large language models (LLMs) could potentially assist in the creation of "saturated DAGs", graphical representations that exhaustively map all possible causal pathways in a system. We'll critically examine if and how LLMs might help identify the full space of plausible causal relationships that traditional approaches may overlook. The presentation will assess the strengths and limitations of prompting LLMs to generate comprehensive causal structures, identify backdoor paths, and navigate complex causal systems.

accuracy of ICD-derived data augmentation algorithms relative to expert chart review (the current gold standard).

The Causal Effect of a Content Moderation Policy Change on the Incidence of LGBTQ2+ Directed Hate Speech on Twitter

Daphna Harel, New York University

Social media companies have amassed a great deal of power and individuals spend a large amount of their time interacting with these platforms and the posts on them. However, some online social media platforms are rampant with hate speech, which individuals must contend with when engaging with the network. Specifically, the impact of hate speech on mental health has been well documented, particularly for marginalized communities, who run the highest risk of being targeted. This work analyzes a total of 92 million tweets collected from 2006 to 2020. First, we explore how to determine whether a tweet can be categorized as hate speech through scalable methodology. Next, we explore whether companies' decisions to either enact or fail to enact content moderation policies has measurable impacts on the incidence of hate speech. During the window of data collection, Twitter changed its content moderation policy regarding hate speech, and, therefore, using interrupted time series methods, we estimate the causal effect of this policy change on the levels of hate speech.

Applying LLMs to Support Data Validation and Augmentation in EHR Research

Sarah Lotspeich, Wake Forest University

Data from electronic health records (EHR) present a huge opportunity to operationalize computable phenotypes, like whole-person health scores, in learning health systems; however, these data are prone to missingness and errors. Validation (e.g., through expert chart reviews) can provide better-quality data, but realistically, only a subset of patients' data can be validated, and most protocols do not recover missing data. Previously, we devised an "enriched" validation protocol that incorporates clinical expertise to promote EHR data quality and completeness. Chart reviewers (clinical research trainees) were given a "roadmap" of auxiliary diagnoses to search for in electronic patient charts to replace missing values (e.g., a diagnosis of impaired glycemic control might imply that a missing hemoglobin A1C value would be considered unhealthy). In this talk, we first compare an algorithmic approach to implementing the same roadmap computationally, on the basis of International Classification of Diseases (ICD) codes from the EHR. Then, we consider new roadmaps that were iteratively refined using large language models (LLMs) in conjunction with clinical expertise to expand the list of auxiliary diagnoses, seeking to improve the



Session Info

S121

TECHNICAL STATISTICAL SESSION

October 14th, 18:30 - 19:30 UTC

Methods with Meaning: Celebrating The "Why" Behind Statistical Innovation

USE SCHED TO ACCESS THE ZOOM ROOM

Statistical innovation is often judged by novelty — but what if we instead celebrated our methods' meaning? This session highlights the real-world motivations, challenges, and impact that drive methodological development. Rather than diving deep into equations, in this session four statisticians share the stories behind their work: why the methods were needed; what gap they fill; and how they contribute to science, policy, or public health. This session is for anyone who believes that statistical methods are not just intellectual exercises, but tools to understand and improve the world. Talks will explore bias in risk prediction, the responsible use of predictions in research, scalable validation of electronic health records, and optimizing cancer screening programs. Each project is rooted in a real-world setting-like healthcare, clinical research, or data science practice-where thoughtful methods can lead to meaningful improvements.

ORGANIZER: Sarah Lotspeich, Wake Forest University**CHAIR:** Sarah Lotspeich, Wake Forest University

Speaker Bios

S121



DR. LUCY D. MCGOWAN

Wake Forest University

<https://www.lucymcgowan.com/>

Lucy D'Agostino McGowan is an associate professor in the Department of Statistical Sciences at Wake Forest University. She received her PhD in Biostatistics from Vanderbilt University and completed her postdoctoral training at Johns Hopkins University Bloomberg School of Public Health. Her research focuses on causal inference, statistical communication, analytic design theory, and data science pedagogy. Dr. D'Agostino McGowan was the 2023 chair of the American Statistical Association's Section on Statistical Graphics and can be found blogging at livefreeordichotomize.com, on Twitter @LucyStats, and podcasting on the American Journal of Epidemiology partner podcast, Casual Inference.



DR. MARISSA ASHNER

Duke University

Marissa Ashner is a collaborative biostatistician and graduate instructor in the Department of Biostatistics and Bioinformatics at Duke University. Marissa is passionate about developing best practices for applied biostatistical methods and enhancing statistics education for both statisticians and non-statisticians. Her collaborative work lies in the Center for Aging and Human Development, where she works with clinical researchers on a variety of aging-related research studies and analyses. Marissa completed her PhD in Biostatistics at University of North Carolina Chapel Hill. When she's not thinking about statistics, Marissa loves to lift weights, hike, and try new local restaurants.



DR. REBECCA HUBBARD

Brown University

<https://vivo.brown.edu/display/rhubbar1>

Dr. Hubbard's research focuses on the development and application of methods to improve analyses using real world data sources including electronic health records (EHR) and medical claims data. The data science era demands novel analytic methods to transform the wealth of data created as a byproduct of digital interactions into valid and generalizable knowledge. Dr. Hubbard's research emphasizes statistical methods designed to meet this challenge by addressing the messiness and complexity of real world data including informative observation schemes, phenotyping error, and error and missingness in confounders. Her methods have been applied to support the advancement of a broad range of research areas through use of EHR and claims data including health services research, cancer epidemiology, aging and dementia, and pharmacoepidemiology.



DR. YATES COLEY

Kaiser Permanente Washington Health Research Institute

<https://kpwashingtonresearch.org/index.php/our-research/our-scientists/coley-yates>

Yates Coley is an Associate Investigator at Kaiser Permanente Washington Health Research Institute. Their research focuses on development and implementation of clinical prediction models that are accurate, equitable, and actionable. Dr. Coley's current R01 aims to develop statistical methods to reduce racial and ethnic disparities in the performance of suicide risk prediction models and, more generally, to further develop the statistical machinery for designing prediction studies and evaluating prediction modeling methodologies. They collaborate in a range of topic areas including mental health, cancer, aging, and health system operations.



The Role of Congeniality in Multiple Imputation for Doubly Robust Causal Estimation

Lucy D'Agostino McGowan, Wake Forest University

This talk provides clear and practical guidance on the specification of imputation models when multiple imputation is used in conjunction with doubly robust estimation methods for causal inference. Through theoretical arguments and targeted simulations, we show that when a confounder has missing data the corresponding imputation model must include all variables used in either the propensity score model or the outcome model, and that these variables must appear in the same functional form as in the final analysis. Violating these conditions can lead to biased treatment effect estimates, even when both components of the doubly robust estimator are correctly specified. We present a mathematical framework for doubly robust estimation combined with multiple imputation, establish the theoretical requirements for proper imputation in this setting, and demonstrate the consequences of misspecification through simulation. Based on these findings, we offer concrete recommendations to ensure valid inference when using multiple imputation with doubly robust methods in applied causal analyses.

Predictors of Physical Resilience in Older Adults Undergoing Total Knee Arthroplasty: Biomarker Analysis From the PRIME-KNEE Study

Marissa Ashner, Duke University

The Physical Resilience Indicators and Mechanisms in the Elderly (PRIME) Collaborative aims to identify predictors of physical resilience, physical resilience as the ability to recover from a health stressor. Total knee arthroplasty (TKA), a common elective surgical intervention for older adults with knee osteoarthritis, provides an opportunity to study physical resilience due to its planned nature. The PRIME-KNEE study enrolled approximately 200 adults over the age of 60 undergoing TKA. We collected pre-stressor physical, cognitive, and psychosocial reserve data, various biomarkers, and other clinical and demographic characteristics. This analysis focuses on four specific outcomes: pain intensity, pain interference, lower extremity physical activities of daily living (PADLs), and step counts. We define physical resilience in these domains using the Expected Recovery Differential (ERDs) approach, which compares actual outcome measures to predicted measures based on a sample-derived model, built using elastic net penalized regression. To identify the most predictive biomarkers of the outcome-specific ERDs, we employed penalized regression models with various penalty parameters. Resampling techniques were used to ensure stability and reliability. Biomarkers were ranked based on their selection probabilities, and the most stable sets were identified based on their frequency of selection across multiple analyses.

Prediction Algorithms and Persistence: The Long Road to a Simple Solution

Rebecca Hubbard, Brown University

In 2014, I had a dream to move research using electronic health records (EHR) from a bias-plagued backwater a colleague once told me she “wouldn’t touch with a ten-foot pole” to something respectable. Armed with a band of plucky collaborators, that

dream, and a grant, I set out to tackle a common problem, how to properly analyze probabilistic phenotypes, outcomes derived from prediction algorithms. Specifically, since many outcomes of interest in EHR research are not explicitly captured in the data, researchers commonly employ prediction algorithms to deduce them from the observed data elements, resulting in outcome measures that are probabilistic rather than dichotomous. The goal sounded simple. The path was anything but. Along the way, I learned that in EHR research, not only are the data messy, but so are the milestones, meetings, and miscommunications. After a long and circuitous journey involving simulations, revisions, and existential doubt, we arrived at an elegant solution, a correction factor to reduce bias in studies using probabilistic phenotypes so simple you can calculate it on the back of a napkin. In this talk, I will share both the method and the non-linear path from inception to publication, including false starts, detours, and persistence, equal parts stubbornness and caffeine, that eventually paid off.

Do Clinical Prediction Models Perpetuate Health Disparities? Assessment of Racial and Ethnic Disparities in Suicide Prediction Models.

Yates Coley, Kaiser Permanente Washington Health Research Institute

Clinical prediction models are frequently estimated using health records data, which reflect inequities in health care access, quality, and outcomes based on race, sexual orientation, gender identity, immigration status, and other identities. Presuming that historical treatment patterns reflect clinical need can perpetuate these inequities. This talk will examine racial and ethnic disparities in the performance of suicide risk prediction models, a context which illustrates several methodological and implementation challenges of applying prediction models across racial and ethnic groups. We will demonstrate how to evaluate potential disparities in the performance and impact of clinical prediction models and discuss the role of statisticians in deploying clinical prediction models that reduce, rather than exacerbate, health disparities.



Session Info

S122

TECHNICAL STATISTICAL SESSION

October 14th, 19:30 - 20:15 UTC

Data at Full Tilt: Accelerating Discovery with Wearables and Smartphones

USE SCHED TO ACCESS THE ZOOM ROOM

The rise of wearable devices and smartphones has ushered in a new era of rich, real-time, and context-aware data. These technologies hold tremendous promise for advancing population health, behavioral science, and precision medicine—yet their integration into rigorous statistical workflows presents both opportunities and challenges. Sponsored by the Caucus for Women in Statistics, this session brings together three researchers leveraging mobile data to drive discovery. The talks will span hospital settings, smartphone-based mobility tracking, and fertility research—each highlighting how statisticians are adapting methods to make sense of high-volume, high-frequency, and sometimes incomplete data from personal devices.

ORGANIZER: Sarah Lotspeich, Wake Forest University**CHAIR:** Sarah Lotspeich, Wake Forest University

Speaker Bios

S122

**DR. DAIQI GAO**

Harvard University

<https://sites.google.com/view/daiqigao>

Daiqi Gao is a postdoctoral fellow in the Department of Statistics at Harvard University. She received her Ph.D. in Statistics and Operations Research from the University of North Carolina at Chapel Hill, and her B.S. in Industrial Engineering and Statistics from Tsinghua University. Her research focuses on optimizing and evaluating personalized policies in sequential decision-making, leveraging techniques from reinforcement learning, machine learning, and causal inference.

**DR. IRINA GAYNANOVA**

University of Michigan

<https://irinagain.github.io/>

Irina Gaynanova is an Associate Professor of Biostatistics at the University of Michigan. Her research focuses on statistical methods for high-dimensional biomedical data, with a particular emphasis on data integration and machine learning, driven by challenges in multi-omics and wearable device data, specifically continuous glucose monitors. Dr. Gaynanova's contributions have been recognized with several prestigious awards, including the David P. Byar Young Investigator Award, an NSF CAREER Award, the IMS Thelma and Marvin Zelen Emerging Women Leaders in Data Science Award, and the COPSS Emerging Leader Award.

**DR. HAOCHANG SHOU**

University of Pennsylvania

Haochang Shou, Ph.D., is Associate Professor of Biostatistics and Associate Director of Penn Statistics in Imaging and Visualization Endeavor (PennSIVE) in the Department of Biostatistics, Epidemiology and Informatics at University of Pennsylvania. Her research focuses on developing novel functional data analysis and machine learning models for integrating multimodal and longitudinal measures from brain imaging and wearable device, and assessing data reproducibility. She has led broad biomedical collaborative research and methodological development in areas of mental disorders, aging, neurodegenerative disease and specializes in analyzing high-density human activity measurements in mobile health (mHealth).



Optimizing and Querying Intermediate Outcomes via Reinforcement Learning for Better Long-Term Health

Daiqi Gao, Harvard University

HeartSteps is a mobile health intervention designed to help sedentary users increase and maintain physical activity. A key component involves delivering activity suggestions via smartphone notifications to encourage short bouts of physical activity (PA). We use online reinforcement learning (RL) algorithms to learn the optimal policy for delivering notifications based on each user's evolving needs. The reward in the RL algorithm is defined as a user's affective association with PA, an intermediate outcome measured through weekly surveys. While there are multiple decision times per week, the reward is only observed at the end of the week, creating a challenge of bagged decision times with non-Markovian and non-stationary dynamics within each bag. To address this challenge, we leverage an expert-provided causal directed acyclic graph (DAG) to construct the states and formulate the problem as a periodic Markov decision process (MDP). Further, the survey provides information about the latent states and rewards, but answering it can be burdensome and may reduce user engagement. We aim to optimize the timing of sending surveys to balance information gain with the risk of disengagement. This active querying problem is formulated as a periodic partially observable MDP, where the latent state is fully revealed only when the survey is answered.

Beyond Fixed Thresholds: Optimizing Summaries of Wearable Device Data via Piecewise Linearization of Quantile Functions

Irina Gaynanova, University of Michigan

Wearable devices, such as actigraphy monitors and continuous glucose monitors (CGMs), capture high-frequency data, which are often summarized by the percentages of time spent within fixed thresholds. For example, actigraphy data are categorized into sedentary, light, and moderate-to-vigorous activity, while CGM data are divided into hypoglycemia, normoglycemia, and hyperglycemia based on a standard glucose range of 70–180 mg/dL. Although scientific and clinical guidelines inform the choice of thresholds, it remains unclear whether this choice is optimal and whether the same thresholds should be applied across different populations. In this work, we define threshold optimality with loss functions that quantify discrepancies between the full empirical distributions of wearable device measurements and their discretizations based on specific thresholds. We introduce two loss functions: one that aims to accurately reconstruct the original distributions and another that preserves the pairwise sample distances. Using the Wasserstein distance as the base measure, we reformulate the loss minimization as optimal piecewise linearization of quantile functions. We solve this optimization via stepwise algorithms and differential evolution. We also formulate semi-supervised approaches where some thresholds are predefined based on scientific rationale.

Harmonization of Longitudinal Physical Activity Intensity Measures From Multiple Sources

Haochang Shou, University of Pennsylvania

Wearable devices and digital phenotyping are increasingly used in observational and interventional studies to assess physical

activity. However, integrating and comparing data across studies and cohorts remains challenging due to variability in device types, acquisition protocols, and preprocessing methods. A key challenge is removing unwanted study- or device-specific effects while preserving meaningful biological signals. These difficulties are exacerbated by the longitudinal and within-day correlations inherent in high-resolution time series data collected from wearable sensors. To address this, we propose INTACT (INtegration of Time series data from weArable sensors for physiCal acTivity), a novel method for harmonizing physical activity intensity time series from accelerometers. INTACT models shared information through common eigenvalues and eigenfunctions while allowing for domain-specific scale and rotation adjustments. We demonstrate the effectiveness of INTACT on two cohorts from the National Health and Nutrition Examination Survey (NHANES), where physical activity measures were collected using different generations of accelerometers and processed into different units. Our results show that INTACT outperforms existing methods in mitigating domain effects while preserving biological signals, enabling more reliable cross-study comparisons of physical activity patterns.



Session Info

S123

TECHNICAL STATISTICAL SESSION

October 14th, 20:00 - 20:15 UTC

Speed Session 1

USE SCHED TO ACCESS THE ZOOM ROOM

This session is a short session featuring three speed talks. We hope that you will contact the speakers if you are interested in discussing their work further.

ORGANIZER: Cynthia Bland, CWS

CHAIR: Jessica Kohlschmidt, The Ohio State University



Speaker Bios

S123



MISS SHARON AYAYO

University of Manchester

Sharon Ayayo is a final-year PhD student in Biostatistics at the University of Manchester, working on Cardiovascular Big Data. Her research focuses on clinical outcomes of patients with acute myocardial infarction in the UK using national electronic health records. She has 2 first author publications from her PhD, and has presented her research at 2 national conferences in the UK (Health Services Research and Young Statisticians Meeting). Before her PhD, she worked as a clinical trial statistician at Cardiff University, and has 3 co-authored publications. Sharon is passionate about using data to drive meaningful improvements in healthcare, particularly in cardiovascular outcomes and health equity. Outside of research, she advocates for inclusive research practices and supports early-career researchers from underrepresented backgrounds.

<https://www.wendychanw.com>


DR. WENDY CHAN

University of Pennsylvania

Dr. Wendy Chan is a statistician whose research focuses on applied statistical methods to improve generalizations from small studies in education. She seeks to improve statistical models and analysis so research can better inform policy and practice. Her work tries to unpack the intellectual and statistical challenges of generalizing the results of localized randomized trials to larger populations. She began her career in education as a member of Teach for America, where she taught sixth- and eighth-grade mathematics in a large middle school in New York City.

<https://www.linkedin.com/in/jacquelynneal/>


DR. JACQUELYN NEAL

WebMD Ignite

Dr. Jacquelyn E Neal completed her PhD in Biostatistics at Vanderbilt University under the supervision of Dr. Dandan Liu, where her research focused on predictive modeling for longitudinal outcomes in Alzheimer's disease. She also has a MS in Biostatistics from Vanderbilt University and a BS in Agricultural and Biological Engineering from the University of Florida. After taking a break to reevaluate her life and career goals, she worked full-time in industry analytics roles while completing her PhD research. From an analyst role using Excel for Operations at Virtual Inc, to a product analytics role at a patient education non-profit, to now leading the development of longitudinal marketing analytics products at WebMD Ignite, she's seen the value of her statistics training in her daily work. This year, she's also served as a Research Mentor for the first cohort of the AIM-AHEAD Bridge2AI for Clinical Care program, where she helps trainees grow into their own ideal careers.



Drivers of 1-Year Mortality Decline After Acute Myocardial Infarction in England and Wales: A 15-Year National Cohort Study

Sharon Ayayo, University of Manchester

Background: One-year mortality following acute myocardial infarction (AMI) has declined over time, yet the reasons for this improvement remain unclear. Understanding the drivers of these changes is essential for informing clinical strategies and health policy.

Methods: We analysed 852,914 adult patients admitted with AMI across England and Wales between 2005 and 2019 using national registry data. We examined changes in 1-year all-cause mortality and quantified the contribution of clinical, treatment, and demographic factors to this trend using decomposition analysis.

Results: Between 2005 and 2019, 1-year all-cause mortality declined from 22.8% (95% CI: 22.4% to 23.2%) to 14.2% (95% CI: 13.7% to 14.7%), an absolute reduction of 8.6 percentage points. Approximately 68.2% of this decline was explained by measured factors. The greatest contributor was increased use of evidence-based pharmacological therapies—including statins, beta blockers, and ACE inhibitors—accounting for 27.8% of the reduction (95% CI: 26.0% to 29.5%). Increased use of percutaneous coronary intervention (PCI) contributed 17.6% (95% CI: 16.3% to 19.0%). Mortality improvements were observed in both STEMI and NSTEMI populations. However, older adults and women continued to experience higher mortality rates throughout the study period.

Conclusion: One-year survival after AMI improved substantially over 15 years in England and Wales, largely due to advances in pharmacotherapy and PCI.

The Energy Distance Between Covariate Distributions and Assessments of Generalization

Wendy Chan, University of Pennsylvania

Randomized controlled trials are known as the gold standard for strengthening the internal validity, or causality, of the relationships between interventions and outcomes. As researchers and policymakers have moved beyond questions of what works to what works, for whom and under what conditions, questions of causality have naturally evolved to include discussions about generalization. For over a decade, statisticians have developed methods to improve generalizations, particularly from non-random study samples. An important strand of this work focuses on study design and ways to assess the generalizability of study samples. In practice, four measures have been proposed to assess the generalizability or similarity between individuals in a sample and those in a target population of inference. These statistics provide single number summaries of the generalizability of the study, but their utility is limited, particularly in small studies where generalization may be difficult to justify. The current study examines the relationship between these four statistics and the energy distance in covariate distributions between the sample and population. Using both a simulation study and an empirical example, we investigate values of the energy distance that are associated with certain values of the four statistics with the goal of providing researchers with a way to identify sources of variation that lead to weak and strong assessments of generalization.

The Accidental Product Manager: How my PhD in Biostatistics led me to a thriving career in healthcare technology

Jacquelyn Neal, WebMD Ignite

Dr. Jacquelyn E Neal didn't set out to work in the healthcare industry as a product manager. After completing her PhD in biostatistics, where her research focused on predictive modeling of longitudinal data, her plan was to work in clinical trials or for a government agency. When life intervened with the unexpected loss of a family member, she took a pause to look after her mental health and began working an entry-level analytics job, which she continued through the end of her PhD. Through a series of jobs she held while finishing her PhD and an unexpected company sale, she landed as a product manager for longitudinal analytics of healthcare marketing campaigns at WebMD Ignite. During this session, she will discuss her unique career journey from highly-specialized statistics research to her current role as a data product leader and what these roles entail. Dr. Neal will focus on how skills from her biostatistics training are uniquely valuable in product management, how she leverages her subject-area knowledge from research in a business role, and how product organizations are a wonderful environment for statisticians and data scientists to build thriving careers. For women wondering how they can leverage their statistics and data science training outside of a traditional research job and build a career of impact that serves their needs, product management can be an ideal fit.



Session Info

S124

CAREER OR PROFESSIONAL DEVELOPMENT SESSION

October 14th, 20:30 - 21:00 UTC

Strategies for Thriving Amid Stress, Conflict, and Uncertainty at Work

USE SCHED TO ACCESS THE ZOOM ROOM

Workplace difficulties are an inevitable part of professional life, whether they stem from high-pressure environments, challenging colleagues, or complex relationships with management. The panelists will discuss practical approaches to managing conflict, fostering resilience, and promoting healthy communication. Topics include recognizing signs of a toxic workplace, setting boundaries, leveraging emotional intelligence, the skill of “managing up” — effectively communicating with and influencing your manager, and knowing when to seek external support or exit. Drawing from real-world examples, the panel will provide actionable insights for employees and leaders to navigate and transform challenging work environments effectively.

ORGANIZER: Umut Ozbek, Eli Lilly and Company**CHAIR:** Deliverance Bougie, U.S. Census Bureau

Speaker Bios

S124



NANCY FLOURNOY

University of Missouri

Nancy Flournoy is Curators Distinguished Professor Emerita and former Department Chair at the University of Missouri and American University. She was the first female Statistics Program Director at NSF. Her Fellowships include ASA, IMS, AAAS and WAAS, and she is the recipient of ASA's Founders Award, COPSS's Elizabeth Scott and F.N. David Awards, a NISS's Distinguished Service Award and NSF's Outstanding Performance Award. Professor Flournoy was first to statistically support the hypothesis of graft-versus-host disease in man, first to implement a dose-finding design framed as an inverse estimation problem, first to implement a Bayesian design, first to propose an adaptive design in the joint context of toxicity and efficacy, first to propose interval designs for dose-finding, and first to recommend isotonic regression for estimating dose-response functions following a dose-finding study. Professor Flournoy was at the forefront of renewed methodological interest in Up-and-Down Design



GOGCE CRYNEN

University of Florida

Dr. Gögçe (pronounced "gohg-che") Ceren Crynen is both the Director of the Bioinformatics Core and the Shared Services Director, overseeing all research cores at The Wertheim UF Scripps Institute in Jupiter, Florida. With advanced training in biological and quantitative genetics from METU in Ankara, Türkiye and the University of Florida, Dr. Crynen has over 19 years of experience in bioinformatics, biostatistics, and the analysis of complex biological data. Much of her work has been centered on building strong collaborations across scientific disciplines, and she is especially interested in fostering teamwork, supporting career growth, and building resilience within research teams.



DELIVERANCE BOUGIE

U.S. Census Bureau

Deliverance Bougie is a Survey Statistician at the U.S. Census Bureau on the Demographic Frame Team. She holds an MS in Measurement and Data Science from the University of Mannheim and served as a Senior Statistician at the Indiana Department of Local Government Finance. Before her career in statistics, she worked in social services, a background that grounds her commitment to ensuring marginalized populations are represented in data.

Outside of work, Deliverance is passionate about cultivating balance through her hobbies. She finds peace in gardening, hiking, and backpacking, and has recently developed a full-blown love for weaving. With six looms (and counting), she creates table runners, hand towels, mug rugs, scarves, and decorative bands—projects that bring both focus and joy. These creative and outdoor pursuits help her decompress from work and recharge for challenges ahead.

Disclaimer: Deliverance is appearing in a personal capacity; her views do not represent the Census Bureau.



Session Info

S125

TECHNICAL STATISTICAL SESSION

October 14th, 21:00 - 22:00 UTC

Harnessing Environmental and Ecological Data: Tools for Navigating Change and Sustainability

USE SCHED TO ACCESS THE ZOOM ROOM

In an era of accelerating environmental change and growing demands for sustainability, the integration of environmental and ecological data has never been more critical. This session, *Harnessing Environmental and Ecological Data: Tools for Navigating Change and Sustainability*, brings together leading researchers who are at the forefront of statistical innovation and applied environmental science. Dr. Marian Scott (University of Glasgow), Dr. Meredith Franklin (University of Toronto), and Dr. Marie Auger-Méthé (University of British Columbia) will share their perspectives on the development and application of cutting-edge statistical and computational tools that illuminate complex ecological processes and inform policy. Together, these talks will offer a compelling vision of how environmental and ecological data science can help navigate the path toward a more sustainable and resilient future.

ORGANIZER: Joanna Mills Flemming, Dalhousie University**CHAIR:** Joanna Mills Flemming, Dalhousie University**SPONSOR:** CANSSI

Speaker Bios

S125

**PROF. MARIAN SCOTT**

University of Glasgow

Dr. Marian Scott is a renowned statistician and Professor of Environmental Statistics at the University of Glasgow. She completed both her BSc and PhD in Statistics there, with a doctoral thesis focused on sources of error in radiocarbon dating. Her academic career has been distinguished by pioneering work at the interface of statistics and environmental science, encompassing air and water quality monitoring, spatio-temporal modeling, pollutant dispersion, radiocarbon calibration, and uncertainty analysis. Professor Scott has received numerous honors, including being appointed Officer of the British Empire (OBE) in 2009. In 2019, she was awarded the Royal Statistical Society's Barnett Award for her innovative application of statistics to environmental challenges. Beyond her research, she plays a significant role in scientific leadership and policy, serving as Vice-President of the Royal Society of Edinburgh, and member of the Scottish Science Advisory Council.

**DR. MARIE AUGER-MÉTHÉ**

University of British Columbia

Dr. Marie Auger-Méthé is an Associate Professor and Canada Research Chair in Statistical Ecology at UBC's Institute for the Oceans and Fisheries and Department of Statistics. She specializes in developing statistical tools—especially state-space and hidden-Markov models—that infer behavioral and population processes from animal-movement data. Passionate about marine and polar species, her work decodes habitat use and migration strategies of narwhals, polar bears, whales, terns, and seals—even integrating Indigenous Knowledge into habitat-selection models. A recent inductee into the Royal Society of Canada's College of New Scholars (2024), she leads interdisciplinary teams to model critical habitats and human impacts on ecosystems. By bridging ecology, statistics, and conservation, Marie's research informs management of marine and polar environments, supports policy decisions, and advances our understanding of animal behavior in a changing world.

**DR. MEREDITH FRANKLIN**

University of Toronto

Meredith Franklin is an Associate Professor in the Department of Statistical Sciences and the School of the Environment at the University of Toronto. With a PhD in Statistics and Environmental Health from Harvard, her work focuses on applying advanced spatio-temporal and machine-learning methods to environmental health research. She investigates the health impacts of exposures such as air pollution, wildfire smoke, and non-tailpipe emissions using satellite data, community monitoring, and novel modeling techniques. She teaches data science and statistical theory, leads interdisciplinary research projects, and mentors students at the Data Sciences Institute. Franklin also serves on editorial boards and collaborates on projects funded by agencies like NASA and NIH. Her work bridges statistical innovation and public health, driving policy on air quality, climate-health interactions, and environmental justice.





Digital Water- Tools to Conserve and Value Water

Marian Scott, University of Glasgow

Increasingly more and more data are being generated on environmental systems, from both new sensors and also earth observation missions. Lakes, reservoirs and rivers are the focus of much monitoring and modelling. There are a number of integrative models, offering holistic views of the "system", in this case the hydrological cycle. However, in this context, digital twins are being developed jointly for water quantity and quality. A digital twin offers a dynamic system view, informed/learning from data, a digital replica of the environmental system with a bi-directional flow of intelligence from real-virtual and virtual- real. Their main power lies in their predictive capacity, and the testing of different future scenarios. I will present a framework for catchment (river basin) modelling of water quality and quantity, and discuss the different statistical tools needed.

Modelling the Movement & Space Use of Marine Species to Support Their Conservation

Marie Auger-Methe, University of British Columbia

Quantifying the distribution of animals and understanding their movement behaviour is fundamental to their conservation. As such, ecologists increasingly collect movement data. However, characterising the distribution and behaviour of marine species is hindered by many formidable challenges. For example, marine species spend most of their life in areas difficult for us to reach (e.g., ocean depths) and many positioning systems (e.g., GPS) are not well suited to the marine environment. Using species such as Arctic terns and narwhals, I will demonstrate how advanced statistical methods can improve our understanding of their ecology and inform management and conservation.

Spatiotemporal Data Science for Climate Change and Public Health: Integrating Satellites, Climate Models, and Physics-Informed Deep Learning

Meredith Franklin, University of Toronto

Many current challenges in environmental change research demand scalable approaches capable of integrating vast and heterogeneous datasets, such as satellite observations and climate model outputs. Deep learning techniques optimized for spatiotemporal data help to uncover patterns and predict trends in critical environmental phenomena with direct public health implications, such as air quality from wildfires and dust storms, and heat exposure. We design and implement advanced neural architectures, including physics-informed models, that embed domain knowledge into learning algorithms, improving both accuracy and interpretability. These systems offer robust, high-resolution estimates, enabling more precise environmental exposure assessments with which we can examine health effects.



Session Info

S126

CAREER OR PROFESSIONAL DEVELOPMENT SESSION

October 14th, 22:00 - 23:00 UTC

Thriving in Multidisciplinary Spaces: Lessons in Leadership, Confidence, and Self-Advocacy for Women in Data Science

USE SCHED TO ACCESS THE ZOOM ROOM

This panel will provide tips for women to thrive in multidisciplinary professional environments. Inspired by the book “Nice Girls Don’t Get the Corner Office” by Lois P. Frankel, our discussion will focus on how women can unlearn internalized behaviors that hold them back, claim space with confidence, and redefine leadership on their terms.

The panel will gather leaders from academia and industry, representing diverse roles and cultural backgrounds, to reflect on power dynamics, communication styles, and self-advocacy in environments that may not have been built with minorities in mind. We will address how to build resilience and influence in settings where norms, expectations, and personalities collide—and how to mentor and support others along the way.

This panel, aligned with the conference theme, will offer nuanced strategies for personal and professional success for women and gender-diverse individuals in data science and targets students, post-docs, and early-career researchers.

Our potential panellists will reflect a diversity of disciplines, ethnicities, and leadership styles. Key Topics:

What lessons can we take from Nice Girls Don’t Get the Corner Office in our careers?

What are effective strategies for navigating multidisciplinary teams and cross-cultural dynamics?

How can we better support and mentor others in interdisciplinary and nontraditional paths?

How do we define leadership in a field that intersects statistics, medicine, computing, and policy?

ORGANIZER: Ghazal Azarfar, University Health Network

CHAIR: IDWSDS Committee Member



Speaker Bios

S126



DR. GHAZAL AZARFAR

University Health Network

Dr. Ghazal Azarfar is passionate about harnessing innovative multimodal AI solutions to transform healthcare. She has joined University Health Network as a part of Transplant AI Initiative. She leverages the team's multidisciplinary expertise in clinical medicine, computer science, and biostatistics to advance the field of transplant medicine—particularly in understanding and managing pre- and post-transplant complications. Her current research focuses on the responsible deployment of a multi-agent AI model that simulates transplant committee deliberations, aiming to generate objective consensus for equitable waitlisting decisions.



DR. MEILING HE

Rockwell Automation

Meiling He holds Ph.D. in electrical engineering, focused on machine learning. She is currently the Senior Data Science Manager at Rockwell Automation, where she leads the advanced data science team in developing AI, GenAI, and automation solutions that drive innovation across sales, engineering, manufacturing, and customer engagement. She has led flagship initiatives including PharmaSuite® Copilot, which dramatically accelerates recipe creation; a GenAI Sales AI Assistant integrated with Salesforce to enhance account planning and sales execution; and AI-powered predictive maintenance for MagneMotion systems that improve operational reliability. Her work also advances foundational AI capabilities in causal AI, graph data science, and machine learning engineering to enable scalable, cross-functional innovation. She also holds an Executive MBA from Northwestern University's Kellogg School of Management and a Graduate Certificate in Product Management.



DR. ELAHEH KHEIRANDISH

globalfoundries

Elaheh Kheirandish is a materials scientist and engineer specializing in advanced semiconductors, characterization, and metrology. She earned her Ph.D. in Electrical Engineering from the University of Wisconsin–Milwaukee and has worked in both academia and industry, including roles at Meta Reality Labs and GlobalFoundries. She is passionate about creating robust electronic materials and devices and mentoring future female innovators.



Session Info

S127

CLOSING SESSION

October 15th, 00:00 - 00:30 UTC

Closing

USE SCHED TO ACCESS THE ZOOM ROOM

In our closing session we will celebrate the 24 hours we spent together in this conference! We will hear from you about sessions that were impactful and engaging, your “ah ha” moments, and lessons learned from our speakers. We will conclude by gathering your feedback and ideas for next year!



CYNTHIA BLAND
Program Co-Chair



JESSICA KOHLSCHMIDT
Program Co-Chair



Session Info

S201

TECHNICAL STATISTICAL SESSION

October 14th, 9:00 - 10:00 UTC

Statistics without Boundaries: Uniting Classical Methods with Artificial Intelligence, Machine Learning and Data Science

USE SCHED TO ACCESS THE ZOOM ROOM

The landscape of statistical analysis has undergone a revolutionary transformation over the past decade, marked by the convergence of traditional statistical methodologies with cutting-edge artificial intelligence and data science techniques.

This session addresses the critical need to bridge the perceived divide between classical statistics and modern computational approaches, demonstrating that these methodologies are not competing paradigms but complementary tools that, when integrated effectively, create a more robust and comprehensive analytical framework. The statistical community increasingly recognizes that the dichotomy between "traditional" and "modern" methods is artificial and counterproductive. Classical statistical techniques—including hypothesis testing, experimental design, regression analysis, and inferential statistics—provide the theoretical foundation and interpretive rigor essential for sound scientific inquiry.

Meanwhile, machine learning algorithms, deep learning architectures, and large-scale data analytics offer unprecedented capabilities for pattern recognition, prediction, and handling complex, high-dimensional datasets. The integration of these approaches represents the next evolutionary step in statistical practice.

ORGANIZER: Wendy Plata, Silver Quality Consulting

CHAIR: Wendy Plata, Silver Quality Consulting



Speaker Bios

S201



DR. ANDREA GARCÍA-ANGULO

Faculty of Natural Sciences and Mathematics, ESPOL

<https://i-research.espol.edu.ec/investigador/59325688400>

Andrea García-Angulo is a Professor of Statistics at the Faculty of Natural Sciences and Mathematics at ESPOL (Escuela Superior Politécnica del Litoral), Ecuador. She holds a Ph.D. in Business Economics with a focus on Statistics and Operations Research from KU Leuven, Belgium. Her research interests include statistical modelling, model selection and post-selection inference, where she has contributed to the literature with new methodological proposals. She has also contributed to interdisciplinary projects and publications in the fields of education, biosciences, and environmental sciences, where she has been responsible for developing study designs and conducting formal statistical analyses. As a statistician, she frequently encounters the realities of messy, incomplete, or inconsistent data. Over time, she has developed strong experience in tackling these challenges through careful data cleaning, preprocessing, and imputation techniques, ensuring that the information is reliable.



DR. KATHERINE MORALES-QUINCA

Khome - La Fabrique by Crédit Agricole

<https://katytamq5.wixsite.com/katherine-morales19>

Katherine Morales is a Data Scientist with a Ph.D. in applied mathematics, specialized in generative models and probabilistic methods. She has worked across real estate, fintech, and climate risk, delivering models from research to production in valuation, risk scoring, and anomaly detection. She has also worked in the medical field with the European Cardiovascular Research Group, Strasbourg-France. She has also collaborated with the Women in Data Science community, contributing to knowledge-sharing and mentoring initiatives.



DR. BRENDA COBEÑA

HEC Montréal

<https://www.brendacobena.com/>

Brenda Cobeña is currently a Postdoctoral Research Fellow at HEC Montréal and an Affiliated Researcher at CIRRELT, one of Canada's leading research centers in logistics and transportation. Her work bridges optimization and machine learning, focusing on real-time decision-making, supply chains, and urban mobility. She recently earned her PhD in Industrial Engineering from Concordia University in Montréal, where her research addressed hub location problems with elastic demand and the design of urban mobility hubs. Before pursuing her doctorate, Brenda gained significant experience in both academia and industry in Ecuador, serving as a Market Intelligence Manager in the banking sector, a senior analyst in the food industry, and a university professor in logistics and transportation. She is the recipient of several distinctions, including the Amazon SCOT/INFORMS Scholarship and IVADO's Scientist-in-Residence program. Passionate about diversity in STEM, promoting women's visibility.



MISS WENDY PLATA

Silver Quality Consulting

<https://www.linkedin.com/in/wendyplata/>

Wendy Plata is a Computer Engineer and Master in Productivity and Quality Management, educated at Escuela Superior Politécnica del Litoral (ESPOL). She brings over 20 years of experience in processes, statistics, and education, leading initiatives that combine technology, quality, and organizational well-being. She is the founder and CEO of Silver Quality Consulting, a firm specialized in quality, productivity, and organizational well-being across Latin America. Currently, she represents the Emotional Paycheck Institute of Canada in Ecuador, the world's first institute 100% dedicated to organizational well-being. Through this partnership, she promotes the use of the TARRIX tool, which enables measuring emotional salary and designing well-being strategies with artificial intelligence support. She has worked in key institutions such as the National Institute of Statistics and Census (INEC) and the National Customs Service of Ecuador (SENAE) contributing to public policies.





"Garbage In, Garbage Out": Same Challenge, New Dimensions

Andrea García-Angulo, Faculty of Natural Sciences and Mathematics, ESPOL

The quality of any analysis—whether traditional statistical modeling or modern artificial intelligence—depends fundamentally on the quality of the input data. The old principle of "garbage in, garbage out" remains as relevant as ever, but today it takes on new dimensions. Messy, incomplete, or inconsistent datasets are the norm rather than the exception, and preparing them requires careful cleaning, preprocessing, and imputation guided by the research questions at hand. While advances in AI and machine learning provide new tools to assist in these processes and to uncover complex patterns, they cannot replace the critical judgment of a statistician. Deciding what to keep, what to transform, and what to question remains an intellectual task that links data back to context and purpose. This talk reflects on the enduring challenge of data preparation, explores how AI can complement rather than replace statistical thinking, and highlights why rigorous preprocessing is key to generating trustworthy insights.

A Probabilistic Semi-Supervised Approach With Triplet Markov Chains

Katherine Morales-Quinga, Khome – La Fabrique by Crédit Agricole

Triplet Markov chains are general generative models for sequential data which consider three kinds of random variables: (noisy) observations, their associated discrete labels and latent variables which aim at strengthening the distribution of the observations and their associated labels. However, in practice, we do not have at our disposal all the labels associated to the observations to estimate the parameters of such models. In this paper, we propose a general framework based on a variational Bayesian inference to train parameterized triplet Markov chain models in a semi-supervised context. The generality of our approach enables us to derive semi-supervised algorithms for a variety of generative models for sequential Bayesian classification.

Bridging Optimization and Machine Learning: Challenges and Opportunities in Agri-Food Hub Networks

Brenda Cobeña, HEC Montréal

The agri-food sector faces increasing pressure to design resilient and sustainable distribution networks that can adapt to uncertainty in demand, supply disruptions, and environmental constraints. Classical optimization approaches—such as hub location models—offer powerful tools to determine efficient network structures, but they often rely on static assumptions and limited data. On the other hand, machine learning (ML) offers predictive power and adaptability; however, its integration into decision-making frameworks for large-scale logistics remains an open challenge.

This presentation examines research gaps and emerging opportunities at the intersection of optimization and ML within agri-food supply chains. Key areas of inquiry include the potential for predictive models to improve hub location decisions under uncertainty, the trade-offs between interpretability and scalability

when integrating ML with mixed-integer optimization, and the contributions of these approaches to food security and sustainability in rapidly changing environments.

This talk focuses on conceptual frameworks, key challenges, and potential research directions. Integrating optimization and machine learning (ML) is expected to enhance the capacity of agri-food systems to operate effectively in complex and dynamic environments.

From Data to Insights: Leveraging Generative Artificial Intelligence for Analytics

Wendy Plata, Silver Quality Consulting

This contribution explores the transformative potential of generative artificial intelligence in revolutionizing data analysis workflows. As organizations increasingly rely on data-driven decision making, the integration of AI-powered tools presents unprecedented opportunities to enhance analytical capabilities and streamline complex processes.

A significant focus is placed on advanced prompt engineering techniques, equipping attendees with the skills to craft effective instructions that maximize AI performance. The session delves into specialized prompts designed for critical data analysis tasks, including data cleaning and preparation methodologies, comprehensive statistical analysis encompassing both descriptive and inferential approaches, and automated code generation in Python and R programming languages.

It concludes with critical guidance on interpreting AI-generated results, addressing accuracy assessment, validation techniques, and best practices for integrating AI insights into decision-making processes. Participants will leave with practical skills and strategic knowledge to effectively leverage generative AI technologies in their analytical work.



Session Info

S202

CAREER OR PROFESSIONAL DEVELOPMENT SESSION

October 14th, 10:00 - 11:00 UTC

Artificial Intelligence: Beyond Bias

USE SCHED TO ACCESS THE ZOOM ROOM

AI is no longer futuristic—it's woven into our lives, shaping what we watch, buy, and how decisions are made in healthcare, finance, and justice. The rise of agentic AI, capable of autonomous perception, reasoning, and action, promises transformation—but also brings challenges, especially AI bias. Bias in AI can harm real lives: from hiring algorithms excluding qualified women to facial recognition misidentifying people of color, or healthcare tools failing certain populations. "AI: Beyond Bias" unites three experts to share actionable strategies for a fairer AI future. The first will urge us to "break away from the familiar," challenging entrenched assumptions and expanding perspectives to tackle the roots of bias. The second will stress the need for authentic data—capturing the complexity of real human experience—over token diversity, to build AI that truly understands. The third will highlight methodological and conceptual diversity, blending technical, social, and ethical expertise to address systemic causes of bias. This isn't just theory—it's a call to action for technologists, policymakers, and citizens to shape an equitable AI future. The session brings together three leading experts to explore this critical challenge. They will offer actionable insights and fresh perspectives on how we can build a more equitable and sustainable AI ecosystem. It's a call to action for anyone who cares about the future of technology and its impact on society.

ORGANIZER: Suhwon Lee, University of Missouri**CHAIR:** Mihye Kim, Chungbuk National University

Speaker Bios

S202



DR. HEISOOK LEE

GISTeR

<https://www.gister.re.kr/front/user/main.do>

Dr. Heisook Lee is a leading voice for inclusive R&D, championing the integration of sex, gender, and intersectional analysis in science and technology. As President of the Korea Center for Gendered Innovations for Science & Technology Research (GISTeR), she has driven the Gender Summit Asia Pacific and the Global Gender Summit for SDGs in Seoul, advancing opportunities for women and girls in STEM. Professor Emeritus of Mathematics at Ewha Womans University, she taught and researched commutative algebra for over three decades. As founding president of WISet (2011–2016), she created pioneering national programs, including the Returnee Program for women re-entering STEM and a large-scale mentoring network for female students. Her current research explores AI for inclusive growth and gendered innovation policy. Dr. Lee's leadership has extended to service on the Presidential Advisory Council on Science & Technology, the National Science & Technology Commission, and as president of KOFWST.



DR. JEONG-HAN KANG

Department of Sociology, Yonsei University

<https://sites.google.com/view/yoda-lab>

Jeong-han Kang received Ph.D. in sociology from University of Chicago, was a post-doc researcher at Cornell University, and currently is Professor of Sociology at Yonsei University, Seoul. He is also the director of CLIO Institute for Social Development Studies in the university and serving as the chief editor of Survey Research of Korean Association for Survey Research. His recent research interests include digital transformation, narrative-driven data science, and science of science with a gender perspective and he is leading YODA(Yonsei Online Data Analysis) Lab in the department (<https://sites.google.com/view/yoda-lab>). His studies have been published in Journal of Informetrics, Social Policy & Administration, Proceedings of the Web Conference (WWW), Journal of Mathematical Sociology, Sociological Methods and Research, and Administrative Science Quarterly.



DR. KEON MYUNG LEE

School of Computer Science, Chungbuk National University

<https://www.lucymcgowan.com/>

Professor Keon Myung Lee received his BS, MS, and Ph.D. degrees in computer science from KAIST(Korea Institute of Science and Technology), Korea and was a Post-doc fellow in INSA de Lyon, France. He was a visiting professor in University of Colorado at Denver and a visiting scholar in Indiana University, USA. After having an industrial career at Silicon Valley, USA, he joined Dept. of Computer Science, Chungbuk National University, Korea in 1995. Now he is a professor. He has served as the Editor-in-Chief of International Journal of Fuzzy Logic and Intelligent Systems. His principal research interests are in artificial intelligence, machine learning, social impacts of AI, and intelligent service systems.



DR. SANG-WOOK YI

Hanyang University

Professor Sang-wook Yi is a tenured professor at Hanyang University, specializing in the philosophy of science and technology. He holds physics degrees from Seoul National University and a PhD in philosophy of science from LSE, where his thesis won the Robert McKenzie Prize. He founded the Center for Ethics, Law and Policy of Science and Technology. His research spans AI ethics, synthetic biology, and posthumanism. He has chaired several South Korean government Technology Assessment committees and served as president of the Korean Society for Philosophy of Science. He is a member of UNESCO's COMEST, where he has been a vice-chair and rapporteur. He also served as rapporteur for the Ad-hoc experts group which drafted UNESCO Recommendation on AI Ethics in 2021.



Session Info

S203

TECHNICAL STATISTICAL SESSION

October 14th, 11:00 - 12:00 UTC

Statistics and Decision Making in the UK

USE SCHED TO ACCESS THE ZOOM ROOM

In an era defined by big data and large language models, the ability to translate statistical insights into evidence-based decisions is more critical than ever. This session explores the intersection of statistical methodologies and practical decision-making in the public health and environmental domains. Presentations will examine the role of statistics in guiding strategic choices under uncertainty, and emphasis will be placed on the challenges of balancing statistical rigor with real world constraints to turn data into actionable insight.

ORGANIZER: Esther Jones, Biomathematics and Statistics Scotland (BioSS)

CHAIR: Helen Kettle, Biomathematics & Statistics Scotland (BioSS)



Speaker Bios

S203

**MANIRA AHMAD**

Public Health Scotland

Manira's career spans private and public sectors, with a focus on systems, emotional architecture, and lived experience. After roles in the financial industry on global workforce planning and offshore trading, she joined the public sector in 2015 to lead the deployment of Local Intelligence across Health and Social Care in Scotland, co-designing linked data to improve decision-making. She contributed to establishing Public Health Scotland and was appointed Chief Officer in 2021. Manira leads strategic, multi-disciplinary programmes, driving collaboration and innovation to support communities. In 2024 she joined the Board of Directors of City of Glasgow College and became an Associate Member of the Centre for Health Policy at the University of Strathclyde. She also serves on boards of local charities, supporting inclusion, physical activity, and reducing health inequalities.

**PROF. SALLIE BAILEY**

Natural England

Sallie Bailey is Chief Scientist at Natural England, the UK government's adviser on the natural environment, leading the Chief Scientist's Directorate to embed science and expertise in decisions, partnerships, and community action for nature recovery. She was formerly Deputy Chief Science Advisor for the Scottish government on environment, resources, and agriculture, developing the First Minister's Environmental Council to guide responses to the climate and ecological crises. With international experience across the EU, UK, and Scotland in evidence-based policy, regulation, and biodiversity management, Sallie has also held senior roles in the state forestry sector. She is a Fellow of the Institute of Chartered Foresters and the British Ecological Society, and Honorary Professor at the University of Stirling. After an MSc in GIS and remote sensing and a PhD in woodland ecology at Nottingham, she completed postdoctoral work at Stanford's Centre for Conservation Biology on biodiversity and natural capital.





A Journey Without LOGOs and EGOs

Manira Ahmad, Public Health Scotland

TBD

Abstract

Sallie Bailey, Natural England

As biodiversity continues to decline globally, effective measurement, monitoring and analysis are critical for understanding the drivers and consequences of environmental change. These tools enable us to identify priority areas for conservation, assess the success of interventions, and guide evidence-based decision making to improve outcomes for nature. Robust analysis and treatment of data, and application of statistics, also play a vital role in tracking progress against international and statutory commitments, supporting evaluation frameworks, and deepening our understanding of ecological systems.

With new rapidly emerging technologies, autonomous monitoring and artificial intelligence creating opportunities to improve how we monitor the natural world, we can utilise this new technology for more effective conservation outcomes.

By describing case studies I'll provide valuable insights from data collection and management through to analytical methodologies which can be harnessed to drive innovation, inform best practice, and make a difference for nature.



Session Info

S204

TECHNICAL STATISTICAL SESSION

October 14th, 12:00 - 12:45 UTC

Rising Stars: Celebrating Student-Driven Research

USE SCHED TO ACCESS THE ZOOM ROOM

In this session, sponsored by the Caucus for Women in Statistics and Data Science, we celebrate our newest statisticians by highlighting a group of students just beginning their research journeys. The speakers will discuss their applications of statistical methods to the fields of social influence, disease transmission, dental care, and mental health. First, first-year Ph.D. student Miriam Vandewater will compare long-term fitness outcomes of competing populations with different levels of sociality using a stochastic agent-based simulation. Next, third-year undergraduate student Lydia Owens will discuss her journey to biostatistics and how she blended epidemiology, programming, and justice-oriented inquiry to develop a model for COVID-19 transmission within prisons. High school junior Carrie LaBarca will follow with an analysis of dental clinic visit trends in the United States. Finally, high school senior Danny Hoberman will close the session by exploring the link between access to healthy food and depression prevalence in southeastern Pennsylvania. All four speakers will blend exciting questions with rigorous statistics, making the session interesting to a general audience. They demonstrate how statisticians can thrive in their environment at any career stage!

ORGANIZER: Ashley Mullan, Vanderbilt University**CHAIR:** Hannah Klinger, Vanderbilt University

Speaker Bios

S204



MS. MIRIAM VANDEWATER

North Carolina State University

Miriam Van de Water is a Ph.D. student in the Department of Biomathematics at North Carolina State University. She graduated in 2025 from the University of Scranton with a double major in Mathematical Sciences and Neuroscience and a minor in Philosophy. Her current research interests include the mathematical modeling of disease spread, collective behavior in dynamical systems, and machine learning. She also enjoys exploring the history and philosophy of mathematics, studying French, and trying out new recipes.



MS. LYDIA OWENS

University of North Carolina at Chapel Hill

Lydia Owens is a Morehead-Cain Scholar, a Barry Goldwater Scholar, and a double major in Biostatistics and Mathematics at the University of North Carolina at Chapel Hill. She is a Student Researcher at the Duke School of Medicine Bellwether Collaborative for Health Justice, where she uses statistical modeling methods to help understand and improve the health of incarcerated populations. She has also served as a Communications Committee Member for the Caucus for Women in Statistics and is a member of the planning committee for the ENAR Fostering Diversity in Biostatistics Workshop. Owens has researched COVID-19 transmission within prisons, used machine learning methods to predict lack of PrEP knowledge among incarcerated individuals, and used Bayesian models to model lung cancer mortality within the United Kingdom. Her goal is to obtain a PhD in Biostatistics and use her statistical knowledge to increase health equity for underserved populations and solve global health problems.



MS. CARRIE LABARCA

St. Anthony's High School

Carrie LaBarca is a rising junior at Saint Anthony's High School. She aspires to go into Dental health, and she has focused much of her student research for the AP Scholars program in this area. She has explored the impact of periodontal disease on the impact of rheumatoid arthritis, thought about the use of nanobot technology on the reduction of periodontal disease, and identified factors that affect patient compliance with post-orthodontic retainer use in Long Island high school students. Outside of school, Carrie is an avid musician and plays multiple instruments including the violin, harp, and piano. She volunteers in the music ministry at her community church and serves as a cantor. Additionally, she is on the Badminton team and loves to sail and rockclimb.



MR. DANIEL HOBERMAN

St. Anthony's High School

Daniel Hoberman is a student at St. Anthony's High School in South Huntington, NY who has a strong passion and interest in public health research and data science. His current research explores how access and proximity to healthy foods relates to mental health outcomes in Pennsylvania. Daniel is currently working under the mentorship of Ashley Mullan, a Ph.D. student at Vanderbilt University. Daniel is planning on entering his research into local and national level competitions and symposiums. Outside of research, Daniel is the president of his school's National Honor Society and Mock Trial Team, is the co-founder of a non-profit organization dedicated to helping older adults and underserved communities access and navigate technology, and also works as a math and science tutor.



Examining Collective Decision-Making Over Time*Miriam Van de Water, North Carolina State University*

Collective decision-making (CDM) and social influence are widely studied mechanisms in biological and artificial systems, often assumed to confer adaptive advantages. In this project, we investigate the long-term fitness outcomes of competing populations exhibiting different levels of sociality and CDM through a stochastic agent-based simulation. The model extends a probabilistic decision framework by incorporating repeated decision cycles and fitness consequences (modeled as reproduction and death probabilities) to evaluate population success over time. We hypothesized that populations with higher levels of sociality and CDM would outperform others. However, our results reveal a nontrivial interaction: CDM improves performance in nonsocial contexts, but the addition of social influence can degrade outcomes. Conversely, sociality can be advantageous in the absence of CDM. We also explore edge cases in which the influence of poor decision-makers is minimized (e.g., when the parameter $k=0$), yielding further deviations from expected behavior. These findings underscore the complex role of social structure in decentralized decision systems and highlight the utility of simulation-based approaches for probing emergent group dynamics. The work contributes to ongoing research in statistical modeling of collective behavior, with potential applications in behavioral ecology, swarm robotics, and collective intelligence.

Integrating Advocacy and Algorithms: My Path from Legal Advocacy to Biostatistical Research*Lydia Owens, University of North Carolina at Chapel Hill*

At sixteen years old, I began my research journey with a question: How can we use data and modeling to protect the health and dignity of incarcerated people? Growing up, I grew my budding identity as an advocate through mock trial and restorative justice programs, always being drawn to the intersections of advocacy and systems thinking. Alongside this, I nurtured a deep love for math and problem solving—two differently seeming interests that found a home in computational public health. With the support of teachers and research mentors, I began developing an agent-based model to simulate COVID-19 transmission within prisons, blending epidemiology, coding, and justice-oriented inquiry. Through this work, I taught myself scientific programming and computational modeling, but also grappled with the uncertainties, setbacks and ethical questions that come with real-world research. In this talk, I'll reflect on that journey and share how combining law, science and math reshaped my path, including what I've learned about using biostatistics to model change.

Exploring Dental Clinic Visit Trends in the United States*Carrie LaBarca, St. Anthony's High School*

Studies have shown that oral health can influence systemic diseases such as diabetes, heart disease, rheumatoid arthritis, and stroke. Organizations such as the National Oral Health Surveillance System (NOHSS) monitor oral disease, healthcare delivery systems, and status of fluoride in community water. This study will examine the relationship between socioeconomic factors, such as age, income, and education, and the prevalence of dental clinic visits among the U.S. population. Data from even

years between 2012 and 2020 collected by the Centers for Disease Control and Prevention (CDC) from the NOHSS will be gathered and analyzed using the R language. These estimates, reported as median prevalence from the Behavioral Risk Factor Surveillance System (BRFSS) public use datasets, will be compared to identify patterns and correlations. Understanding the role of socioeconomic factors can help inform policies aimed at improving oral health outcomes nationwide.

Mapping the Mind: The Link Between Proximity to Healthy Foods and Depression in Southeastern Pennsylvania*Daniel Hoberman, St. Anthony's High School*

Access to healthy foods is a well-studied factor in physical health outcomes, but its connection to mental health remains less understood. This project examines how the availability of healthy food retailers relates to depression prevalence across Southeastern Pennsylvania, aiming to uncover patterns that could inform public health policy. A two-phase design model was implemented in which Haversine (straight-line) distances to healthy food retailers were measured from the population center of each census tract in the chosen region, and driving distances were also calculated for select randomly queried tracts. Using data available from the CDC's PLACES, depression prevalence was also compiled and plotted for the selected region of PA. Poisson regression models were then applied to assess how different food access measures, as well as other socioeconomic factors, influence the observed association between food access and depression. In this talk, I will walk through how I collected and integrated these datasets, why measurement choice matters, and what my results reveal about the complex relationship between food environments and mental health.



Session Info

S205

TECHNICAL STATISTICAL SESSION

October 14th, 12:30 - 12:45 UTC

Speed Session 1

USE SCHED TO ACCESS THE ZOOM ROOM

This session is a short session featuring three speed talks. We hope that you will contact the speakers if you are interested in discussing their work further.

ORGANIZER: Cynthia Bland, CWS

CHAIR: Jessica Kohlschmidt, CWS



Speaker Bios

S205



EMILY GENTLES
RTI

<https://www.linkedin.com/in/emily-gentles-336959157>

Emily Gentles graduated from Duke University with a Masters Degree in Statistical Science and now works at Research Triangle Institute (RTI). Emily is an expert in data linkage, including entity resolution and privacy preserving record linkage. She has conducted research regarding efficient privacy preserving record linkage methods, worked to develop secure linkage systems, and designed innovative record linkage procedures, including manual review protocol. She has experience with project leadership and often collaborates with colleagues from various fields. Additionally, she provides data processing, sampling, statistical programming, and analysis for support on various projects. Other significant areas of experience include Bayesian methods and hierarchical modeling. Ms. Gentles is also knowledgeable about experimental design, information theory, and data privacy. In her free time, Emily enjoys running, crafting, and reading.



MISS CHEYENNE AMOROSO
Universidade da Coruña

Cheyenne is a researcher at the Centro de Investigación en Tecnoloxías da Información e as Comunicaciós (CITIC) at the University of A Coruña. She holds a Mathematics degree from the University of Santiago de Compostela and a Master's in Statistical Techniques and Operations Research, specialising in statistics. She has worked as a data scientist at the Spanish bank ABANCA, focusing on data extraction, text mining and macroeconomic time series analysis. She is currently collaborating with the National Statistics Institute and the University of A Coruña while pursuing her PhD.



DR. MARÍA BUGALLO
Miguel Hernández University of Elche

<https://orcid.org/0000-0001-6290-3510>

María Bugallo holds a Bachelor's Degree in Mathematics and a Master's Degree in Statistical Techniques from the University of Santiago de Compostela (Spain). In December 2024, she earned her Ph.D. in Statistics, Applied Mathematics, and Computer Science from the Miguel Hernández University of Elche (Spain). She is currently working as a Postdoctoral Research Support Technician on a project funded by the Spanish National Institute of Statistics. Her research lies at the intersection of methodological statistics, economics, and social and environmental applications. She focuses on statistical modeling, small area estimation, and computational statistics to support data-driven decision-making in complex real-world situations.



Record Linkage: Recent Developments in Efficiency and Privacy*Emily Gentles, RTI*

Record linkage aims to connect records across multiple databases referring to the same entity (Steorts, 2020). Bayesian methods have the advantage of providing uncertainty estimates for the linkage process, although the tradeoff is often computational scalability. While there are multiple common methods aimed at reducing the computational burden of record linkage algorithms, such as blocking and deduplication, the cost remains high for both frequentist and Bayesian approaches. Additionally, unique identifiers are typically unavailable for use in linkage tasks, so the process utilizes personally identifying information, known as fields, such as first and last name, date of birth, sex, and residential address. Oftentimes the results of record linkage contain sensitive personal information such as medical conditions, sexual orientation, or political affiliation which potentially could be damaging if made public. The goal of Privacy Preserving Record Linkage (PPRL) is to ensure that sensitive information, including identifying information, is not revealed through the linkage process or the resulting matched records (Christen, 2020). In this presentation we will discuss methods for increasing the computational efficiency of record linkage, improving privacy, and the how these methods interact. We'll cover new developments in the literature and discuss the usefulness of some of these methods in real-data situations.

BSTFA: An R Package for Spatio-Temporal Data

by National Statistical Offices, enabling the production of reliable indicators even under complex data conditions. Incorporating robustness into the estimation process helps to ensure that Official Statistics remain accurate, robust and meaningful in a data-rich, yet often imperfect, real-world context.

Beyond the Crisis: Seasonal Adjustment in Long Time Series with Structural Changes*Cheyenne Amoroso, Universidade da Coruña*

Long time series are rarely static: events such as the 2008 financial crisis can drastically alter their behaviour. This talk introduces an approach to seasonal adjustment that accounts for structural changes over time, combining two different models – before and after the event – linked by a dynamically modelled transition period. The aim is to make seasonal adjustment more reliable when data patterns evolve.

The Role of Robust Small Area Estimation in Official Statistics*María Bugallo, Miguel Hernández University of Elche*

National Statistical Offices and private institutions are becoming increasingly interested in detailed data on specific subpopulations to support more informed decision-making. While direct estimates from surveys are generally reliable when the sample size is large, they often become unstable in small domains due to limited data. In such cases, Small Area Estimation is a valuable solution, integrating survey data with other sources to improve precision. However, traditional Small Area Estimation methods can be highly sensitive to outliers and slight deviations from the model's assumptions, which can lead to biased or misleading results. Robust approaches address these issues by reducing the influence of extreme values and enhancing the method's resilience to anomalies. Techniques such as M-quantile models produce more stable and credible estimates. The use of robust Small Area Estimation is becoming increasingly important in Official Statistics, where high-quality, disaggregated data are essential for evidence-based policymaking. Furthermore, robust Small Area Estimation aligns with the quality standards adopted



Session Info

S206

TECHNICAL STATISTICAL SESSION

October 14th, 13:00 - 14:00 UTC

Advances in High-Dimensional Bayesian Modeling

USE SCHED TO ACCESS THE ZOOM ROOM

This session, organized by the International Society for Bayesian Analysis, highlights recent advances in Bayesian modeling of high-dimensional data. In the first talk, Dr. Gemma Moran (Rutgers) will discuss sparse variational auto-encoders for estimating provably identifiable low-dimensional representations of complex, dependent data commonly seen in genomics. Then, Dr. Cecilia Balocchi (Edinburgh) will present a novel approach for approximating posterior distributions over the combinatorially large space of partitions that is based on minimizing the Wasserstein distance between the true and approximate posterior distributions. Finally, Dr. Elizabeth Bersson (MIT) will introduce a new way to estimate large covariance matrices when the sample size is small by carefully leveraging auxiliary information from meta-covariates.

ORGANIZER: Sameer Deshpande, University of Wisconsin–Madison

CHAIR: Sameer Deshpande, University of Wisconsin–Madison

SPONSOR: ISBA



Speaker Bios

S206

**DR. GEMMA MORAN**

Rutgers University

<https://www.gemma-moran.com>

Gemma Moran is a tenure-track Assistant Professor in the Rutgers Statistics Department. Prior to that, she was a postdoc at the Columbia Data Science Institute, working with David Blei. She received her PhD in Statistics from the University of Pennsylvania, advised by Edward George and Veronika Rockova.

Gemma's research develops flexible Bayesian models for analyzing high-dimensional data. Some of her recent projects include developing identifiable and interpretable deep generative models (especially variational autoencoders) and improved tools for Bayesian model criticism.

**DR. CECILIA BALOCCHI**

University of Edinburgh

<https://cecilia-balocchi.owlstown.net>

Cecilia Balocchi is a Lecturer (Assistant Professor) in Statistics at the School of Mathematics, University of Edinburgh. Previously, she was a postdoctoral researcher in Statistics at the University of Torino, working with Stefano Favaro. She is also a member of the "de Castro" Statistics Initiative at Collegio Carlo Alberto. She received her PhD in Statistics from the University of Pennsylvania under the supervision of Ed George and Shane Jensen. Her research interests include Bayesian nonparametrics and hierarchical modeling, clustering, Bayesian regression trees and spatial methods.

**DR. ELIZABETH BERSSON**

MIT

<https://betsybersson.github.io>

Elizabeth Bersson is a postdoctoral fellow at Massachusetts Institute of Technology working with Tamara Broderick. She earned her Ph.D. in Statistical Science from Duke University in May 2024, under the supervision of Peter Hoff. Her research focuses on hierarchical modeling, covariance estimation, small-area estimation, and conformal prediction.





Nonlinear Multi-Study Factor Analysis

Gemma Moran, Rutgers University

In many domains, high-dimensional data exhibits variability that can be summarized by low-dimensional latent representations, or factors. In genomics, for example, genes that comprise a biological pathway may exhibit coordinated expression patterns; by learning which genes are co-expressed, we can develop new hypotheses about the underlying biology. To learn such factors, we propose the sparse variational autoencoder. The underlying model is sparse in that each observed feature (i.e. each dimension of the data) depends on a small subset of the latent factors. In the genomics example, this means each gene is active in only a few biological processes. We prove such sparse deep generative models are identifiable, and apply the sparse VAE to movie ratings, text and genomics data. We then introduce an extension of the sparse VAE to multi-study data. In this data, we expect some factors to be shared across studies, and some factors to be specific to a single study. We prove that the shared factors can be identified. Finally, we apply the multi-study sparse VAE to blood platelet gene expression data from patients across different disease conditions.

can infer the number of factors adaptively. However, they have the disadvantage of shrinking towards a diagonal covariance, often under estimating important covariation patterns in the data. Alternatively, the dimensionality problem is addressed by collapsing the detailed exposure data within chemical classes, potentially obscuring important information. We apply a covariance meta regression extension of Bayesian factor analysis, which improves performance by including information from features summarizing properties of the different exposures. This approach enables shrinkage to more flexible covariance structures, reducing the over-shrinkage problem, as we illustrate in the TESIE data using various chemical features as meta covariates.

Interpreting Uncertainty in Bayesian Cluster Analysis

Cecilia Balocchi, University of Edinburgh

The Bayesian approach to clustering is often appreciated for its ability to provide uncertainty in the partition structure. However, summarizing the posterior distribution over the clustering structure can be challenging, due the discrete, unordered nature and massive dimension of the space. While recent advancements provide a single clustering estimate to represent the posterior, this ignores uncertainty and may even be unrepresentative in instances where the posterior is multimodal. To enhance our understanding of uncertainty, we propose a WASSerstein Approximation for Bayesian clusterIng (WASABI), which summarizes the posterior samples with not one, but multiple clustering estimates, each corresponding to a different part of the partition space that receives substantial posterior mass. Specifically, we find such clustering estimates by approximating the posterior distribution in a Wasserstein distance sense, equipped with a suitable metric on the partition space. An interesting byproduct is that a locally optimal solution can be found using a k-medoids-like algorithm on the partition space to divide the posterior samples into groups, each represented by one of the clustering estimates. Using synthetic and real datasets, we show that WASABI helps to improve the understanding of uncertainty, particularly when clusters are not well separated or when the employed model is misspecified.

Covariance Meta Regression

Elizabeth Bersson, MIT

The motivation of this talk is to improve inferences on the covariation in environmental exposures, motivated by data from a study of Toddlers Exposure to SVOCs in Indoor Environments (TESIE). The challenge is that the sample size is limited, so empirical covariance provides a poor estimate. In related applications, Bayesian factor models have been popular; these approaches express the covariance as low rank plus diagonal and



Session Info

S207

TECHNICAL STATISTICAL SESSION

October 14th, 14:00 - 15:00 UTC

Women in Biostatistics: Insights from the BIOSTATNET Network

USE SCHED TO ACCESS THE ZOOM ROOM

In celebration of the International Day of Women in Statistics and Data Science, this session highlights the contributions of women researchers in the field of biostatistics through the BIOSTATNET network — the Spanish Network for Biostatistics. We will begin by introducing the mission and objectives of BIOSTATNET, followed by presentations from four women statisticians of the network. Each researcher will present her current project, offering insights into diverse methodological approaches and applications in health and biomedical research. This session aims to connect and elevate the visibility of women in biostatistics at both national and international levels.

ORGANIZER: Jaione Etxeberria, Public University of Navarre**CHAIR:** Irantzu Barrio, University of the Basque Country

Speaker Bios

S207



DR. NATALIA VILOR-TEJEDOR

(IRAS), Utrecht University,

<https://github.com/GeneticNeuroStats>

Dr. Natalia Vilor-Tejedor holds a PhD in Biomedicine, a specialized MSc in omics data analysis, and a foundational background in mathematics and statistics. Her research focuses on understanding the aetiology and prevention of complex diseases by exploring the interplay between genetic and environmental risk factors. She currently leads the Exposomics and Data Science group at the Institute for Risk Assessment Sciences (IRAS), Utrecht University, Netherlands. Dr. Vilor-Tejedor also holds dual appointments at the Radboud University Medical Center and the Barcelonaβeta Brain Research Center. Her work integrates multiomics, environmental exposures, and data science to uncover insights into complex molecular diseases. She is deeply committed to inclusive science, actively fostering mentorship, interdisciplinary collaboration, and public engagement.



MISS NAOMI DIZ ROSALES

CITIC - Universidade da Coruña.

<https://orcid.org/0000-0002-6932-7912>

I graduated in Biology from the Universidade da Coruña (UDC) with the Extraordinary Prize and the Galician End of Degree Prize. I specialised in microbial bioenergy in my Final Degree Project and continued in biotechnology by taking the Master in Bioinformatics for Health Sciences (UDC Extraordinary Prize), developing AI tools for macular oedema diagnosis. I began research at the Biomedical Institute of A Coruña, applying machine learning and proteomics in a European project on osteoarthritis biomarkers. I am currently doing a PhD in Statistics in the Modelling, Optimisation and Statistical Inference group (MODES, UDC), developing methodology for mixed models with applications in small area estimation: hospital resource prediction and poverty analysis with a gender perspective. I also combine research with university teaching and science communication to promote statistics and data science among new generations.



DR. ALBA MUÑOZ DEL RÍO

Universidad de Castilla-La Mancha

Doctor in Physics and Mathematics from the University of Castilla-La Mancha, specialized in optimal experimental design applied to nonlinear models in predictive microbiology. Her research develops statistical approaches to improve the accuracy of parameter estimation and methodologies to strengthen designs against uncertainty in nominal parameter values.

She has participated in applied projects in the energy and agri-food sectors and has worked as a Data Scientist at Orange and NTT Data. She is currently part of the Statistics Department at the University of Castilla-La Mancha, where she teaches statistics and programming courses in Python and R across various undergraduate and master's programs. She is a member of the "DoE + Bioinformatics" node of Biostatnet and collaborates in outreach activities such as StatWars, Estalmat, and the Survey Incubator.

<https://github.com/bio2ds-ucm>



DR. SILVIA PINEDA SANJUAN

Universidad Complutense de Madrid

I am a Assistant Professor in the Department of Statistics and Data Science at the Complutense University of Madrid (UCM), where I recently established the Biomedical Data Science & Biostatistics Research Group. I hold a degree in Statistics from the Complutense University of Madrid (UCM) and a PhD in Genetic and Molecular Epidemiology from the Spanish National Cancer Research Center (CNIO), where I completed a dissertation entitled Advanced Statistical Methods for the Integration of Omics Data in Cancer. I completed a postdoctoral fellowship at the University of California, San Francisco (UCSF) in the Bakar Computational Health Sciences Institute, and later returned to Spain after being awarded a grant from the American Association for Cancer Research (AACR) to start my own research group.



Stochastic Journey in Precision Genetic Neuroepidemiology*Natalia Vilor-Tejedor, BarcelonaBeta Brain Research Center, Utrecht University*

This talk will present biostatistical approaches that are helping to advance the field of genetic neuroepidemiology. Recent developments in data science and statistical modeling have improved our ability to identify genetic and environmental factors related to brain health, with implications for risk prediction, disease prevention, and targeted interventions.

I will discuss the use of multivariate models applied to large-scale genomic, environmental, and neuroimaging data, along with statistical tools designed to integrate these diverse sources. A focus will be placed on methods that account for omics data integration, population diversity and sex-specific effects among others, which are critical for producing accurate and generalizable findings.

These advances support the identification of biological pathways and contribute to more precise approaches in medicine and public health. The session will also highlight how data science is enabling progress in this area, and the importance of inclusive research practices in science.

Small Area Estimation for Social and Biomedical Equity: Unity is Strength*Naomi Diz-Rosales, Universidade da Coruña / CITIC*

Small area estimation is a powerful statistical tool for overcoming data scarcity and making accurate estimates for socio-economic and public health policy making. In contexts such as poverty monitoring or pandemic response, traditional methods fail when sample sizes are insufficient, which can mask inequalities in vulnerable groups. This research addresses the challenges of small area estimation using generalised linear mixed models with random slopes, exploiting the principle that 'there is strength in numbers'. By combining information across areas or domains using random effects, these mixed models allow small areas to 'borrow strength' from similar areas, effectively increasing sample sizes and improving precision. We apply this framework to two critical scenarios:

- Estimates of poverty disaggregated by sex at the provincial level in Spain.
 - Real-time ICU occupancy prediction during COVID-19, optimising resource allocation in undersampled health districts.
- Our results demonstrate how small area estimation transforms scarce data into actionable information, reducing biases in policy decisions, from gender-sensitive budget allocations to pandemic triage.

Design Smarter, Experiment Less: Predictive Microbiology Meets Optimal Experimental Design*Alba Muñoz, Universidad de Castilla-La Mancha*

In predictive microbiology, modeling the growth or inactivation of microorganisms under varying environmental conditions is essential for applications such as food safety, shelf-life prediction, and industrial microbiology. These models—commonly referred to as primary and secondary models—are typically fitted using standard experimental procedures that, while widely used, do not always lead to the most accurate parameter estimates. Since the reliability of model-based predictions depends heavily on the

precision of these estimates, improving how data are collected is a key challenge. This work evaluates common experimental strategies from the perspective of Optimal Experimental Design, a statistical approach that enhances efficiency and precision. Applying these techniques allows researchers to obtain more accurate estimates with fewer experiments, reducing both time and material costs. The study highlights how statistical thinking can drive smarter experimentation and support more sustainable, impactful scientific practices across disciplines.

From Genes to Clinical Decisions: The Machine Learning Pathway*Silvia Pineda San Juan, Universidad Complutense de Madrid*

The rapid expansion of omics technologies—such as genomics, transcriptomics, and microbiomics—has generated massive volumes of data that, together with non-omics sources like hospital records and cohort studies, present new challenges for biomedical analysis.

Traditional statistical techniques are often insufficient to capture the complexity inherent in these high-dimensional data. In this context, machine learning emerges as a key tool for transforming this information into knowledge that guides medical decision-making. Using examples from complex conditions such as cancer and transplant rejection, we will explore approaches that tackle these challenges with the goal of developing models that not only predict accurately but also foster interpretability and potential clinical utility.



Session Info

S208

TECHNICAL STATISTICAL SESSION

October 14th, 15:00 - 16:00 UTC

Empowering Progress in Infectious Disease Research through Innovation and Algorithmic Solutions

USE SCHED TO ACCESS THE ZOOM ROOM

Infectious disease research is constantly evolving, and staying ahead requires fresh ideas and practical solutions. This session focuses on analytical examples showcasing how data-driven methods and algorithmic solutions are helping researchers understand and manage infectious diseases.

Our speakers will share their experiences with innovative tools and techniques that are changing the way we approach assay qualification, statistical modeling of infectious diseases, and treatment trials.

The first speaker will introduce a novel algorithm designed to automate the linearity assessment process in assay qualification for vaccine development and disease diagnostics progression. The second speaker will present findings from a meta-analysis using individual participant data to identify viral load outcomes as surrogates for hospitalization and death in COVID-19 treatment trials. The third speaker will explore the spatial diffusion of COVID-19 variants across the continental United States using correlation statistics and mathematical modeling. The fourth speaker will discuss the challenges of using prognostic tools to predict patient outcomes during the COVID-19 pandemic. The final speaker will review the PALM007 trial results, focusing on the efficacy of tecovirimat treatment for clade I Mpox virus in the Democratic Republic of Congo. This session aims to provide researchers with examples of innovative solutions, practical insights, and tools to advance infectious disease research.

ORGANIZER: Ana Ortega-Villa, NIAID**CHAIR:** Ana Ortega-Villa, NIAID

Speaker Bios

S208



MISS MEGAN C. GRIECO

National Institute of Allergy and Infectious Diseases

Megan Grieco is a biostatistician in the Biostatistics Research Branch at the National Institute of Allergy and Infectious Diseases. She graduated from the University of Virginia in 2021 with a BS in Systems and Information Engineering, and from the Rollins School of Public Health at Emory University in 2024 with a MSPH in Biostatistics. Her projects focus on health outcomes in pediatric food allergies and immunologic conditions, as well as vaccine immunogenicity studies for influenza, malaria, and COVID-19.

<https://www.linkedin.com/in/allyson-mateja-44723374>


MISS ALLYSON MATEJA

Frederick National Laboratory for Cancer Research

Allyson Mateja received her bachelor's degree in biology and mathematics from the College of William and Mary in 2015, and her M.S.P.H. in biostatistics from the Rollins School of Public Health at Emory University in 2019. Since then, she has worked as a biostatistician supporting the Biostatistics Research Branch at the National Institute of Allergy and Infectious Diseases. Allyson enjoys data visualization, statistical programming, and statistical methods for assay validations and qualifications.



DR. VIVIANE CALLIER

Leidos Biomedical Research/NIH

Viviane is a biostatistician in the Office of Biostatistics Research at NIAID, where she supports studies related to COVID-19, Ebola, malaria, influenza, dengue, and other infectious diseases. Her work here spans analysis of clinical trials and observational studies, as well as assay development and infectious disease modeling. She is especially interested in malaria. She was a lab biologist and science writer before becoming a statistician.



MISS JING WANG

Frederick National Laboratory for Cancer Research

Jing has been a biostatistician with the Biostatistics Research Branch since late 2013. She holds an M.S. in Biology, with a concentration in immunology and molecular biology, from the University of Wisconsin-Milwaukee, and a second M.S. in Biostatistics from the George Washington University. Jing is passionate about statistical programming and exploring machine learning techniques.



MISS AMY TILLMAN

Frederick National Laboratory for Cancer Research

Amy Tillman is a biostatistician who currently provides statistical support for the National Institute of Allergy and Infectious Diseases; within the Clinical Trials Research Section, she's involved in conducting clinical trials for emerging infectious diseases, with a particular emphasis on trials conducted in resource-limited settings. She holds an MS in Statistics and a quantitative certificate in Fisheries Science from Virginia Tech and a BS in Mathematics with a minor in Statistics from the University of Georgia. Over the years, she's contributed to trials in areas such as cancer, diabetes, and neurodegenerative, inherited, and infectious diseases. In previous roles, Amy served as an independent reporting statistician for phase II+III trials and worked on trials focusing on vaccines and antibodies targeting emerging infectious diseases.



Automating Assessments of Linearity Acceptance Criteria in Assay Qualifications: A Novel Algorithmic Approach

Megan C. Grieco, National Institute of Allergy and Infectious Diseases

In vaccine development and disease diagnostics, assay qualifications play an essential role in the accuracy and reliability of antibody measurements through ensuring that criteria for the standard curve, precision, linearity, and specificity are met. Traditionally, assessing whether antibody data meets the acceptance criteria for linearity has been a manual process in which non-linear concentration points are identified visually and iteratively removed until the criteria are met. This ad-hoc approach, while functional, is time-consuming and prone to subjective biases. This talk introduces a novel algorithm designed to automate the linearity assessment process, offering an objective solution that significantly reduces the time and effort required in traditional methods. The algorithm iterates through possible concentration ranges of the data, starting with those with the widest intervals and that include median concentrations. The algorithm checks R^2 , standardized residuals, and the slope, proceeding to the next possible range if criteria are not met. The algorithm's output will be demonstrated through a Shiny dashboard interface, allowing users to visualize results and summary statistics of a premade dataset or upload their own. This talk will also explore the progression of problem-solving in data science projects, from identifying areas for automation to developing efficient coding solutions.

The Choice of Viral Load Endpoint in Early Phase Trials of COVID-19 Treatments Aiming to Reduce 28-Day Hospitalization and/or Death

Allyson Mateja, Frederick National Laboratory for Cancer Research

While the US FDA allows virologic endpoints in trials for COVID-19 therapeutics, they are not recommended as surrogates for clinical endpoints. No meta-analysis using individual participant data (IPD) has been undertaken to identify viral load outcomes best associated with effects on hospitalization/death. This meta-analysis combined IPD from 23 COVID-19 treatment vs. control comparisons to calculate R^2 , a surrogacy measure quantifying the relationship between the treatment effect on 28-day hospitalization/death and the treatment effect on the surrogate. We estimated R^2 for viral load outcomes at Days 3, 5, and 7, including change-from-baseline, slope, and average area under the curve minus baseline (AAUCMB). R^2 was numerically highest for the change-from-baseline to Day 3 (0.53 [0.26, 0.79]), slightly lower for change-from-baseline to Day 5 (0.49 [0.24, 0.75]) and numerically lower for change-from-baseline to Day 7 (0.40 [0.15, 0.65]). All were statistically significant. Our study is the first to use IPD, allowing us to evaluate viral load collected on various study days as a surrogate to clinical outcomes. Change in $\log_{10}$ (viral load) from baseline to Day 3 or 5 are moderate surrogates for 28-day hospitalization/death and suitable primary endpoints in Phase 2 clinical trials and are preferred over change-from-baseline to Day 7. Slope and AAUCMB require more calculation but didn't improve prediction so aren't recommended.

Spatio-Temporal Spread of COVID-19 Over Three Variant Waves in the Continental United States

Viviane Callier, Leidos Biomedical Research/NIH

Understanding the spatio-temporal spread of infectious diseases is important to improve control; yet most studies lack granular epidemiological data and are complicated by heterogeneities in pre-existing immunity. The COVID-19 pandemic provides a unique opportunity to study the invasion of a novel pathogen in a naive population. Here we apply correlation statistics and mathematical modeling to daily city-level COVID-19 case data between 2020-2022 to dissect the spatial diffusion of different virus variants across the

US. We find that between-city correlation in case incidences varies with geographic distance, supporting the idea of diffusive disease spread up to a radius of 1000-1500 km. The strongest diffusive spread is observed for the Delta wave and the weakest for Omicron. We fit mechanistic transmission models to onset times and find that models combining geographic distance with the population size of recipient cities produce the best fit to all 3 waves, while indicators for pre-existing immunity and vaccination do not improve model fit. Across the three waves, a few large cities like Chicago, Atlanta and Detroit, but also mid-size and smaller cities, mostly in the Midwest, disproportionately contributed to the onward spread of COVID-19. These hotspots of disease transmission could be useful for sentinel surveillance.

Impact of Surge Strain and Pandemic Progression on Prognostication by an Established COVID-19-Specific Severity Score

Jing Wang, Frederick National Laboratory for Cancer Research

Many U.S. crisis standards of care (CSC) guidelines adopted the Sequential Organ Failure Assessment (SOFA) score for triage during the COVID-19 pandemic, though it performed poorly in predicting outcomes for patients with COVID-19. In contrast, disease-specific prognostic tools such as the modified 4C (m4C) score initially showed stronger predictive ability. Yet, the durability of m4C's performance across evolving waves of the pandemic and under varying hospital surge conditions remains unclear. We evaluated the performance of the m4C score among 298,379 adults hospitalized with COVID-19 across 281 U.S. hospitals between March 1, 2020, and January 31, 2022. A hospital-surge index, representing a severity-weighted measure of local COVID-19 caseloads, was computed for each hospital-month to reflect real-time care strain. We assessed model discrimination using area under the receiver operating characteristic curves (AUC) and evaluated calibration over time. The m4C score showed good discrimination early but declined in later waves. Calibration worsened over time, and performance did not significantly improve with surge adjustment. Although relatively robust to hospital strain, m4C's predictive performance declined as the pandemic evolved. These findings highlight the need for CSC protocols to regularly re-evaluate prognostic tools, ensuring they remain accurate and equitable under changing clinical conditions.

A Deeper Dive into the Results of the PALM007 Trial for Clade I MPox Virus (MPXV) Infection in the Democratic Republic of Congo (DRC)

Amy Tillman, Frederick National Laboratory for Cancer Research

The PALM007 trial enrolled adults and children with clade I MPXV in the DRC. Randomization was stratified according to trial site and days since symptom onset, and patients were randomly assigned in a 1:1 ratio to receive tecovirimat or placebo for 14 days, along with supportive care. The trial's primary endpoint was time to lesion resolution. Although treatment with tecovirimat was safe, primary efficacy results did not show any significant differences between tecovirimat and placebo. Several baseline factors were associated with slower lesion resolution, and adjustments for each factor did not alter conclusions about treatment efficacy. Prespecified subgroup analyses suggested a treatment-by-site interaction which led us to perform an exploratory, post hoc analysis that adjusted for site-specific imbalances. With these adjustments, no interaction of treatment effect with site was observed. During this talk, I will briefly review the overall PALM007 trial results and take a deeper dive into baseline risk factors and the exploratory, post hoc analysis mentioned above. Research identifying and evaluating additional therapeutics is still needed, but our findings highlight the importance of conducting rigorous clinical trials during an outbreak, especially for investigational treatments that are available for use but have unknown efficacy.



Session Info

S209

TECHNICAL STATISTICAL SESSION

October 14th, 16:00 - 16:30 UTC

Statistical Innovations for Error-Prone Data

USE SCHED TO ACCESS THE ZOOM ROOM

Measurement error is a pervasive challenge in statistical analysis, with the potential to bias results and undermine inference across diverse applications. This session highlights recent methodological advances for addressing measurement error in complex settings. The first talk presents a Bayesian approach to zero-inflated Poisson models with error-contaminated count responses, examining when ignoring error can still yield valid results and how to handle identifiability issues. The second talk introduces an improved regression calibration method for external validation study designs that enhances transportability and reduces bias in exposure-disease association estimates. Together, these presentations showcase innovative tools for producing more reliable and interpretable results when working with imperfect data.

ORGANIZER: Sarah Lotspeich, Wake Forest University**CHAIR:** Sarah Lotspeich, Wake Forest University

Speaker Bios

S209

**DR. GRACE YI**

University of Western Ontario

Grace Y. Yi is a professor at the University of Western Ontario where she currently holds a Tier I Canada Research Chair in Data Science. She is recognized as one of the influential women in Statistics. Her research interests focus on statistical methodology to address challenges concerning measurement error, causal inference, missing data, high-dimensional data, and statistical machine learning. She authored the monograph *Statistical Analysis with Measurement Error or Misclassification: Strategy, Method and Application* (2017, Springer) and co-edited *Handbook of Measurement Error Models* (Grace Y. Yi, Aurore Delaigle, and Paul Gustafson, 2021, Chapman & Hall/CRC).

**DR. DONNA SPIEGELMAN**

Yale University

Donna Spiegelman is the Susan Dwight Bliss Professor of Biostatistics at the Yale School of Public Health and Professor of Statistics and Data Science at Yale University. She is the founding director of the Yale Center for Methods in Implementation and Prevention Science (CMIPS) and leads the Global Oncology program at Yale Cancer Center. One of the few with a joint doctorate in biostatistics and epidemiology, she bridges the two disciplines to address complex problems in public health research.

Her work focuses primarily on developing and applying innovative statistical methods for implementation science, including pragmatic trials, causal inference, measurement error correction, and the Learn as You Go (LAGO) adaptive trial design. She advances the design and analysis of large-scale public health interventions to improve effectiveness, efficiency, and generalizability.





Zero-Inflated Poisson Models with Measurement Error in Response

Grace Yi, University of Western Ontario

The analysis of zero-inflated count data is often based on a mixture model which facilitates excess zeros in combination with a Poisson distribution, and various inference methods have been proposed under such a model. Such analysis procedures, however, are challenged by the presence of measurement error in count responses. We propose a new measurement error model to describe error-contaminated count data. We show that ignoring the measurement error effects in the analysis may generally lead to invalid inference results, and meanwhile, we identify situations where ignoring measurement error can still yield consistent estimators. Furthermore, we propose a Bayesian method to address the effects of measurement error under the zero-inflated Poisson model and discuss the identifiability issues. Numerical studies are conducted to evaluate the performance of the proposed method.

Improving Transportability of Regression Calibration Under the Main/External Validation Study Design

Donna Spiegelman, Yale University

In epidemiology, obtaining accurate individual exposure measurements can be costly and challenging. Thus, these measurements are often subject to error. Regression calibration with a validation study is widely employed as a study design and analysis method to correct for measurement error in the main study due to its broad applicability and simple implementation. However, relying on an external validation study to assess the measurement error process carries the risk of introducing bias into the analysis. Specifically, if the parameters of regression calibration model estimated from the external validation study are not transportable to the main study, the subsequent estimated parameter describing the exposure-disease association will be biased. In this work, we improve the regression calibration method for linear regression models using an external validation study. Unlike the original approach, our proposed method ensures that the regression calibration model is transportable by estimating the parameters in the measurement error generating process using the external validation study and obtaining the remaining parameter values in the regression calibration model directly from the main study.



Session Info

S210

OTHER

October 14th, 16:30 - 17:00 UTC

Data Science for All: Creating a Thriving Community for Interdisciplinary Students via Industry Experiential Learning Projects

USE SCHED TO ACCESS THE ZOOM ROOM

The demand for data literacy now spans every discipline yet students arrive with vastly different levels of computing experience and confidence. How can universities help students, from any major, thrive in a data-driven business environment?

Entering its eighth year, The Data Mine at Purdue University engages more than 2,000 graduate and undergraduate students from over 130 majors in academic year-long, interdisciplinary teams that learn data science under the mentorship of industry partners. Students collaborate on 80+ projects annually with corporate partners across industries such as aerospace, agriculture, manufacturing, and pharmaceutical science. The students meet with the corporate partners every week as part of their class.

This session will show how The Data Mine model, rooted in hands-on learning, teamwork, and mentoring, creates an inclusive pathway into data science. By embedding industry experience into the academic year, students develop technical fluency, problem-solving skills, business acumen, and confidence, regardless of prior programming background.

The Data Mine is a repeatable, scaleable, and financially self-supported model and this session will demonstrate how universities can create environments where students of all majors thrive as future data leaders while generating genuine value for industry partners.

ORGANIZER: Margaret Betz, Purdue University

CHAIR: Margaret Betz, Purdue University



Speaker Bios

S210

**MARGARET BETZ**

Purdue University

<https://www.linkedin.com/in/margaretbetz/>

Margaret "Maggie" Betz is Managing Director of The Data Mine at Purdue University in Indianapolis. Maggie launched a new location of The Data Mine in Indianapolis in fall 2024 after a successful four years building the program in West Lafayette. In her previous role of Managing Director of Corporate Partners, Maggie grew the Corporate Partners cohort of The Data Mine from 1 partnership in 2018 to more than 65 corporate partners and 800+ students in 2023. Maggie is changing the way students prepare for career readiness in a data-driven world and how companies recruit and interact with talent. She is an alumna of Purdue University (B.S. Statistics) and Indiana University (M.S. Biostatistics). Maggie serves on the board of the Central Indiana Chapter of the American Statistical Association and member for the Committee for Membership Retention and Recruitment (CMRR).

**JESSICA GERLACH**

Purdue University

Jessica Gerlach is a Corporate Partner Liaison for The Data Mine at Purdue University in West Lafayette and Indianapolis. She drives industry outreach, connecting companies with talented students to bring real-world projects into The Data Mine Corporate Partners program. This creates a mutually beneficial structure: students gain invaluable, hands-on experience, while companies interact with and recruit top talent in a deeper, more involved way. Jessica is an alumna of Indiana University.



Session Info

S211

TECHNICAL STATISTICAL SESSION

October 14th, 18:00 - 18:45 UTC

SEIO Women in Statistics and Data Science for the Common Good

USE SCHED TO ACCESS THE ZOOM ROOM

This session, organised by the SEIO Women Commission of the Spanish Society of Statistics, Operations Research and Data Science (SEIO), presents 3 talks of relevant women researchers whose work on Statistics and Data Science is oriented to the common good:

- Begoña Vitoriano, associate professor
- M. Dolores Martínez Miranda, full professor
- Jessica Rodriguez-Pereira, Ramón y Cajal researcher in the Universitat Politècnica de Catalunya

ORGANIZER: María D. Ruiz-Medina, University of Granada

CHAIR: María D. Ruiz-Medina, University of Granada



Speaker Bios

S211



DR. BEGOÑA VITORIANO

Universidad Complutense de Madrid

Begoña is Associate Professor at the Department of Statistics and Operational Research and the Interdisciplinary Mathematics Institute (IMI) of Complutense University of Madrid (UCM). PhD in Mathematics (Operational Research), she has been university lecturer since 1990. Begoña has been involved in research projects all along her career, especially in knowledge and technology transfer, leading projects since 1998. Currently, she leads the UCM research group on Decision Aid Models for Logistics and Disaster Management, the UCM team of the Horizon Europe project HURRICANE, and a university cooperation for development project in Mozambique. She is author of more than 60 publications, some of them being references in her area. She has been coordinator of several PhD and Master programs of Complutense University, and several masters in Statistics in El Salvador, Peru and Mozambique since 1996. She has been the SEIO President (2022-25), being currently the SEIO Past President (2025-26).

<https://www.ugr.es/~mmiranda/>



PROF. MARÍA D. MARTÍNEZ MIRANDA

University of Granada

María Dolores Martínez-Miranda is Full Professor of Statistics at the University of Granada (Spain), where she has been a member of the Department of Statistics and Operations Research since 1994. She started her research career with a PhD on non-parametric regression at the University of Granada. In 2012 she was awarded with a Fellowship for career development in the Marie Curie Intra-European Fellowships framework for two years. As Senior Research Fellow in Bayes Business School, City, University of London, she developed an interdisciplinary research project in Mathematics with high impact in the insurance industry. She has published more than 40 papers in JCR-indexed journals. She serves as Associate Editor for several statistical journals, and has been President of SEIO since June 2025.

<https://cientificasinnovadoras.fecyt.es/cientificas/jessica-rodriguez-pereira>



DR. JESSICA RODRIGUEZ-PEREIRA

Universitat Politècnica de Catalunya

Investigadora Ramón y Cajal en la Universitat Politècnica de Catalunya, donde también curso sus estudios de Ingeniería Civil (2012), máster en logística (2014) y doctorado en Estadística e Investigación Operativa (2018). Estancias de investigación en el Centre Interuniversitaire de Recherche sur les Réseaux d'Entreprise, la Logistique et le Transport (Canadá), la Universität Heidelberg (Alemania), y postdoctoral en el HEC Montréal (Canadá) y en la Universitat Pompeu Fabra (España). Su investigación se centra en el estudio de modelos de optimización matemática para la resolución de problemas relacionados con la logística, el transporte y el diseño de redes. Galardonada con diversos premios. Fruto de su trabajo de investigación, es coautora de artículos de investigación publicados en prestigiosas revistas científicas y ha participado en congresos tanto nacionales como internacionales. Participa en proyectos de investigación y colabora con entidades para el bien social, y divulgación.





Data Science for Decision Making in Disaster Management: Scenarios Generation for Stochastic Optimisation

Begoña Vitoriano, Complutense University of Madrid

The disaster management cycle is a process involving several phases, some of them before a disaster occurs (Prevention and Preparedness) and others during or afterwards (Response and Recovery). Decision-support models for disaster management must deal with a high level of uncertainty, and stochastic optimisation is a powerful tool for obtaining robust solutions. They can incorporate the uncertainty through quantified and georeferenced scenarios of potential disasters in the targeted area. This presentation introduces a methodology for generating scenarios for a multi-stage stochastic model for the location and sizing of warehouses (strategic decisions) and the budget allocation and pre-positioning of relief aid (tactical decisions), considering response scenarios (operational decisions). The methodology is based on historical data, which are usually scarce and incomplete, especially for disasters in developing countries. The methodology is illustrated in a case study of Mozambique.

Monitoring and Forecasting a Developing Pandemic With Available Data

María Dolores Martínez Miranda, University of Granada

We present a full dynamic system designed to describe and forecast both the spread and severity of a developing pandemic. Our proposal is rapid to implement, works with available—even low-quality—data, and can be applied across countries. It is simple to communicate and allows for a division of labor in which epidemiological expertise can be incorporated transparently. Understanding pandemic dynamics is crucial for decision-making and policy design. Potential applications include vaccine prioritisation and planning, hospital occupancy and mortality forecasting, elder care strategies, and resource optimisation during health crises. This highlights the value of robust benchmark models that can be used from the very beginning of a pandemic, providing reliable early insights without requiring detailed data, and facilitating communication across national and regional contexts.

Los Modelos Matemáticos Y Las Operaciones Humanitarias

Jessica Rodríguez-Pereira, Universitat Politècnica de Catalunya

En contextos de crisis humanitaria —ya sea por desastres naturales, conflictos armados o desplazamientos forzados— la toma de decisiones rápida y basada en evidencia es crucial para salvar vidas y optimizar recursos. Esta ponencia explora el papel transformador de las herramientas y modelos matemáticos en el diseño, planificación y evaluación de operaciones humanitarias. Se presentan casos de aplicación donde técnicas como la optimización o el análisis geoespacial han permitido mejorar la distribución de ayuda, anticipar necesidades logísticas y evaluar el impacto de las intervenciones.



Session Info

S212

TECHNICAL STATISTICAL SESSION

October 14th, 19:00 - 19:45 UTC

Advances in Bayesian and Spatial Methods for Environmental and Public Health Research

USE SCHED TO ACCESS THE ZOOM ROOM

Modern public health challenges often involve data that are misaligned across time and space, contain complex dependencies, and require flexible modeling strategies. This session showcases recent methodological innovations in Bayesian, spatial, and spatiotemporal modeling that address these issues in diverse environmental and health applications. The featured approaches improve alignment and integration of multi-source data, account for complex temporal patterns and interactions, and provide more accurate, fine-scale estimates of population health outcomes.

ORGANIZER: Sarah Lotspeich, Wake Forest University**CHAIR:** Sarah Lotspeich, Wake Forest University

Speaker Bios

S212



JIHYEON KWON

Drexel University

<https://drexel.edu/dornsife/academics/departments/epidemiology-biostatistics/biostatistics-student-profiles/>

Jihyeon (Ji) Kwon is a PhD candidate in Biostatistics at Drexel University, advised by Dr. Harrison Quick. Her research focuses on developing methods to evaluate and restrict the informativeness of spatial models in disease mapping. She holds a BA in Statistics from Korea University and an MS in Statistics from Wake Forest University. Previously, she worked as a post-master's student at Los Alamos National Laboratory, where she developed a spatiotemporal model for Histoplasmosis. Her research interests include Bayesian statistics, spatiotemporal statistics, and disease mapping, with applications in public health, infectious diseases, and social sciences.



DR. KAYLEIGH KELLER

Colorado State University

<https://www.kellerbiostat.com/>

Dr. Kayleigh Keller is an Associate Professor in the Department of Statistics at Colorado State University and she also has an appointment in the Colorado School of Public Health. Her research is primarily in environmental biostatistics, where she develops statistical methods for studying the human health impacts of environmental exposures. Her projects involve applications across the lifecourse, ranging from perinatal health to cognitive aging.



DR. EVA MURPHY

Wake Forest University

<https://evamurphy100.github.io/>

Eva Murphy is a Teacher-Scholar Postdoctoral Fellow at Wake Forest University, where she teaches statistics courses and is mentored by Dr. Staci Hepler. She earned her PhD in Mathematical Sciences from Clemson University, SC, under the supervision of Dr. Whitney Huang. Her research develops statistical models to understand complex environmental and public health problems through latent process modeling, particularly for spatial and spatio-temporal data. Current projects include modeling opioid-related outcomes across space and time, integrating data from multiple geographic levels, and studying how extreme events disrupt addiction treatment access.



A Spatio-Temporal Hierarchical Model to Account for Temporal Misalignment in American Community Survey Explanatory Variables

Jihyeon Kwon, Drexel University

The American Community Survey (ACS) provides essential demographic and socioeconomic data for U.S. communities, publishing 5-year estimates for all areas and 1-year estimates only for large populations each year. Because the 5-year estimates are available for all years and geographies, they are widely used in public health research; however, they average over multiple years, which can mask important year-to-year changes and lead to biased estimates. This bias is particularly problematic when studies aim to estimate cross-sectional associations at a specific time, as many explanatory variables—especially economic and social variables—vary over time. As interest in social determinants of health continues to grow, accounting for this temporal misalignment between multi-year covariates and yearly outcomes becomes increasingly important. To address this challenge, a Bayesian spatio-temporal model is proposed that incorporates ACS sampling uncertainty while aligning covariates and outcomes in time. The method is illustrated using county-level data from North Carolina, estimating the yearly association between community characteristics and frequent mental distress from 2014 to 2018.

areal units, which provides a more granular understanding of the opioid epidemic. Furthermore, this approach allows us to identify areas with varying levels of opioid burden and reveals local regions with relatively high burden that county-level analyses might miss. Finally, we highlight the need for careful consideration when relying solely on ZIP code level data for naloxone, as it may lead to misinterpretations, particularly in rural regions.

Distributed Lag Interaction Model with Index Modification

Kayleigh Keller, Colorado State University

Associations between exposure to air pollution during pregnancy and birth and child health outcomes are frequently analyzed using a distributed lag model. However, these associations may be modified by multiple factors of unknown relative importance and varying measurement accuracy. In this talk, I will present a distributed lag interaction model with data-driven index modification that allows for effect modification of a functional predictor by a weighted average of multiple modifiers. Our model allows for simultaneous estimation of modifier index weights and the exposure-time-response function via a spline cross-basis in a Bayesian hierarchical framework. Through simulations, we showed that our model out-performs competing methods when there are multiple modifiers of unknown importance. We applied our method to two birth cohorts to estimate impacts of air pollution on perinatal and birth parent health outcomes.

Integrating Data at Multiple Spatial Scales to Estimate the Local Burden of the Opioid Syndemic

Eva Murphy, Wake Forest University

The opioid epidemic has been particularly severe in Ohio, prompting significant efforts to understand its spatial patterns, mainly using available data at the county level. However, relying solely on county-level analysis can overlook crucial information relevant to localized effects. To address this, we integrate spatially misaligned data observed at the county and ZIP code levels to explore the complex interaction of five opioid-related outcomes, providing a more detailed local understanding of the opioid epidemic. We demonstrate how to map ZIP-code level data to ZIP-code Tabulation Areas (ZCTAs) and relate the county-level and ZCTA-level outcomes to a spatially correlated latent factor. The latent factor is defined on the intersection of the misaligned



Session Info

S213

HISTORY OF WOMEN STATISTICIANS SESSION

October 14th, 19:45 - 20:30 UTC

Statistician Woman in Iran

USE SCHED TO ACCESS THE ZOOM ROOM

We will discuss the situation of women in the workplace, especially in statistics and data science.

ORGANIZER: Mina Aminghafari, University of Calgary

CHAIR: Mina Aminghafari, University of Calgary



Speaker Bios

S213



MISS HANIEH SAEIDI

Amirkabir University of Technology

PhD candidate at Amirkabir University of Technology



DR. FERDOS GORJI

Amirkabir University of Technology

Professor (Assistant) at Amirkabir University of Technology



DR. MINA AMINGHAFARI

University of Calgary

Dr. Mina Aminghafari is an Associate Professor in the Department of Mathematics and Statistics at the University of Calgary. Her research lies at the intersection of high-dimensional statistics, machine learning, and applied data science, particularly on clustering theory and statistical learning and their applications in health and environmental sciences. She has contributed significantly to methodological advances in regularization techniques, co-clustering, and functional data analysis. Her work frequently bridges theory and practice, motivated by real-world challenges in biomedical diagnostics, autoimmune disease research, and environmental monitoring. Dr. Aminghafari actively mentors highly qualified personnel at undergraduate, graduate, and doctoral levels, focusing on their development. She also leads several interdisciplinary research projects and training programs, including co-leading the Prairie hub of STAGE (Statistical Genetics and Genetic Epidemiology) and one of the organizers of M2PI (Math to Power Industry) for the Pacific Institute of Mathematical Sciences (PIMS).



Session Info

S214

CAREER OR PROFESSIONAL DEVELOPMENT SESSION

October 14th, 20:30 - 21:00 UTC

Role of Brazilian Women in Statistics

USE SCHED TO ACCESS THE ZOOM ROOM

Currently, there are 35 higher education institutions in Brazil that offer undergraduate programs in Statistics. The first to do so was the Federal University of Rio de Janeiro, which began offering the program in 1946. The profession of Statistician was officially established by law in 1965.

The Brazilian Statistical Association (Associação Brasileira de Estatística – ABE) was founded in 1984, following several years of productive discussion within the national statistical community regarding its feasibility and relevance. Its creation emerged naturally from the organization of the National Symposiums on Probability and Statistics, which began in 1974.

Since its founding and continuing to the present day, the statistical community in Brazil has experienced significant growth, both academically and professionally. Professors Vera Tomazella and Silvia Regina Lopes served as presidents of ABE during the 2014–2016 and 2008–2010 terms, respectively. They addressed the role of Brazilian women in statistics, their areas of interest, and their most recent publications.

ORGANIZER: Jeniffer Johana Duarte Sanchez, Federal University of Ceará

CHAIR: Jeniffer Johana Duarte Sanchez, Federal University of Ceará



Speaker Bios

S214



DR. VERA TOMAZELLA

Universidade Federal de São Carlos

<http://lattes.cnpq.br/8870556978317000>

Vera Tomazella is a Full Professor at the Federal University of São Carlos-UFSCar. She holds a degree in Mathematics from the Federal University of Maranhão, a Master's in Computer Science and Computational Mathematics from the University of São Paulo (1994) and a PhD in Computer Science and Computational Mathematics from the University of São Paulo (2003). She has a postdoctoral degree from the University of Valencia-Spain and a postdoctoral degree from the University of Manchester-UK. Elected member of the International Statistical Institute (ISI) and member of the IBS Education Committee. President of the Brazilian Region of the International Society of Biometrics (RBras). She works mainly in the following areas: Survival and Reliability Analysis, Risk Models, Diagnostic Analysis, Bayesian Inference, Non-Linear Models. In 2024, she was honored as a Professor at the 7th Workshop on Survival Analysis and Applications at UnB – University of Brasília.



DR. SILVIA COSTA LOPES

Universidade Federal do Rio Grande do Sul

<http://lattes.cnpq.br/0811754132547633>

Graduate at Bacharelado em Matemática from Universidade Federal do Rio Grande do Sul (1972), master's at Mathematics from Associação Instituto Nacional de Matemática Pura e Aplicada (1976) and ph.d. at Estatística Matemática from University Of Maryland College Park (1991). Has experience in Probability and Statistics, acting on the following subjects: spectral analysis, pontos fixos, chaotic time series, forecasting and simulations

In 2023, she was honored as a professor at the 20th School of Time Series and Econometrics by the Brazilian Statistical Association (ABE). In 2019, she was honored as a professor at the Brazilian Meeting of Women Mathematicians by IMPA (Institute of Pure and Applied Mathematics).



Modelagem de Sobrevivência com Fragilidade e Fração de Cura: Fundamentos, Avanços e Aplicações em Dados Médicos

Vera Lucia Damasceno Tomazella, Universidade Federal de São Carlos

Esta apresentação tem como objetivo discutir avanços metodológicos recentes na análise de dados de sobrevivência, com ênfase em cenários em que a heterogeneidade não observada e a presença de indivíduos curados desempenham papel central. Os modelos de fragilidade permitem captar fatores latentes que influenciam o risco de falha, enquanto os modelos de fração de cura consideram a existência de uma parcela da população que não apresenta o evento de interesse, mesmo após longos períodos de acompanhamento. Serão exploradas algumas abordagens teóricas, bem como exemplos práticos com dados médicos. A ênfase recai sobre como esses modelos podem aprimorar a compreensão dos mecanismos de risco em doenças crônicas e em estudos clínicos, oferecendo ferramentas estatísticas mais flexíveis e realistas para subsidiar decisões em saúde pública, orientar o planejamento de tratamentos e apoiar a formulação de políticas de cuidado.

Hellinger Integral and Large Deviations Properties in Testing Parameters in CIR and Vasicek Models

Silvia Regina Costa Lopes, Universidade Federal do Rio Grande do Sul

Proper representation formulas for the Hellinger integral of the distribution of the Cox-Ingersoll-Ross (here denoted by CIR) and Vasicek models are presented with their likelihood and least squares functions. Under the assumption that only the mean level parameter is unknown both a maximum likelihood type and a least squares type estimator are derived for this parameter, for both models. Large Deviation properties for Vasicek and CIR models are illustrated as possible applications for the famous stochastic differential equations in Finance. We also present some simulations for both, the maximum likelihood and the least squares estimators, for both models. This is joint work with A.V. Medino (Mathematics Department, University of Brasília) and F.S. Quintino (Statistics Department, University of Brasília).



Session Info

S215

TECHNICAL STATISTICAL SESSION

October 14th, 21:00 - 21:45 UTC

Next Generation Data Scientists: Empowering Young Women Through Research

USE SCHED TO ACCESS THE ZOOM ROOM

This session will be organized by the Florence Nightingale Day for Statistics and Data Science, a 501(c)(3) Non-Profit Organization.

Florence Nightingale was a pioneer in data visualization. She used data in a novel and effective way to improve care for wounded soldiers during the Crimean War and to advance public health more broadly. The Florence Nightingale Day (FNday) for Statistics and Data Science non-profit organization was established to honor her legacy and to inspire young people—especially girls—to pursue careers in statistics and data science by demonstrating how data can be used to solve real-world problems and improve lives.

This session aims to empower the next generation of women in statistics and data science by highlighting the work of three exceptional high school students. They will present original research projects that apply statistical and data science methods to tackle real-world challenges, ranging from exploring the biological effects of synthetic fibers, to uncovering psychological strategies that enhance athletic performance, to analyzing how nutrition affects weight gain.

First, second, and third prizes will be awarded to the presenters, with the selection made by the FNday Committee. This session offers an excellent opportunity to witness the future of data science and reinforces the importance of encouraging young women to pursue careers in statistics and related fields.

ORGANIZER: Xiaokui MO, The Ohio State University

CHAIR: Asuman Turkmen, The Ohio State University



Speaker Bios

S215

**MISS HEERA NAIR**

Bellbrook High School

Heera Nair is a senior at Bellbrook High School with a strong interest in science, leadership, and the arts. She has earned two State Superior titles at the Ohio State Science Fair and is a proud Girl Scout who has completed her Gold Award, the highest honor in Girl Scouting. Outside of academics, Heera is passionate about theater and enjoys acting in plays. In her free time, she also loves to read and explore new stories both on the stage and on the page.

**MISS SOPHIA SZOLOS**

Athens High School, Ohio

Sophia Szolosi is currently a Senior at Athens High School where she has engaged in various research projects related to social sciences. She has earned the ranking 'Superior' at the Ohio State Science Fair five times and won the American Psychological Association Award for her numerous research studies. She is also an avid runner and a member of her high school's cross country/track program where she won seven state titles across different distance events. She is committed to Duke University to run track and cross country and continue her academic career in the Fall of 2026.

**MISS MCKENSIE MCCOY**

Northwestern High School

McKensie McCoy is a first-year nursing major at The Ohio State University. She has a passion for science, learning, and caring for others. McKensie previously graduated from Northwestern high school in West Salem, Ohio, where she conducted research for her senior capstone project. She has also been involved in Health Occupations Students of America (HOSA), band, and cross country.



General Onset of Symptoms in RP Patients*Heera Nair, Bellbrook High School*

Retinitis pigmentosa, a group of inherited retinal disorders, typically starts with nyctalopia (night blindness) and gradually progresses to constriction of visual field, reduction in visual acuity and total blindness in affected individuals. The age at which onset of symptoms happens varies, depending on the specific genetic mutation. However, as we analyze current research data on RP, a few general patterns emerge. The objective of this research is to analyze these patterns and develop/test hypotheses related to onset of RP regardless of genetic mutation. Based on our analysis of currently available research, we hypothesized that symptom onset follows a predictable timeline, with night blindness appearing by third decade, visual field constriction by fourth decade & visual acuity reduction appearing by the fifth decade. An online survey was distributed through RP-focused social media groups to collect self-reported data from 160+ individuals on the age of onset for these key symptoms. Statistical tests were performed to compare the online survey data with expected distributions derived from available research data. The performed statistical tests indicated no significant difference between online survey and expected onset of symptoms. Night blindness consistently emerged by age 30, followed by visual field constriction by age 40 and visual acuity reduction by age 50. These findings may aid in developing prognostic models and early intervention strategies for RP patients.

Investigating the Effects of Two Emotional Regulation Strategies on Runner Performance*Sophia Szolosi, Athens High School*

The ability of a runner to control their emotions prior to a race could be a critical component to their performance and success. The use of different emotional regulation strategies may play an important role in a runner's performance. The following study investigated the effect that two specific emotional regulation strategies had on a runner's performance. A total of 15 self-identified runners participated in the project. Runners completed two 2k time trials following an intervention that involved listening to an audio training session that aligned with either a cognitive reappraisal or visual imagery emotional regulation strategy. Results from a series of paired samples t-tests revealed that participants' overall time performance was statistically better for the cognitive reappraisal condition compared to the visual imagery condition. Information obtained from this science fair project could have implications in the how coaches prepare runners or how runners prepare for a race.

Munchin Mice: The Effects of Artificial Sweeteners on Weight Gain in Mice*McKensie McCoy, The Ohio State University*

Today, the obesity rate in America has skyrocketed to an all time high, with reports claiming that 81% of Americans are obese. In order to combat this epidemic, companies have turned to artificial sweeteners as an additive in foods and beverages with the marketing promise that these artificial sweeteners help aid in weight loss. Meanwhile, other research suggests that artificial sweeteners may increase weight due to their negative impacts on the food reward pathway. To test these contradicting studies, 15

Mus musculus mice were divided into 5 cages. Over the course of 11 weeks, each cage received the acceptable daily intake according to the Food and Drug Administration of either aspartame, sucralose, stevia, table sugar, or no sweetener. The sweeteners were added into their water, and their weights were recorded weekly. The starting and ending weight of each mouse was not statistically different for the control and each sweetener group. It was found that after 11 weeks of experimentation there was no statistical significance between the control and any of the other sweeteners, failing to reject the null hypothesis. This concludes that artificial sweeteners do not impact the weight of mice.



Session Info

S216

TECHNICAL STATISTICAL SESSION

October 14th, 21:45 - 22:30 UTC

Innovative Statistical Approaches to Mitigate Data Challenges in Women's and Maternal Health

USE SCHED TO ACCESS THE ZOOM ROOM

Women's and maternal health research is crucial for advancing our understanding of reproductive health, improving health outcomes, and shaping effective interventions. However, many studies face significant challenges due to error-prone and missing data. This session, organized by the Caucus for Women in Statistics, brings together five expert speakers from various backgrounds and career stages to discuss innovative statistical approaches that they have designed to address these challenges, offering new tools and methodologies to enhance data reliability and inform better health interventions.

ORGANIZER: Sarah Lotspeich, Wake Forest University**CHAIR:** Sarah Lotspeich, Wake Forest University

Speaker Bios

S216



DR. SADIE MIRZAEI

St. Jude Children's Research Hospital

<https://www.stjude.org/directory/m/sedigheh-mirzaei.html>

Dr. Sadie Mirzaei has a broad background in statistics with specific training and experience in analyzing survival data. She received her PhD in Statistics from Indian Statistical Institute in Kolkata, India and completed fellowship training at Duke-NUS Medical School and the NIH. Dr. Mirzaei joined St. Jude as an Assistant Member in 2018 and since then, has built robust collaborations with the Department of Epidemiology and Cancer Control. She serves as a co-investigator for several studies within CCSS and SJLIFE. Dr. Mirzaei is passionate about the role biostatistical analysis plays in further St. Jude's mission, and welcomes exchange with researchers across, and beyond, the institution.



DR. ANA ORTEGA-VILLA

National Institute of Allergy and Infectious Diseases

<https://www.niaid.nih.gov/about/brb-staff-ortega-villa>

Dr. Ana M. Ortega-Villa joined the Biostatistics Research Branch (BRB) in 2018 and serves as a mathematical statistician. Prior to joining the BRB, Dr. Ortega-Villa obtained her Ph.D. in Statistics from Virginia Tech and completed post-doctoral fellowships at both the Eunice Kennedy Shriver National Institute of Child Health and Human Development and the National Cancer Institute. Her interests include longitudinal data, mixed models, vaccines, immunology, research capacity building, statistics education, and initiatives that foster a culture of belonging.



DR. KIMBERLY A. H. WEBB

University of Pittsburgh

<https://www.gim-crhc.pitt.edu/people/kimberly-hochstedler-webb-phd>

Kimberly A. H. Webb, PhD, is an Assistant Professor of Medicine in the Division of General Internal Medicine. Prior to joining the University of Pittsburgh, Dr. Webb developed novel bias-correction methods for misclassified outcome and mediator variables in observational studies. She used these methods to study patterns of misdiagnosis in diseases like myocardial infarction and gestational hypertension, as well as algorithmic fairness concerns in the pretrial detention system. In addition to her methodological contributions, Dr. Webb is interested in applied research in social determinants of health and healthcare decision-making.

Dr. Webb was a varsity swimmer in college and now likes to stay active by hiking, running, and swimming with the Pittsburgh Elite Aquatics "Masters" group. She also enjoys baking and trying new restaurants.



Association Analysis of Imperfectly Recalled Health Conditions among Childhood Cancer Survivors

Sadie Mirzaei, St. Jude Children's Research Hospital

The Childhood Cancer Survivor Study (CCSS) is an enriching resource for survivorship research. It provides valuable information on the incidence and prevalence of various health conditions and the ages at their first occurrence. However, the data may exhibit partial incompleteness due to survivors' challenges in recalling the exact onset age, leading to interval censoring. A complicated cancer treatment received, which potentially affects their memory, together with the evident fact of memory fading with time, makes the censoring informative. We propose a model where the probability of not recalling the onset age is influenced by the time elapsed since the onset of GHD and radiotherapy to the brain. This talk proposes a Cox regression model. The model introduces a novel semi-parametric inference for estimating regression parameters and such data's baseline hazard (or distribution). To demonstrate the method's applicability, we apply it to the analysis of the time to deficiency of growth hormones using data from the CCSS.

model and provide an R package to implement the proposed methods. Using these methods, we assess the misclassification-corrected effect of maternal age on preterm birth risk, while simultaneously estimating misclassification rates of gestational hypertension.

Estimating Onset Time From Longitudinal Data in the Presence of Measurement Error With Application to Estimating Gestational Age From Maternal Anthropometry During Pregnancy

Ana Ortega-Villa, Biogen

Accurate assessment of gestational age at birth is necessary for optimal pediatric care. In high resource countries, several methods using ultrasound have been proposed to assess gestational age at birth; however, these methods are not easily accessible for low-resource populations. We develop a shared random parameter model for estimating gestational age at birth from longitudinal maternal anthropometry that incorporates additional maternal information from the last menstrual period, a measure of gestational age collected with sizable measurement error. The proposed methodology is evaluated using simulation studies under a training-test set paradigm. In addition, we propose methodology to validate prediction when some measurements of the gold standard are collected with measurement error. We illustrate the proposed methodologies with data from the NICHD Fetal Growth Studies.

The Misdiagnosed Mediator: Estimating the Effect of Maternal Age on Preterm Birth Risk in the Presence of Misclassified Gestational Hypertension

Kimberly Hochstedler Webb, University of Pittsburgh

The risk of preterm birth increases with maternal age, and it is possible that hypertensive disorders, like gestational hypertension, mediate this maternal age-preterm birth relationship. Previous studies, however, have found low diagnostic accuracy of gestational hypertension. Thus, any mediation analysis conducted with this potentially misclassified binary mediator variable may be severely biased. This bias is especially challenging to address when the misclassification is covariate-dependent and when no gold standard measures are available. In this study, we develop methods to handle misclassification in the gestational hypertension mediator variable by modelling misdiagnosis based on patient-level factors. We present an expectation-maximization algorithm to estimate the



Session Info

S217

CAREER OR PROFESSIONAL DEVELOPMENT SESSION

October 14th, 22:30 - 22:45 UTC

Thriving with Data: Harnessing Machine Learning Techniques to Transform Global Cancer Research

USE SCHED TO ACCESS THE ZOOM ROOM

People have unique experiences in their fight against cancer, these journeys often involve uncertainty. Treatment is most effective when cancer is caught in its earliest stages. However, data can be leveraged to develop models specifically tailored to address this concern. Cancer is a global challenge, underscoring a need to develop statistical techniques and models that can address early detection, screening and treatment. This session explores global data and research on different types of cancer to better understand risk factors, how they vary across ages and gender, how they differ worldwide, and how they have evolved over time. By using machine learning techniques, Country-level cancer data will be analyzed to identify emerging trends, predict future burdens, and highlight regions where survival rates and disease severity are improving despite high environmental and lifestyle risk factors or costly treatment. The session will also use statistical and predictive modeling to simulate key factors influencing cancer diagnosis, treatment, and survival, offering valuable insights for researchers, policymakers, and healthcare providers. By uncovering country-specific cancer risks and effective interventions, this research will show how data-driven strategies can help both individuals and countries adapt and overcome challenges. These strategies will enable them not just to survive, but to thrive in their unique environments.

ORGANIZER: Christianah Olanrewaju, Brooks Insights, Abuja**CHAIR:** Christianah Olanrewaju, Brooks Insights, Abuja

Speaker Bios

S217

**CHRISTIANAH OLANREWAJU**

Brooks Insights, Abuja

<https://www.linkedin.com/in/christianafolashadeolanrewaju>

Christianah Olanrewaju is a First-Class Graduate of Statistics from Ladoke Akintola University of Technology, Nigeria. She was recognized as the Best Female Graduating Student in her department. She has worked as a Teaching Assistant at Bowen University, Nigeria, and she is currently a Data Analyst at Brooks Insights, FCT Abuja, Nigeria. She is the newly elected CWS Program Committee representative for Africa. She founded "The Beauty of Statistics (TBS)" to promote the relevance of statistics across sectors. Her research interest includes Biostatistics, Survival Analysis, Health Data Science and Machine Learning.



Session Info

S218

TECHNICAL STATISTICAL SESSION

October 14th, 22:45 - 23:00 UTC

Speed Session 3

USE SCHED TO ACCESS THE ZOOM ROOM

This session is a short session featuring three speed talks. We hope that you will contact the speakers if you are interested in discussing their work further.

ORGANIZER: Cynthia Bland, CWS

CHAIR: Jessica Kohlschmidt, CWS



Speaker Bios

S218



MS. FARAH AHMED

Equitable Advisors

I am a Senior Analytics & Insights Leader with over a decade of experience delivering data-driven solutions across Fortune 500 companies and federal clients. I hold a Master's in Data Analytics from The Ohio State University and multiple industry credentials, including PMP, CSM, CSPO. My expertise spans predictive analytics, machine learning, and process optimization, with a track record of bridging technical analysis and strategic decision-making. I have led high-impact projects in financial services, healthcare, and public sector environments—developing dashboards, predictive models, and survey-based insights that drive measurable business outcomes. I was honored to receive a travel award at the American Statistical Association's Women in Statistics and Data Science Conference (2023) and also participated in McKinsey's Asian Management Accelerator (2024) program. I am passionate about translating complex analytics into actionable strategies that improve efficiency.



MRS. OLUWAFUNMILAYO DAWODU

PIPGES UFScar-USP

Oluwafunmilayo is a PhD student in the Interinstitutional Graduate Program at the Federal University of São Carlos and the University of São Paulo, Brazil (PIPGES UFScar-USP). She holds a B.Tech in Statistics from the Federal University of Technology, Akure (FUTA), Ondo State, Nigeria, and an M.Sc. in Mathematical Sciences from the African Institute for Mathematical Sciences (AIMS), Ghana. She has worked with several organisations, including the National Population Commission and the World Health Organisation (Akure Chapter), among others. Currently, she is a member of the International Statistical Institute (ISI) and the Institute of Mathematical Statistics (IMS), and a research member of the Biostatistics and Spatial Modelling Research Group.



DR. ANNA KYE

University of California Irvine

Anna Kye earned her Ph.D. in Research Methodology from Loyola University Chicago, specializing in quantitative methods, educational measurement, and predictive modeling. She also holds a B.A. in Statistics from Iowa State University, an M.S. in Statistics from the University of Massachusetts, and an M.A. in Mathematics Education from the University of Iowa. Her research focuses on improving higher education through data-driven approaches that deepen understanding of student outcomes and inform strategies for institutional improvement. She applies statistical modeling and machine learning techniques to examine issues such as college access, transfer pathways from community colleges to four-year institutions, and disparities in academic achievement. Her work bridges research and practice by contributing to the development of tools and programs that support institutional effectiveness, faculty development, and equitable student success.



Thriving in Direct Marketing: Predicting Term Deposit Subscriptions with Ensemble Machine Learning

Farah Ahmed, Equitable Advisors

In today's competitive marketing environment, thriving means making smarter, more transparent decisions with data. Using a Portuguese bank's campaign dataset, this session demonstrates how ensemble machine learning can both predict which clients are likely to subscribe to term deposits and explain the factors driving those decisions. By combining open data, reproducible methods, and community-driven practices, this session will show how data science enables organizations — and practitioners themselves — to grow and succeed in their environments.

Predictive goal: Build an accurate and robust ensemble model to identify clients most likely to subscribe to a term deposit, enabling marketing teams to focus efforts where they are most impactful.

Interpretability goal: Translate machine learning outputs into actionable marketing insights by uncovering the key demographic, financial, and behavioral factors driving subscription decisions.

Through exploratory analysis, stacking models, and interpretability tools, I show how the combination of technical rigor and transparent analytics empowers decision-makers to thrive in their environment.

Resilience Measures for the Surrogate Paradox

Bayesian Spatial Quantile Regression for Countable Responses: An Application to Avian Influenza Viruses

Oluwafunmilayo Dawodu, PIPGES UFSar-USP

A considerable body of research has focused on the H5N1 strain of highly pathogenic avian influenza (HPAI) virus affecting wild birds and marine species, including penguins, cetaceans, mustelids, pinnipeds, and sea turtles. However, a comprehensive methodological framework for evaluating the effects of covariates across the entire distribution of count responses remains underexplored. A major challenge with this approach is the count nature of prevalence; thus, classical quantile regression is unsuitable. Hence, we adopted a jittered transformation technique to unravel the spatial patterns related to the zoonotic dissemination.

The presentation explores the application of the asymmetric Laplace distribution (ALD) within a Bayesian quantile regression framework, incorporating fixed effects, random effects, and spatial components at each quantile level. This approach improves robustness to outliers and enables the detection of under- and over-estimations in the response distribution. Although the observed count outcomes are assumed to follow a Poisson distribution, the model adopts the more flexible ALD as a working likelihood, situating the method within the class of generalised linear mixed models (GLMMs). Prior distributions are specified for all model parameters to account for uncertainty, and Markov chain Monte Carlo (MCMC) sampling is performed using Hamiltonian Monte Carlo in RStan until convergence is achieved.

Empirical Analysis of STEM Faculty Productivity: Using NbClust and Logistic Regression to Explore Interactions Among Faculty

Teaching and Research Productivity Metrics, Demographic, and Disciplinary Characteristics

Anna Kye, University of California Irvine

This study investigates the nexus between research and teaching productivity among STEM faculty at a public research-intensive university, analyzing data from 553 faculty members across four STEM disciplines: Biological Sciences, Engineering, Information and Computer Sciences, and Physical Sciences. Through the combined application of cluster analysis using the NbClust package and logistic regression, the research explores the correlation between productivity metrics and faculty demographics, including position type, rank, gender, and discipline. The analysis reveals distinct productivity clusters characterized by varying levels of research and teaching productivity outcomes across demographic groups, underscoring significant disparities. The findings emphasize the imperative for institutional policies that holistically support both teaching and research to foster faculty success. By offering a nuanced understanding of faculty productivity profiles, this study informs strategies for equitable resource allocation, faculty development, and evaluation, ultimately contributing to the advancement of STEM education and the fulfillment of institutional missions.



Session Info

S301

TECHNICAL STATISTICAL SESSION

October 14th, 9:00 - 9:45 UTC

New Advances in Relational Event Models

USE SCHED TO ACCESS THE ZOOM ROOM

Relational Event Models (REM) are widely popular for analysing streams of events such as messages and online interactions. They are particularly useful to analyse social networks and understand the dynamics of communication over time.

This session features three researchers who have advanced the set of methodologies for REM. Martina Boschi has developed a new subsampling algorithm for reducing the computational cost of estimating an REM and scaling up their application to large event datasets. Federica Bianchi has developed a statistical test for detecting changes in the dynamics of relational network and selecting relevant features related to those changes. Viviana Amati has developed a procedure for jointly analysing and comparing multiple event sequences from different sources, a problem often arising in meta-analysis.

ORGANIZER: Deborah Sulem, Università della Svizzera italiana

CHAIR: Deborah Sulem, Università della Svizzera italiana



Speaker Bios

S301

**MRS. MARTINA BOSCHI**

Università della Svizzera italiana

<https://search.usi.ch/en/people/e76b502ae6d763925d17c2293e3b53fe/boschi-martina>

Martina Boschi is a third-year PhD candidate in Computational Science at Università della Svizzera italiana (USI). Her research focuses on network science, particularly on the statistical modeling of time-stamped relational data. She serves as a teaching assistant and contributes to outreach for USI's Data Science Bachelor program. Martina holds BSc and MSc degrees in Statistical Sciences from the University of Bologna. Her academic journey has been supported by several grants, including Erasmus+ at Université Paris II, the Swiss European Mobility Programme at USI, and the NGI Enrichers Fellowship at McGill University.

**DR. FEDERICA BIANCHI**

Università della Svizzera italiana

<https://search.usi.ch/en/people/9c95575b7ab95d437d74b925bfe8fe0d/bianchi-federica>

Federica Bianchi is a researcher at the University of Cambridge (UK) and the Università della Svizzera italiana in Lugano, Switzerland (CH). Her research interests include statistical network models and methods for the analysis of markets and organizations.

**PROF. VIVIANA AMATI**

University of Milano-Bicocca

<https://www.unimib.it/viviana-amati>

Viviana Amati is an associate professor of social statistics in the Department of Statistics and Quantitative Methods at the University of Milano-Bicocca (Italy). Before joining the University of Milano-Bicocca, she was a postdoctoral researcher in the Department of Humanities, Social and Political Sciences at ETH Zurich and in the Department of Computer Science at the University of Konstanz. She received a Ph.D. in Statistics from the University of Milano-Bicocca. Her research focuses on developing statistical methods for analyzing network data, with a particular emphasis on network models, and their application across various fields, including social and organization sciences.



Scaling Relational Event Models via Graph Sampling Methods*Martina Boschi, Università della Svizzera italiana*

For nearly twenty years, Relational Event Models (REMs) have been a robust framework for analyzing streams of time-stamped events. These models explain the dynamics of underlying relational processes using statistics calculated from previously observed events. However, as temporal networks grow in size and complexity, REMs face serious computational challenges. Although several inference strategies have been developed to reduce the computational cost of estimation, there has been limited progress in optimizing the calculation of history-dependent statistics.

In this work, we introduce several estimators for network statistics based on sampled subsets of past events. By formally deriving their theoretical properties, we can measure their measurement errors and apply corrections during the estimation process. In particular, we fit REMs using error-in-variables methods, including simulation-extrapolation techniques. We test our approach on both synthetic and real-world data. Our method allows REMs to be estimated on much larger datasets than before, increasing the practical use of these models.

entities.

We propose procedures based on descriptive statistics and random coefficient models for the joint analysis of multiple independent relational event sequences. We also evaluate how well the observed event dynamics generalize across various contexts and groups. Our examination includes various methodological approaches, and we utilize simulations to identify situations where meta-analysis and random coefficient models are more effective. To demonstrate the application of the proposed procedures, we analyze multiple sequences of relational events collected in different social settings.

Relational Event Models with Polymorphic Network Times*Federica Bianchi, University of Cambridge*

Relational Event Models (REM) describe how an interaction order may emerge from complex temporal dependencies in sequences of social interaction events. Most empirical applications assume that the social processes that undergird this order remain constant and operate at uniform velocity. This paper introduces a novel approach that transforms these assumptions into testable hypotheses about how network mechanisms evolve and change over time. By employing a generalized additive inference framework (GAM) for relational event modeling, we estimate smooth, non-linear event rate functions that capture the dynamic influence of relevant network mechanisms on sequences of relational events. This new specification provides a more detailed and accurate account of how polymorphic temporal factors interact to shape network evolution. We illustrate the empirical value of our proposal in an analysis of high-frequency trading data in a major online financial market. Our results show that key network mechanisms, such as direct and generalized reciprocity, operate on different time scales and evolve at changing velocities. These new models expand the analytical possibilities of REM by redefining time as an intrinsic dimension of social structure.

Random Coefficient Models for Relational Event Data*Viviana Amati, University of Milano-Bicocca*

Traditionally, the analysis of relational event data has focused on case studies that examine a single sequence of relational events. In this study, we shift our focus to the multilevel analysis of relational events, where the data consists of multiple sequences that are conceptually similar but collected from distinct groups of entities, with no interactions occurring between these groups. These sequences can be viewed as independent replications of the same event dynamics, reflecting the underlying social processes that generated the events. Such data often arises from relational event sequences collected across different geographical regions or from repeated studies involving separate groups of



Session Info

S302

CAREER OR PROFESSIONAL DEVELOPMENT SESSION

October 14th, 9:45 - 10:30 UTC

Women in Data Science Driving Industries in the African Context

USE SCHED TO ACCESS THE ZOOM ROOM

African women in data science are transforming industries, agriculture and reshaping the future of innovation in various businesses. This session focuses on how African women plays a pivotal in the use and application of Statistics and data science to solve real-world challenges in agriculture, healthcare, finance, education, and governance within an African Context. Although African women in data science make up less than 20% of data science professionals as of 2025, women are leading groundbreaking initiatives that are both locally relevant and globally significant. Participants will explore inspiring case studies, learn from leading female data scientists, and engage in dialogue on how to overcome systemic barriers to inclusion. The session will also spotlight programs and partnerships that are equipping women with the skills and leadership tools needed to thrive in this dynamic field.

ORGANIZER: Chipso Zidana, Botswana International University of Science and Technology

CHAIR: Chipso Zidana, Botswana International University of Science and Technology



Speaker Bios

S302



MS. MUSA MBEDZI

Croco Motors Zimbabwe

Musa Mbedzi is a Senior Data Scientist with more than 10 years' experience in the Data Science and Analytics space having worked as a Senior Data Scientist, Data Analyst, Business Analyst and an Analyst, in the Telecoms, Mobile Money, Banking, Engineering, Waste Management, Motor Industry, Retail, Insurance and Consultancy spaces. She holds an MSc in Data Science and Analytics (Brunel University, London), MBA (NUST) and a BSc in Maths & Statistics(UZ), an Executive Certificate in Project Monitoring and Evaluation(UZ), among other certifications. Musa has worked on several projects including stock markets for the New York Stock exchange, property analysis for investors in the London market, customer data for predicting churn and cross sell opportunities in the South African market. She received the Chevening Scholarship (a UK government's international scholarships programme, funded by the Foreign, Commonwealth and Development Office and partner organisations) in 2019.



MS. SIMISANI NDABA

University of Botswana

Simisani Ndaba is a part-time Spatial Data Scientist Researcher in the Department of Engineering and Technology and serves as a Teaching Assistant in the Department of Computer Science at the University of Botswana. She is currently pursuing a PhD at the Botswana International University of Science and Technology, with her research focused on bio-inspired approaches to optimization challenges. Simisani is an active member of the R-Ladies Global community and the founder of the R-Ladies Gaborone chapter



MS. PENDO MANJELE

Ubuntu AI

Pendo Manjele is a data science professional in Zambia's financial sector, with a background in Computer Science and Applied Statistics and a career rooted in technology. Over the past seven years, she has worked across banking, telecoms, and a Big 4 audit firm, earning recognition including the Genius Woman in Tech Award (2021), Push Woman in ICT Award (2023), and TechTrends Rising Star Award (2024). Pendo is also a TEDx Lusaka speaker and co-founder of Ubuntu AI, a community initiative promoting inclusive and accessible AI in Zambia. She is passionate about AI, its ethical implications, and advancing inclusion in technology.



Amplifying Voices: African Women in R-Ladies Rotating Curation

Simisani Ndaba, University of Botswana

This talk explores how African women in data science have used the R-Ladies rotating curation (rocur) platform to share their work, highlight impactful projects, and inspire audiences worldwide. Through curated weeks of storytelling, technical insights, and interactive engagement, these curators have showcased not only their successes but also the challenges they have faced and the strategies they used to overcome them—with valuable support and mentorship from more experienced women in the community. This aligns closely with the mission of The African Carpentries, which promotes inclusivity, accessibility, and skill-focused education in data science across the continent.

Rocur complements this mission by creating an open, real-time, peer-to-peer learning space where practitioners exchange tools, methods, and experiences. The talk will present success stories from the rocur program, highlight how it fosters mentorship and collaboration, and demonstrate how it advances the learning and inclusion goals championed by R-Ladies and The African Carpentries.

What is the state of women's participation in data-driven technologies in Zambia? Insights from the State of AI in Zambia Survey

Pendo Manjele, Ubuntu AI

In 2024, Ubuntu AI Community, Bongohive, and ZICTA conducted a national survey to assess the state of artificial intelligence in Zambia. The findings revealed significant gender-specific trends, highlighting both gaps and opportunities for women's participation in data-driven technologies. These insights point to an urgent need for inclusive design in emerging technologies and proactive measures to address the risks of non-inclusive systems. In this talk, I will present key gender-related findings from the survey—insights that may also resonate with other ecosystems and regions. I will outline practical steps to increase women's participation in AI and data-driven fields, explain why inclusivity is essential, and explore the critical roles stakeholders must play to drive meaningful change.



Session Info

S303

TECHNICAL STATISTICAL SESSION

October 14th, 10:30 - 10:45 UTC

Enhancing Mental Workload Classification from N-Back Task Data

USE SCHED TO ACCESS THE ZOOM ROOM

This study investigates the classification of mental workload levels using data from an n-back cognitive task, employing deep learning approaches that can accommodate both dense and sparse features. The n-back paradigm captures rich temporal information on working memory and cognitive demand from the same participants across multiple conditions. The models incorporate dense features, consisting of hemodynamic measures and experimental variables, together with sparse features that represent individual subjects. The extreme Deep Factorization Machine (xDeepFM) model, applied with stratified 5-fold cross-validation, was identified as the most effective approach. For the 0- versus 1-back classification task, xDeepFM achieved improvements over the baseline model, with increases of 67.50% in accuracy, 68.74% in sensitivity, 66.24% in specificity, and 68.48% in F1-score. Applying Principal Component Analysis prior to modeling led to further gains compared to a baseline logistic regression model, resulting in increases of 53.03% in accuracy, 98.03% in sensitivity, 24.14% in specificity, and 70.37% in F1-score for the same classification task. These findings demonstrate the potential of deep learning, particularly approaches that integrate dense and sparse features, to enhance workload assessment in cognitive neuroscience and to provide valuable insights into cognitive functioning.

ORGANIZER: Serenay Cakar, Middle East Technical University**CHAIR:** Serenay Cakar, Middle East Technical University

Speaker Bios

S303



MISS SERENAY ÇAKAR

Research and Teaching Assistant

Serenay Çakar is a fourth-year Ph.D. candidate in Statistics at Middle East Technical University (METU), where she also earned her Bachelor's and Master's degrees, graduating at the top of her class. Her Master's thesis on longitudinal data analysis in cognitive science, integrating statistical and machine learning methods, received the METU Graduate Thesis Award. As a research and teaching assistant, she has taught Applied Statistics, Linear Models, and Statistical Computing, combining theory with practical programming in R and Unix. Her research focuses on applying advanced statistical and deep learning approaches to cognitive neuroscience data, particularly analyzing hemodynamic responses in n-back experiments. She has expertise in MATLAB for data preprocessing and in implementing statistical and machine learning algorithms in R, with a commitment to advancing data-driven solutions in neuroscience and related fields.

Session Info

S304

TECHNICAL STATISTICAL SESSION

October 14th, 10:45 - 11:00 UTC

An Overview of Record Linkage and Examples in Probabilistic Matching in Public Health

USE SCHED TO ACCESS THE ZOOM ROOM

This presentation explores the evolution and modern applications of record linkage, also known as matching. Record linkage combines data from various sources that refer to the same individual or entity. In healthcare, it may link patient records across providers to better track patient history or improve the understanding of disease trends for public health departments integrating surveillance data with external data sources. Advances in computer systems and programming have transformed linking systems throughout the past century to more effectively parse and evaluate thousands of records at once. Modern algorithms implement the following computerized methods: deterministic and probabilistic matching. Deterministic matching requires records to match one-to-one, which is ideal for unique identifiers like SSN. However, in the context of public health, overly conservative methods may lead to missed true matches, especially when given poor quality data. Probabilistic, or “fuzzy”, matching addresses these small mismatches using statistical methods to calculate scores based on similarity and assign greater weight to more reliable variables. In this talk, I will explain probabilistic matching in practice, exploring how probability weights are assigned, why records succeed or fail to link and software tools that are used to facilitate the process.

ORGANIZER: Sejal Padmonkar, RTI International**CHAIR:** Sejal Padmonkar, RTI International

Speaker Bios

S304



MS. SEJAL PADMONKAR

RTI International

I am a statistician at RTI International and earned my bachelor's degree in Biostatistics from the University of North Carolina at Chapel Hill. At RTI, I work on large-scale federal surveys in healthcare and education. In undergrad at the Gillings School of Public Health, my research interests lay in the intersection of air quality and environmental justice, where I led efforts to analyze particulate matter data we collected in rural North Carolina. I hope to soon pursue graduate education in the field of biostatistics to continue to expand my theoretical knowledge base and learn more about statistical applications in R/python programming.



Session Info

S305

TECHNICAL STATISTICAL SESSION

October 14th, 11:00 - 11:15 UTC

The Critical Role of XAI in Safety-Critical Applications: From Aviation to Healthcare & Digital Safety

USE SCHED TO ACCESS THE ZOOM ROOM

In safety-critical domains such as aviation, healthcare, and digital safety, opaque AI systems can create operational, ethical, and compliance risks. Explainable AI (XAI) transforms these “black boxes” into transparent, trustworthy tools that decision-makers can understand and validate.

This talk explores XAI applications across three sectors:

Aviation : Predictive maintenance, anomaly detection, and safety assurance, informed by experience in Aviation to Digital Safety & Fraud Prevention

Healthcare : Interpretable clinical decision support to improve patient safety and meet regulatory standards.

Digital Safety & Fraud Prevention : Large-scale fraud detection and abuse prevention at Microsoft, integrating XAI with Generative AI and multimodal systems.

Attendees will learn practical methods such as SHAP, LIME, causal inference, and human-AI feedback loops for building interpretable models without sacrificing performance. The session will provide a comparative framework for evaluating XAI approaches, strategies for scaling them across sectors, and guidance on aligning with global regulations.

By bridging performance with transparency, XAI enables AI systems that can be trusted when the stakes are highest.

ORGANIZER: Akshata Kishore Moharir, Microsoft

CHAIR: Akshata Kishore Moharir, Microsoft



Speaker Bios

S305

**AKSHATA KISHORE MOHARIR**

Microsoft

<https://www.linkedin.com/in/akshatak/>

I am a Lead Data Scientist at Microsoft, working in the fraud detection and Generative AI space. I specialize in Responsible AI, Explainable AI (XAI), and AI-driven fraud prevention, with experience spanning academia, industry, and government collaborations. I have worked on projects with the University of Maryland, NASA, USDA, and Boeing, applying AI to safety-critical domains such as aviation maintenance, environmental hazard prediction, and financial fraud prevention. Since 2022, I have served as an Adjunct Professor and authored multiple papers on LLMs, XAI, multimodal AI, and Responsible AI, targeting high-impact conferences such as AAAI, IJCAI, and IEEE venues. My expertise extends to real-time AI auditing, transparency frameworks, voice biometrics, and secure AI deployment in regulated sectors. I also hold seven granted patents in AI and analytics .



Session Info

S306

HISTORY OF WOMEN STATISTICIANS SESSION

October 14th, 11:15 - 11:30 UTC

Through Her Data Lens: Women and the Books That Shift the View

USE SCHED TO ACCESS THE ZOOM ROOM

Some of the most important conversations in statistics, data science, and public life have been sparked by books written by women. These works challenge bias, reveal overlooked patterns, and reframe how data is collected, interpreted, and used. This 30-minute live “book club” session will explore influential titles such as *Invisible Women* (Caroline Criado Perez), *Data Feminism* (Catherine D’Ignazio & Lauren Klein), *More Than a Glitch* (Meredith Broussard), *Everyday Sexism* (Laura Bates), and *Weapons of Math Destruction* (Cathy O’Neil).

Through short summaries, and interactive discussion prompts, we’ll connect each author’s ideas to practical ways women in statistics and data science can thrive in their own environments by advocating for better data practices, designing fairer algorithms and building more inclusive teams. The session is jointly proposed by the Celebrating Diversity and History of Statistics sections of the Royal Statistical Society.

ORGANIZER: Edith Milanzi, Femanalytica**CHAIR:** Edith Milanzi, Femanalytica

Speaker Bios

S306

**DR. EDITH MILANZI**

Femanalytica

Dr. Edith B. Milanzi is a Malawian medical statistician and data scientist. She is the founder of Femanalytica, a Malawian data science initiative dedicated to the social sector. The lab provides analytical solutions to non-profit organizations and promotes data literacy, with a particular emphasis on training women in data and technology in Malawi. Recognized as a CIPHER Growth of Leaders of Tomorrow Fellow and serving as a Royal Statistical Society Future Leaders Fellow.



Session Info

S307

TECHNICAL STATISTICAL SESSION

October 14th, 11:30 - 12:15 UTC

Statistics and Data Science in Portugal: Thriving Across Generations

USE SCHED TO ACCESS THE ZOOM ROOM

This session, organized by the Portuguese Statistical Society (SPE), highlights the diverse and impactful work of three researchers in statistics and data science, representing different stages of academic careers. Aligned with the theme Thriving in Your Environment and the session title "Statistics and Data Science in Portugal: Thriving Across Generations", it illustrates how innovation emerges through both early-career and senior contributions.

The first presentation, part of a PhD project, explores the integration of statistical and machine learning tools in dinosaur paleontology to address spatial uncertainty in the fossil record. The second presents a flexible modelling approach for estimating wildlife mortality around wind turbines, combining non-parametric and parametric techniques. The final talk introduces a new Singular Spectrum Analysis-based method for detecting structural changes in time series.

Together, these talks demonstrate the breadth and societal relevance of modern statistical methods and celebrate the strength of a research community thriving across generations.

ORGANIZER: Raquel Menezes, Centre of Mathematics, University of Minho

CHAIR: IDWSDS Committee Member



Speaker Bios

S307



CAROLINA MARQUES

University of Lisbon

Carolina Marques is a PhD student in Biostatistics and Bioinformatics at the Faculty of Sciences, University of Lisbon, supported by the Centro de Estatística e Aplicações. Her research focuses on developing innovative statistical and machine learning methods for palaeontological data analysis. She is focusing on Bayesian spatial models to explore the geographic distribution of fossil records and on developing classification tools for diverse data types, including imagery and morphometric data, which are applied to dinosaur teeth and footprints. She also develops interactive visualization tools to support reproducible research. Her work involves collaborations with researchers across disciplines (statistics, machine learning, palaeontology, and biology).



PROF. REGINA BISPO

CREEM, and NOVA

Regina Bispo has PhD in Statistics and Operational Research (specialization in Experimental Statistics and Data Analysis) from Universidade de Lisboa. She is currently a Senior Lecturer in Statistics at the School of Mathematics and Statistics, University of St Andrews. RB is an active member of MOVAMath Center for Mathematics and Applications, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa and a research member at CREEM Centre for Research into Ecological and Environmental Modelling. Her main work is related to Environmental Statistics with a focus on impact assessment studies and wildfire's modelling.



PROF. ADELAIDE FREITAS

CIDMA, and University of Aveiro

Adelaide Freitas has a bachelor's degree in Mathematics (University of Aveiro, UA), a master's degree in Probability and Statistics (University of Lisbon) and a PhD in Mathematics (UA). She is currently an Associate Professor in the Department of Mathematics at UA and serves as the scientific leader of the Probability and Statistics Group at the Center for Research & Development in Mathematics and Applications (CIDMA, UA). Her research interests include Multivariate Statistics (Dimensionality reduction, Analysis of compositional data and Analysis of other multidimensional complex data), Extreme Value Theory, and Teaching Probability and Statistics. She holds over 35 publications indexed in SCOPUS. Presented over 50 oral communications at international conferences. Served on the local organizing committee for one national and two international conferences in Statistics, as well as served as president of the scientific committee for a national congress.



Statistical and Machine Learning Tools in Dinosaur Paleontology

Carolina Marques, Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa (CEAUL)

The fossil record is inherently incomplete and spatially biased, limiting our ability to reconstruct patterns of biodiversity and paleoecology in deep time. In vertebrate paleontology, and particularly in dinosaur research, statistical and machine learning (ML) approaches are increasingly being used to overcome these challenges. We applied supervised ML techniques, such as random forests and convolutional neural networks, to tasks including the taxonomic classification of different fossil records and the prediction of fossil occurrence probabilities based on environmental variables. These models offer robust frameworks for managing high-dimensional, sparse, and noisy datasets. In addition, we used spatial point process models, including Log-Gaussian Cox Processes, to model the spatial distribution of fossil discoveries. These methods enable the identification of regions with high discovery potential and allow for the quantification of spatial uncertainty. The integration of ML and spatial statistics in vertebrate palaeontology presents opportunities for understanding macroevolutionary trends and for guiding field exploration in under-sampled regions.

This is a joint work with S Pereira, V F Santos, E Malafaia, P Mocho, J Órfão (U. Lisboa), E Dufourq (U. Northern Iowa), D Castanera (U. Zaragoza), I Díaz-Martínez (U. Cantabria), L Sciscio (U. Fribourg), M Belvedere (U. degli Studi di Firenze), and J Gracia-Cobeña (Fundación Conjunto Paleontológico de Teruel-Dinópolis)

Flexible Modelling of Carcass Density Gradients Around Wind Turbines

Regina Bispo, School of Mathematics and Statistics, University of St Andrews

The rapid global expansion of wind energy has raised increasing concerns about its environmental impacts, particularly on bird and bat populations. Post-construction fatality monitoring is a key tool for assessing collision-related mortality at wind energy facilities. However, accurate mortality estimation critically depends on accounting for spatial bias in carcass detection. Carcasses rarely fall in a uniform radial pattern around turbines. Because field surveys typically cover limited radii around turbines, failing to adjust for non-uniform spatial distribution and partial coverage can lead to substantial bias in mortality estimates. Existing approaches often rely on parametric assumptions about carcass fall distributions, which may not hold across different contexts. We present a flexible modelling approach that combines a bulk non-parametric density estimator, based on penalized splines, with a parametric tail to model carcass distances from turbines. In this talk, I will summarize main results of a simulation study that varies mortality rates, spatial distributions, detection probabilities, and truncation distances to examine how well the method recovers the true underlying distribution and under which conditions it provides the most reliable estimates. Our results highlight the importance of flexible models in addressing partial coverage in wind farm mortality estimation, offering improved tools for environmental impact assessments and wildlife conservation planning.

An SSA-Based Procedure for Exploring Structural Changes in Time Series

Adelaide Freitas, University of Aveiro – Portugal

Some procedures based on Singular Spectral Analysis focus on detecting potential structural changes in a time series by comparing a single decomposition method applied to two trajectory matrices (base and test). We propose a new approach that compares two different decompositions, both using the NIPALS algorithm, applied to the same trajectory matrix of the time series. Examples are provided to illustrate this new structural change detection procedure. This is a joint work with Alberto Silva (CIDMA, University of Aveiro).



Session Info

S308

TECHNICAL STATISTICAL SESSION

October 14th, 12:15 - 13:00 UTC

Advances from Junior Bayesian Statisticians

USE SCHED TO ACCESS THE ZOOM ROOM

The junior section of the International Society of Bayesian Statistics (j-ISBA) is excited to present a session for the International Day of Women in Statistics and Data Science. Featuring early-career women statisticians from Italy, France, and the US, this session showcases a blend of theory and applied work.

The session is built around three intertwined themes: methodologies for dynamic systems (first and second talk), applications of Bayesian methods in biology and ecology (second and third) and studies of random partition models (first and third). Indeed, the first talk will explore dynamic Bayesian mixture models for temporal data, delving into flexible frameworks for capturing changes in cluster structure and highlighting novel computational strategies for efficient inference. Along the biological and dynamic axes, the second talk will present causal discovery methods in ecological time series, an emerging area of interest that seeks to uncover causal relationships within dynamic ecological systems, discussing how these methods can be adapted to observational data with limited timestamps. Finally, the third talk will showcase how Bayesian nonparametric models can help uncover complex spatial relationships in biological systems, designing with random partitions the co-localization of immune, stromal, and tumor cell populations.

These diverse topics highlight the innovative contributions of women Bayesian statisticians to the field.

ORGANIZER: Francesca Panero, Sapienza University of Rome

CHAIR: IDWSDS Committee Member



Speaker Bios

S308

<https://beatricefranzolini.github.io/>


DR. BEATRICE FRANZOLINI

Bocconi University

Beatrice is an Assistant Professor at the Bocconi Institute for Data Science and Analytics and a core faculty member of the Bayesian Learning Laboratory at Bocconi University in Milan, Italy. Her research focuses on developing probabilistic representations for statistical modeling of complex phenomena, particularly in Bayesian nonparametrics and machine learning. She aims to deepen the understanding of probabilistic structures (dependent random measures and multivariate random partitions) while also exploring their applications to modeling dynamic systems, including change-point detection and network data analysis. Beatrice earned her PhD in Statistics from Bocconi University in 2022. She held research positions at the Agency for Science, Technology and Research and the Division of Biomedical Data Science at the National University of Singapore. She served as Chair of the Junior Section of the International Society for Bayesian Analysis and received the IMS New Researcher Travel Award.



DR. DARIA BYSTROVA

Pierre Louis Institute of Epidemiology and Public Health

Daria is a data scientist specializing in causal inference at the Pierre Louis Institute of Epidemiology and Public Health and France Cohortes. Her work focuses on methodological and applied problems like the development and application of causal inference methods in public health. She was a postdoctoral researcher at the Laboratoire d'Informatique de Grenoble, where she worked on hybrid causal discovery in time series and the identification of difference graphs between populations, with applications in ecology and health. She obtained her PhD in Applied Mathematics in 2023 from the University of Grenoble Alpes and the Inria Grenoble Rhône-Alpes research center, in the Statify team at Inria and the Laboratoire d'Écologie Alpine under the supervision of J. Arbel and W. Thuiller. There, she studied the theoretical properties of Bayesian nonparametric clustering and its application to inferring high-dimensional dependence structures among species from observed co-occurrence data.



DR. YUNSHAN DUAN

Johns Hopkins University

Yunshan Duan recently achieved a PhD in Statistics at the University of Texas Austin under the supervision of Prof. Peter Mueller. She is now a postdoctoral fellow at Johns Hopkins University, working on Bayesian nonparametrics and statistical machine learning methods with applications in biomedical applications.



Dynamic Bayesian Mixture Models*Beatrice Franzolini, Bocconi University*

This talk begins by discussing the foundational role of Bayesian mixture models in clustering and density estimation, focusing on their use in standard settings such as the analysis of cross-sectional data. Building on this foundation, the talk explores some recent advances and extensions that enable dynamic clustering in temporal data, addressing the challenges inherent in modeling evolving phenomena over time with mixture models. These developments include flexible frameworks for capturing changes in cluster structure, accommodating time-dependent data, and improving interpretability in dynamic contexts. Additionally, the talk highlights theoretical advances in extending classical Bayesian symmetry assumptions, such as exchangeability, toward more complex symmetry structures. Finally, computational strategies for efficient inference in dynamic Bayesian mixture models are discussed, including scalable algorithms that facilitate practical implementation on large datasets.

partition models for the different types of cells, with a hierarchically structured prior introducing the desired dependence. Specifically, we use Pitman-Yor priors and add dependence in the base measure for spatial features, while leaving the base measure corresponding to gene expression features a priori independent across different types of cells. Details of the model construction are designed to lead to a convenient MCMC algorithm for posterior inference. Simulation studies show favorable performance in identifying co-localization between types of cells. We apply the proposed approach with colorectal cancer (CRC) data and discover subtypes of immune and stromal cells that are spatially aligned with specific tumor regions.

Causal Discovery From Ecological Time Series With One Timestamp and Multiple Observations*Daria Bystrova, Sorbonne University*

Ecologists often aim to uncover causal relationships within ecosystems—such as species interactions, ecosystem functions, or services—using observational data. However, many studies still rely on correlation-based methods, which lack causal interpretability. Recently, causal discovery methods have gained attention, particularly for analyzing ecological time-series. Yet, the scarcity of such data remains a challenge due to the intensive effort required for long-term monitoring.

This paper explores the use of causal discovery methods on point-in-time (snapshot) observational data from dynamic ecological systems. We focus on the PC algorithm, which is theoretically valid under assumptions like the causal Markov condition, faithfulness, and causal sufficiency. We also examine the FCI algorithm, an extension that accounts for possible violations of causal sufficiency.

Using theoretical analysis and simulation experiments, we identify conditions under which these methods provide meaningful insights. Notably, we show that the PC algorithm—despite its assumptions—can still reveal certain causal relationships even when causal sufficiency is violated. Our findings are demonstrated through simulations and real-world data on bird species abundances, climate, and land-cover variables.

Spatially Aligned Random Partition Models on Spatially Resolved Transcriptomics Data*Yunshan Duan, Johns Hopkins University*

We propose spatially aligned random partition (SARP) models for clustering multiple types of experimental units, incorporating dependence in a subvector of the cluster-specific parameters, e.g., a subvector of spatial information, as in the motivating application. The approach is developed for inference about co-localization of immune, stromal, and tumor cell sub-populations. The aim is to understand the recruitment of immune and stromal cell subtypes by tumor cells, formalized as spatial dependence of the corresponding homogeneous cell subpopulations. This is achieved by constructing Bayesian nonparametric random



Session Info

S309

TECHNICAL STATISTICAL SESSION

October 14th, 13:00 - 14:00 UTC

Integrating Emerging Data Sources for Official Statistics in Latin America and the Caribbean

USE SCHED TO ACCESS THE ZOOM ROOM

This session focus on the expanding need to bring emerging data sources and tools into official statistical production. In a rapidly evolving digital environment, conventional approaches must be augmented by innovative information inputs. It emphasizes linking methodological progress with data innovations to reinforce Latin American and Caribbean official statistics, assessing how new sources and tools can improve statistical workflows and outputs. Topics include projects that integrate alternative sources such as mobile phone data, remote sensing, and web scraping to improve statistical outputs in LAC countries; approaches to enhance data collection, processing, and accessibility for climate change data production; the use of Earth Observation and artificial intelligence for crop mapping and agricultural monitoring; and the application of small area estimation techniques to facilitate data integration. The session aims to showcase statistical innovation in the scope of official statistics, foster cross-country learning and collaboration, address data quality and governance considerations, and identify practical pathways for institutions to adopt data innovations in the production of official statistics.

ORGANIZER: Maria Luíza Toledo, Brazilian Institute of Geography and Statistics

CHAIR: Maria Luíza Toledo, Brazilian Institute of Geography and Statistics



Speaker Bios

S309



MS. MARIA LUÍZA TOLEDO

IBGE - Brazilian Institute of Geography and Statistics

Maria Luíza holds a PhD in Production Engineering from the Federal University of Minas Gerais (UFMG), with a Bachelor's and Master's in Statistics also from UFMG. She is currently a Professor and Researcher in the undergraduate and graduate programs at the National School of Statistical Sciences (ENCE/IBGE). Her research areas include Probability and Statistics, with a focus on Survival Analysis and Reliability, Data Quality in Surveys, and Big Data in Public Statistics. She is the Manager of the ENCE/IBGE Research Hub, Coordinator of the CDBDIA Learning Track, and Leader of the UN Regional Hub for Big Data and Data Science in Brazil.



MS. ESTEFANÍA PIZARRO ARIAS

National Institute of Statistics of Chile

Estefanía is an Agricultural Engineer and a Master of Agricultural Sciences from the University of Chile, with a specialization in remote sensing, data science, and geomatics. She currently focuses on integrating these tools into the production of official statistics. She works as a technical analyst at the National Institute of Statistics of Chile, as part of the team responsible for developing a land use and land cover map to update the institution's area frames. She has actively promoted the use of remote sensing as a key input for the design and updating of statistical frames.



MS. ANJALI KISOENSINGH

General Bureau of Statistics in Suriname

Anjali is a statistician working at the General Bureau of Statistics since 2008, holding a Master's Degree in Research Methods and a Bachelor's degree in International Economy. She has experience in data collection and compilation environments for Statistics, including Climate Change indicators and statistics, the Sustainable Development Goals (SDGs), and Poverty statistics. Since 2016, she has been a member and co-chair of the UNSD Expert Group on Environment Statistics (EG-ECCS). Since 2017, she has been a member of the Technical Working Group of CARICOM Environment Statistics and CARICOM SDGs. As of 2022, she is a member of the VNR/SDG committee and the Poverty Committee in Suriname, and as of 2024 she is a member of the Census 2024 Committee.



MS. DENISE SILVA

Society for the Development of Scientific Research

Denise Britz do Nascimento Silva is a Principal Researcher at the Society for the Development of Scientific Research (SCIENCE). Previously, she worked for many years at the National School of Statistical Sciences (ENCE) at the Brazilian Institute of Geography and Statistics (IBGE). She completed her PhD in Statistics at the University of Southampton and has worked as a survey methodologist for over 35 years and as a lecturer at the graduate and undergraduate levels. Denise is currently ISI vice president, editor of the Statistical Journal of the IAOS, and was president of the International Association of Survey Statisticians (IASS) in 2019–2021. Her main areas of interest are survey methods, official statistics, and statistical education.



Session Info

S310

TECHNICAL STATISTICAL SESSION

October 14th, 14:00 - 15:00 UTC

Brazilian Women Thriving in Research

USE SCHED TO ACCESS THE ZOOM ROOM

This session showcases the work of three inspiring Brazilian women in academia. Among the speakers are recipients of the Brazilian L'Oréal-UNESCO Award for Women in Science, holders of the CNPq Research Productivity Fellowship, and accomplished software developers and maintainers. These scholars exemplify how to balance the challenges of being a woman in academia with excellence in teaching, service, and impactful research. The first presentation will explore Bayesian methods for estimating and selecting variables in mixtures of logistic regression models. The second will focus on Small Area Estimation for Sustainable Development Goal indicators. The third will address the challenges of staying connected to research during early motherhood.

ORGANIZER: Gabriela Bettella Cybis, Universidade Federal do Rio Grande do Sul

CHAIR: IDWSDS Committee Member



Speaker Bios

S310



DAIANE ZUANETTI

Universidade Federal de São Carlos

She holds a Bachelor's and Master's Degree in Statistics from the Federal University of São Carlos (2003) and a PhD in Statistics from the Inter-institutional Graduate Program of the Department of Statistics of the Federal University of São Carlos and the Institute of Mathematical Sciences and Computer Science of the University of São Paulo – PIPGEs (2016). She is currently an Adjunct Professor in the Department of Statistics of the Federal University of São Carlos and president of PIPGEs. She has been a member of the International Network for Research and Public Policy (RP3CD) since 2023. She has worked in the area of Probability and Statistics, with emphasis on the selection and estimation of independent or dependent mixture models (HMM), QTL mapping with familial dependence, parametric and non-parametric Bayesian inference, and MCMC computational methods with applications mainly in Genomics and Molecular Biology.



KELLY CRISTINA GONÇALVES

Universidade Federal do Rio de Janeiro

Associate Professor in the Department of Statistics at the Federal University of Rio de Janeiro (UFRJ) in Brazil. In 2020 I was a visiting researcher in the Department of Statistics at University of Florida, USA. From 2018 to 2023, I was a Visiting Researcher at Instituto de Pesquisa Econômica Aplicada (IPEA) and from 2023 to 2025 I worked as a technical consultant for a cooperation project between Instituto Brasileiro de Geografia e Estatística (IBGE) and United Nations Population Fund (UNFPA). Currently, I am a CNPq Productivity Fellow and a "Jovem Cientista do Nosso Estado" Fellow (FAPERJ). I am part of the editorial board of the Brazilian Journal of Probability and Statistics and Spatial and Spatial-temporal epidemiology and current president of the Brazilian Chapter of the International Society for Bayesian Analysis (2025 to 2027).



TAIANE SCHAEGLER PRASS

Universidade Federal do Rio Grande do Sul

She holds a degree in Mathematics (teaching major), as well as a Master's and a PhD in Mathematics, both with an emphasis on Probability and Mathematical Statistics. She has worked as a Data Scientist at StatSoft South America and is currently a professor in the Department of Statistics at the Universidade Federal do Rio Grande do Sul, as well as a permanent member of the Graduate Program in Statistics (PPGEst) at the same university. She has experience in the field of Mathematical Statistics, focusing primarily on the following topics: time series and nonlinear stochastic processes, long-memory models, and imputation of missing data in time series. She is the main developer of the R packages PPMiss, BTSR, PTSR, and DCCA, and a contributor to the PRTree package.





Bayesian Variable Selection in Mixture of Logistic Regressions With Pólya-Gamma Data Augmentation

Daiane Aparecida Zuanetti, Federal University of São Carlos

We present Bayesian methods for estimating and selecting variables in a mixture of logistic regression models. A common issue with the logistic model is its intractable likelihood, which prevents us from applying simpler Bayesian algorithms, such as Gibbs sampling, for estimating and selecting the model since there is no conjugacy for the regression coefficients. We propose to solve this problem by applying the data augmentation approach with Pólya-Gamma random variables to the logistic regression mixture model.

For selecting covariates in this model, we investigate the performance of two prior distributions for the regression coefficients. A Gibbs sampling algorithm is then applied to perform variable selection and fit the model. The conjugacy obtained for the distribution of the regression coefficients allows us to analytically calculate the marginal likelihood and gain computational efficiency in the variable selection process. The methodologies are applied to both synthetic and real data. Joint work with Mariella Ananias Bogoni.

approaches, including traditional regression trees and Probabilistic Regression Trees.

Small Area Estimation for Sustainable Development Goal indicators in Brazil

Kelly Gonçalves, Universidade Federal do Rio de Janeiro

National statistical agencies around the world have emphasized the growing need to produce reliable estimates of economic and social indicators for more disaggregated segments of the population based on survey sample data. In Brazil, for example, the country's vast territory, diversity, and inequality highlight the importance of geographically disaggregated estimates to better inform targeted public policies. In the context of Goal 1 of the 2030 Agenda for Sustainable Development, for example, which aims to eradicate poverty for all people in all its forms everywhere, reliable state level poverty estimates can be obtained from the Brazilian National Household Survey for Brazil. However, producing more detailed statistics at lower geographic levels requires the use of small area estimation methods. This work presents the development of Bayesian small area models to estimate poverty and labor force indicators for Brazilian municipality strata.

DFA and DCCA under Missing Data: From Classical Imputation to Probabilistic Regression Trees

Taiane Prass, Universidade Federal do Rio Grande do Sul

Detrended Fluctuation Analysis (DFA) and Detrended Cross-Correlation Analysis (DCCA) are widely used indirect methods for quantifying variance and cross-correlation in trend-stationary time series, particularly for detecting long-range dependence. However, their computational requirements make them unsuitable for application to time series with missing data, as they rely on complete datasets. To address this limitation, imputation techniques can be employed to estimate missing values. This study investigates the performance of DFA and DCCA in time series exhibiting short-range dependence and containing a significant proportion of missing data. A comparison is made between classical imputation methods and more advanced



Session Info



CAREER OR PROFESSIONAL DEVELOPMENT SESSION

October 14th, 15:00 - 16:00 UTC

Career Insights from Women in Statistics at MRC-BSU, University of Cambridge

USE SCHED TO ACCESS THE ZOOM ROOM

What helped you to thrive in your environment? We pose this question to four female biostatistical researchers at the MRC-Biostatistics Unit, University of Cambridge: Sofía Villar, Anne Presanis, Hélène Ruffieux, and Elena Vigorito. These panelists, representing diverse career phases and areas of application within a leading academic biostatistics setting, have been invited to this session for their lived experience navigating and excelling within this demanding environment.

Through a panel discussion chaired by Mia Tackney, we will hear from these researchers about how they have defined and achieved a "thriving" level in their careers. The discussion will cover practical topics including navigating the grant application process, building collaborations, effectively supervising students and early-career researchers, contributing to the research community and maintaining a crucial balance between work and personal responsibilities. This session aims to share invaluable insights and actionable strategies for success, fostering a supportive environment for women navigating their careers in statistics. We hope attendees will gain practical advice on how to leverage community support and identify what more is needed to thrive in their own unique statistical and data science environments.

ORGANIZER: Sofía Villar, MRC Biostatistics Unit, University of Cambridge

CHAIR: Mia Tackney, MRC Biostatistics Unit, University of Cambridge

SPONSOR: Medical Research Council Biostatistics Unit - University of Cambridge



Speaker Bios

S311



DR. SOFIA VILLAR

MRC Biostatistics Unit, University of Cambridge

Sofia Villar is a Programme Leader (Efficient Study Design Theme) at the MRC Biostatistics Unit. She is the co-lead of the MRC-NIHR Trials Methodological Partnership (TMRP) Adaptive Design Working Group. She was the recipient of the first Biometrika postdoctoral fellowship in 2013 and she is the 2025 recipient of the Frank Hansford-Miller fellowship. Her recent work focuses on methodological challenges of implementing and analyzing response-adaptive designs in which there is a high frequency of adaptation. Her applied work has resulted in the application of innovative trial designs to different areas, including rare diseases.



DR. ELENA VIGORITO

MRC Biostatistics Unit, University of Cambridge

Elena Vigorito obtained a PhD in Immunology at the University of Barcelona, Spain, and worked for several years on gene regulation on the immune system at the Babraham Institute, UK. After re-training in Epidemiology, she joined Chris Wallace's group at the MRC Biostatistics Unit to apply statistical and computational methods to further understanding causes of autoimmune diseases.



DR. ANNE PRESANIS

MRC Biostatistics Unit, University of Cambridge

Anne Presanis is a Senior Investigator Statistician in the Population Health theme at the MRC Biostatistics Unit. Her research lies at the intersection of biostatistics and infectious disease epidemiology, focusing on Bayesian evidence synthesis models to estimate prevalence, incidence and severity for pathogens such as HIV, influenza and SARS-CoV-2, from multiple observational data sources. Her research interests are in data integration, model assessment and design of disease monitoring systems, motivated by the need for and challenges of disease burden estimation to monitor infectious diseases, design and evaluate interventions to mitigate the impact of epidemics, and therefore inform healthcare policy.



DR. HÉLÈNE RUFFIEUX

MRC Biostatistics Unit, University of Cambridge

Hélène Ruffieux is a Senior Research Associate at the MRC Biostatistics Unit. She holds a PhD in Mathematics from EPFL, Switzerland. Her research interests broadly lie in the development of Bayesian methods and their application to open problems in biomedicine, with a focus on scalable hierarchical modelling approaches for variable selection, latent structure discovery and network estimation, in high-dimensional or temporal data settings.



Session Info

S312

TECHNICAL STATISTICAL SESSION

October 14th, 16:00 - 16:15 UTC

Designing Cluster Randomized Trials: A Practical Guide

USE SCHED TO ACCESS THE ZOOM ROOM

An important consideration in the design stage of randomized controlled trials is whether individuals within each site should be randomized to a specific treatment arm (an individually randomized controlled trial) or whether entire sites should be randomized to a specific treatment arm (a cluster randomized controlled trial). Recently, cluster randomized trials have grown in popularity; however, investigators have expressed a need for guidelines related to selection of the best design – especially when considering more complex designs such as stepped wedge. We address common misconceptions surrounding the appropriate use of cluster randomization and crossover designs (such as the stepped wedge design) and provide practical recommendations. Finally, the intraclass correlation coefficients for different levels of clustering are presented, and the impact of different designs on sample size discussed.

ORGANIZER: Kendra Plourde, Yale School of Public Health**CHAIR:** Kendra Plourde, Yale School of Public Health

Speaker Bios

S312

**DR. KENDRA PLOURDE**

Yale School of Public Health

<https://ysph.yale.edu/profile/kendra-plourde/>

Kendra Plourde earned her PhD in Biostatistics at Boston University in 2020, and joined the Yale School of Public Health faculty in 2022 as an Assistant Professor of Biostatistics after completing a postdoc at Yale. Her methodological research interests are in the design and analysis of observational studies and randomized controlled trials with correlated data, for example, longitudinal studies, cluster randomized and stepped wedge designs with complex sources of clustering, and clinical trials with multivariate outcomes. Her collaborative research interests are in studies of Alzheimer's disease and related dementias and aging. Kendra is also a member of the Yale Center for Analytical Sciences and a collaborator with the Yale Program on Aging.



Session Info

S313

TECHNICAL STATISTICAL SESSION

October 14th, 16:15 - 17:00 UTC

Data, Ethics, and Statistical Reasoning: Theoretical Insights and Practical Applications

USE SCHED TO ACCESS THE ZOOM ROOM

As the scale and scope of data collection grow exponentially and as statistical tools become increasingly embedded in decision-making systems across disciplines, the responsibilities of statisticians have never been more critical. This session will explore the ethical dimensions of statistical work from both a theoretical and application perspective. From biased algorithms to privacy breaches and misleading interpretations, the implications of statistical decision-making affect not only scientific integrity but real-world outcomes in healthcare, criminal justice, education, and more. The session will include 3 presentations, each approx. 17 mins. in length + 3 mins. for Q&A. The first talk will discuss the implications of a threshold in establishing statistical significance in hypothesis tests and suggest a pathway to a more transparent statistical approach. The second speaker will discuss semiparametric model fitting method that allows unbiased assessment of model improvement without requiring a representative sample from the target population, addressing a key practical challenge in data analytic applications. The third presentation will showcase statistical analysis of health and fitness data collected by electronic fitness trackers, highlighting the need to address ethical issues concerning privacy and security in such data. It will provide insights for enhancing students' health and well-being through the responsible use of these devices.

ORGANIZER: Sinjini Mitra, California State University, Fullerton**CHAIR:** Sinjini Mitra, California State University, Fullerton

Speaker Bios

S313



DR. NICOLE LAZAR

Pennsylvania State University

Dr. Nicole Lazar is Professor and Head of the Department of Statistics at Penn State University and affiliated faculty with the Huck Institutes of the Life Sciences at Penn State. Dr. Lazar received her PhD in statistics from the University of Chicago, her MS in statistics from Stanford University, and her BA in statistics and psychology from Tel Aviv University. She is a Fellow of the ASA and the IMS, an Elected Member of the ISI, served as President of the Caucus for Women in Statistics in 2019, and as Editor-in-Chief of The American Statistician in 2014-2016. Her research interests include likelihood theory, the statistical analysis of functional neuroimaging data, topological data analysis, and the foundations of statistics. In recent years, Prof. Lazar has been involved in the statistical reform movement and become an advocate for Open Scholarship practices.



DR. JINBO CHEN

University of Pennsylvania Perelman School of Medicine

Dr. Jinbo Chen is a Professor of Biostatistics and Director of the Statistical Center for Translational Research in Medicine in the Department of Biostatistics, Epidemiology & Informatics at the University of Pennsylvania Perelman School of Medicine. Her research focuses on statistical methods for risk modeling, innovative analysis of electronic health record (EHR) data, and designing efficient sampling schemes under resource constraints. Dr. Chen's recent work has concentrated on practically important analytical challenges in risk modeling using EHR data, including robust methods for addressing EHR phenotyping inaccuracy, assessing and improving the fairness of predictive algorithms, enhancing risk modeling through data integration, and inferring variable importance. Her methodologies have been applied to advance research across a wide range of scientific fields, including breast cancer risk prediction, health services research, and cardiovascular health research.



DR. SINJINI MITRA

California State University, Fullerton

Dr. Sinjini Mitra is Professor in the Information Systems & Decision Sciences (ISDS) department at the College of Business and Economics, California State University, Fullerton (CSUF). Previously, she was a postdoctoral research fellow at USC's Information Sciences Institute. Dr. Mitra earned her Ph.D. in Statistics from Carnegie Mellon University. She is a faculty fellow of CSUF's Catalyst Center, Center for Information Technology and Business Analytics, and Cybersecurity Center. Her research interests encompass statistical applications in security, healthcare, business, and education, emphasizing interdisciplinary settings. Dr. Mitra has authored two books, several journal articles, and book chapters, and served as a Co-PI on two NIH collaborative grants focused on student mentoring and research. She also serves as the Director of Undergraduate Assessment at her college. Dr. Mitra is also affiliated with several professional organizations, including IMS, ASA, and INFORMS.



The Ethical Problems of a Threshold

Nicole Lazar, Pennsylvania State University

The use of "statistical significance" thresholds is ubiquitous in science, but such use raises a key question: How should one treat values that don't quite attain the threshold? Given current incentives in publishing and grantsmanship, by which only "statistically significant" results are deemed worthy of attention, the ethical dilemma is clear. The issue is not restricted to p-values and their cutoffs; the same problem arises, for example, with thresholds on effect sizes or Bayes factors. Inherently, of course, there is no real difference between, say, a p-value of 0.049 and a p-value of 0.051, yet in the presence of the common (artificial) "significance threshold" of 0.05, a researcher either needs to treat them differently or resort to convoluted and statistically meaningless terminology such as "trending towards significance." A third, ethically problematic, path is to engage in a variety of so-called questionable research practices (QRPs). QRPs may appear to many researchers as valid analysis strategies, but in fact they - along with other practices such as p-hacking - distort findings. These spurious results in turn will often fail to replicate. In this talk, I will discuss these implications of a threshold and suggest a pathway to a more transparent statistical approach.

factors influencing tracker usage. Multiple linear regression models identified determinants affecting tracker use, while logistic regression assessed factors driving sustained use. The study revealed that positive perceptions of health benefits encouraged tracker use, but privacy and data security concerns acted as deterrents. Additionally, we analyze how the COVID-19 pandemic influenced fitness tracker usage and perceptions of their utility for health management before, during, and after the pandemic. These findings underscore the importance of addressing ethical concerns related to privacy and security in health data, offering insights for improving students' health and well-being through responsible use of fitness trackers.

Toward Well-Calibrated Risk Estimation with Biased Training Data

Jinbo Chen, University of Pennsylvania Perelman School of Medicine

The added value of candidate predictors for risk modelling is routinely evaluated by comparing the performance of models with or without including candidate predictors. Such comparison is most meaningful when the estimated risk is unbiased in the target population. Oftentimes, data for standard predictors in the base model is richly available from the target population, but data for candidate predictors are available only from nonrepresentative convenience samples. While the base model can be naively updated using the study data without recognizing the discrepancy between the underlying distribution of the study data and that in the target population, the resultant risk estimates and the evaluation of the candidate predictors are biased. We proposed a semiparametric method for model fitting that enables unbiased assessment of model improvement without requiring a representative sample from the target population, thereby overcoming a major bottleneck in practice. I will discuss how a data analysis project inspired this methodological effort, leading to a novel approach tailored to practical needs. I will describe how this method underpinned a recently well-scored scientific grant proposal, demonstrating how a novel statistical methodology can drive and enable innovative scientific endeavors.

Ethical Aspects in Analyzing Health and Fitness Tracker Usage Data Among College Students

Sinjini Mitra, California State University, Fullerton

This case study examines the ethical considerations surrounding college students' use of fitness trackers, focusing on privacy and security concerns associated with health and fitness data. An online survey of 86 students at a U.S. public university provided data for analysis. Descriptive statistics and statistical hypothesis tests were employed to explore associations between various



Session Info

S314

OUTREACH EVENTS

October 14th, 18:00 - 19:00 UTC

Florence Nightingale Day: Lessons in Longevity

USE SCHED TO ACCESS THE ZOOM ROOM

Florence Nightingale Day (FND) is an international outreach initiative that encourages girls and underrepresented students to explore statistics and data science (<https://fndaystats.org/>). In Canada, CANSSI has been central to bringing FND to life, supporting and coordinating events across the country (<https://canssi.ca/program/florence-nightingale-day/>). While running a single event is challenging, creating a sustainable model that works year after year, and at multiple institutions, is even more complex.

This panel brings together organizers from both Canada and the United States to share experiences and reflect on what it takes to keep FND going over the long term. We'll discuss practical challenges like recruiting and engaging students, securing institutional support, and managing the workload for organizers—especially as events grow or when new hosts take them on for the first time. We'll also look at how events can be adapted to different institutional contexts and student populations while staying connected to the international goals of FND.

Rather than focusing on solutions, the panel will highlight key challenges and lessons learned to help others who are planning or running FND. The goal is to open up a conversation about what it takes to sustain this kind of initiative and to recognize the important role of CANSSI in supporting and strengthening the network of FND events internationally.

ORGANIZER: Tatiana Krikella, York University**CHAIR:** Tatiana Krikella, York University**SPONSOR:** CANSSI

Speaker Bios

S314



DR. JOANNA MILLS FLEMMING

Dalhousie University

Dr. Joanna Mills Flemming is a Professor in the Department of Mathematics and Statistics, and Acting Dean of Graduate and Global Relations in the Faculty of Science, at Dalhousie University. Her research interests center on the development of statistical methodologies for data exhibiting spatial and/or temporal dependencies, with particular interests in marine animal movement and fisheries science. Over the last decade, she has advanced the supporting statistical theory and inference tools for stock assessments in Canada. She currently serves as Regional Director of CANSSI Atlantic, and as an Associate Editor for the Canadian Journal of Statistics, Movement Ecology, and Environmetrics.



DR. SAMANTHA-JO CAETANO

University of Toronto

Samantha-Jo (Sam) Caetano is an Assistant Professor, Teaching Stream, in the Department of Statistical Sciences at the University of Toronto. She earned her PhD in Statistics from McMaster University. Sam has been widely recognized for her commitment to teaching and engagement, including receiving the 2024 Early Career Educator Award from the Statistical Society of Canada. Her teaching focuses on making statistics accessible and engaging, emphasizing communication, collaboration, and real-world applications. She has introduced several initiatives to foster student connection and enthusiasm for statistics, including the Data Science Café and DataFest. She also led the launch of Florence Nightingale Day at U of T, a celebration aimed at highlighting the role of women and diversity in data science. Beyond university teaching, Samantha-Jo is passionate about statistical outreach and literacy, actively working with high school and elementary students to spark interest in data.



DR. WEI (BECKY) LIN

Simon Fraser University

Wei (Becky) Lin is a Lecturer in the Department of Statistics and Actuarial Science at Simon Fraser University (SFU). She earned her PhD in Statistics from the University of Toronto and has held academic and research roles at institutions including U of T, the University of Southern California, and Thompson Rivers University. Her teaching is informed by diverse experience, including work as a mathematical statistician at Statistics Canada and co-founding an AI startup. At SFU, Becky is known for her dynamic teaching and commitment to student engagement. She also leads the annual Florence Nightingale Day at SFU, a community outreach event aimed at inspiring high school students — especially those from underrepresented groups—to explore careers in statistics and data science. Passionate about broadening participation in STEM, Becky champions initiatives that promote statistical literacy and spark curiosity about data.



DR. SHILI LIN

Ohio State University

Dr. Shili Lin is a Professor in the Department of Statistics at The Ohio State University. Her research centers on developing and applying statistical methodologies to analyze genomic data from both population- and family-based studies. Before joining Ohio State, Dr. Lin served as a Neyman Visiting Assistant Professor in the Department of Statistics at the University of California, Berkeley. She is an active contributor to the statistical community through editorial service for several journals, participation on NIH study sections, and leadership roles in professional organizations. Dr. Lin has been instrumental in organizing Florence Nightingale Day events, which encourage middle and high school students to explore careers in statistics and data science. She is also a Fellow of the American Statistical Association, the Institute of Mathematical Statistics, the American Association for the Advancement of Science, and an Elected Member of the International Statistical Institute.



Session Info

S315

TECHNICAL STATISTICAL SESSION

October 14th, 19:00 - 20:00 UTC

Statistical Methods for the Validation of Surrogate Markers in Clinical Trials

USE SCHED TO ACCESS THE ZOOM ROOM

Surrogate markers—biomarkers or intermediate outcomes that substitute for a clinical endpoint—can accelerate clinical trials and reduce costs, but their use demands careful statistical validation to ensure accurate and reliable inference. This session brings together researchers developing innovative methods for evaluating and applying surrogate markers in clinical trials. Together, these presentations will showcase cutting-edge tools to enhance the efficiency and rigor of clinical research through better use of surrogate endpoints.

ORGANIZER: Sarah Lotspeich, Wake Forest University

CHAIR: Sarah Lotspeich, Wake Forest University



Speaker Bios

S315

**MS. ANH NGUYEN**

University of Wisconsin - Madison

Anh Nguyen is a current graduate student in Statistics at University of Wisconsin – Madison. She obtained her MS in Statistics at Wake Forest University.

**MS. EMILY HSIAO**

University of Texas at Austin

Emily Hsiao is a PhD student in the Department of Statistics and Data Science at the University of Texas at Austin. Emily's research focuses on developing an empirical framework to test assumptions that ensure protection from the surrogate paradox via nonparametric testing methods, and a simulation-based approach to assess resilience to the surrogate paradox in a future study.

**DR. EMILY ROBERTS**

University of Iowa

Emily Roberts is an assistant professor of biostatistics at the University of Iowa. She is excited about statistical methodology that can be meaningfully used in a clinical setting. She completed my PhD at the Department of Biostatistics, University of Michigan in 2022 with my PhD dissertation “Causal Inference Methods and Intermediate Endpoints in Randomized Clinical Trials”. She earned my MS in biostatistics from Michigan in 2018 and graduated Summa Cum Laude in 2016 from Coe College with a BA in mathematics and psychology. Her passion for medical research and ultimate career goals stem from her diagnosis of type I diabetes at the age of four.

<https://emilykroberts.github.io/>

Correcting for Missing Data When Validating Surrogate Markers in Clinical Trials*Anh Nguyen, University of Wisconsin - Madison*

Evaluating treatment effects is critical in clinical trials but sometimes involves lengthy, invasive, or costly follow-up procedures. In these cases, surrogate markers, which provide intermediate measures of the long-term treatment effect, allow clinicians to obtain results faster and more efficiently than would have otherwise been possible. Prior to adoption, it is vital that the utility of surrogate markers (i.e., their ability to capture the treatment effect on the primary outcome) is statistically validated. Many frameworks for validating surrogate markers have been proposed, but they do not account for missing data. Instead, they rely on complete cases, which can be inefficient and biased. To improve on this, we propose methods to accommodate missing data in nonparametric and parametric surrogate validation via inverse probability weighting (IPW) and maximum likelihood estimation (MLE). Through simulations, we demonstrate how the proposed methods remain unbiased under more types of missing data than complete case analysis. Moreover, our missing data corrections offer complementary strengths, with IPW being more robust and MLE providing better efficiency.

baseline patient characteristics and consider how these may be relevant in the context of determining for whom the surrogate works well and use the proposed method within a prostate cancer clinical.

Resilience Measures for the Surrogate Paradox*Emily Hsiao, UT Austin*

Surrogate markers are often used in clinical trials to evaluate treatment effects when primary outcomes are costly, invasive, or take a long time to observe. However, reliance on surrogates can lead to the "surrogate paradox," where a treatment appears beneficial based on the surrogate but is actually harmful with respect to the primary outcome. In this paper, we propose formal measures to assess resilience against the surrogate paradox. Our setting assumes an existing study in which the surrogate marker and primary outcome have been measured (Study A) and a new study (Study B) in which only the surrogate is measured. Rather than assuming transportability of the conditional mean functions across studies, we consider a class of functions for Study B that deviate from those in Study A. Using these, we estimate the distribution of potential treatment effects on the unmeasured primary outcome and define resilience measures including a resilience probability, resilience bound, and resilience set. Our approach complements traditional surrogate validation methods by quantifying the plausibility of the surrogate paradox under controlled deviations from what is known from Study A. We investigate the performance of our proposed measures via a simulation study and application to two distinct HIV clinical trials.

Incorporating Patient Characteristics for Surrogate Marker Evaluation with Time-to-Event Endpoints*Emily Roberts, University of Iowa*

It's common to assess the impact of a clinical trial treatment on an intermediate outcome, especially when the primary outcome is challenging or expensive to measure. We demonstrate causal methods to evaluate intermediate outcomes when the trial outcomes are time-to-event using principal stratification. We propose illness death models to accommodate the semi-competing risk structure of the outcomes and characterize a valid surrogate using a causal effect predictiveness curve. We include



Session Info

S316

TECHNICAL STATISTICAL SESSION

October 14th, 20:00 - 21:00 UTC

Statistical Learning and Local Impact: Applications in Costa Rica

USE SCHED TO ACCESS THE ZOOM ROOM

This session explores the local impact of statistical learning through diverse applications in Costa Rica. Presentations include an interpretable neural network model for credit scoring among MSMEs, a machine learning approach to predict soil biodiversity in sustainable coffee production, and a cluster analysis of linguistic attitudes among Costa Rican Spanish speakers. Together, these studies demonstrate how interdisciplinary, data-driven methods can address complex social, environmental, and economic challenges with local relevance and global insight.

ORGANIZER: Shirley Rojas-Salazar, University of Costa Rica**CHAIR:** IDWSDS Committee Member

Speaker Bios

S316



TATIANA QUIRÓS MUÑOZ

University of Costa Rica

I hold a bachelor's in Actuarial Science and a Master's in Mathematics and its Applications from the Universidad de Costa Rica. For the past six years, I've worked as a senior mathematical model developer at Banco Nacional de Costa Rica—the largest bank in Costa Rica and second largest in Central America by assets. My work focuses on market and credit risk, actuarial valuation models, and portfolio optimization. What I consider my most significant contribution to data science is a credit scoring model for MSMEs, where I implemented an interpretable neural network to address the “black box” issue in machine learning. As a woman thriving in male-dominated fields like banking and mathematics, I value being a role model for girls and women in science. I also practice pole dance, seeking balance between discipline and expression. I hope my career path will allow me to mentor other women and contribute meaningfully to academia and Costa Rica's economic development.



MARIANNE PEÑA WÜST

University of Costa Rica

Marianne is a third-year Statistics student from the University of Costa Rica with a background in competitive mathematics. Currently, she is a research assistant at the Centro de Investigación en Matemática Pura y Aplicada (CIMPA), where she collaborates with projects related to environmental data science, particularly agriculture and hydrology. She is also a climate activist and has been a part of Fridays for Future Costa Rica and the Frente Ecologista Universitario (FECOU).



KENDY VALVERDE VARGAS

University of Costa Rica

Senior student in the Statistics major at the University of Costa Rica and professor at the School of Philology, Linguistics and Literature at the same institution. Has participated in research projects that combine statistical and linguistic approaches. Passionate about both fields and interested in exploring their intersections to enrich academic and applied work..



NELLYZA RAMÍREZ ROBLETO

University of Costa Rica

Senior student in the Statistics at the University of Costa Rica. She has participated in academic projects focused on data analysis and survey design, with an interest in applying statistical methods in various contexts.



Innovating the Banking With Machine Learning: Credit Score for MSMEs With Explainable Neural Networks

Tatiana Quirós, Actuarial Science, Universidad de Costa Rica

The use of innovative machine learning techniques has significantly transformed the banking sector, including the credit origination process. The adoption of more robust approaches has led to the creation of more accurate classification models. However, this progress has been accompanied by a dilemma in high-level banking discussions: the lack of explainability and interpretability in black-box models involved in credit access decisions. This is a critical issue in banking, as customers and regulators expect to have a reasonable understanding of the variables that could improve or deteriorate their chances for credit access. In this study, we discuss a methodological approach for a credit score for micro, small, and medium-sized enterprises (MSMEs) at a commercial bank in Costa Rica. To address the interpretability challenge, we introduce both local and global interpretability perspectives, highlighting the GamiNet method as an enhanced example for neural networks. Developed in recent years, GamiNet aims to maintain the robustness of a generalized feedforward neural network with multiple additive subnetworks. Each subnetwork consists of several hidden layers, allowing us to capture main effects and pairwise interactions. GamiNet is globally interpretable by design and demonstrates competitive accuracy compared to other machine learning methods such as Random Forests. We also perform various sensitivity analyses on the model to assess its robustness.

A Machine Learning Model for the Prediction of Soil Biodiversity Using Agricultural Practices, Shade Levels, and Plant and Soil Properties

Marianne Peña, University of Costa Rica

By 2050, the worldwide demand for coffee—the second most consumed beverage globally—is expected to be three times greater. Conventional, intensive agricultural methods negatively impact soil chemistry and physical structure. Hence, the development of sustainable coffee farming is essential, especially in developing tropical countries, including Costa Rica. This investigation aims to calibrate a machine learning model for predicting coffee quality based on soil biodiversity metrics (Faith's phylogenetic diversity, Shannon diversity index, and species richness), chemical soil properties, agricultural practices, and shade levels. To this end, an interdisciplinary team, including agronomists, biologists, chemists, microbiologists, and data scientists, is working on the project. Statistics contributes by providing a machine learning model, following a comprehensive analysis of the most pertinent dataset variables. Data science not only provides technical expertise but also helps translate complex ideas into clear, accessible concepts for team members from other disciplines. This facilitates effective collaboration in solving complex, real-world problems.

Cluster Analysis of Costa Rican Spanish Speakers According to Their Linguistic Attitudes: A Comparative Approach by Area of Origin

Kendy Valverde, Universidad de Costa Rica

The objective of this article is to identify clusters of Costa Rican Spanish speakers based on components of linguistic attitudes, to

identify similar characteristics according to sociodemographic variables such as age, gender, educational level, place of origin, perceptions of peculiar speech, and whether they consider themselves to be from an area outside the Greater Metropolitan Area. Three components of linguistic attitudes are analyzed: the cognitive, affective, and behavioral dimensions. The Linguistic Attitudes Scale of Costa Rica (EAL-CR) was used, developed by Kendy Valverde Vargas, a philologist and professor at the School of Philology, Linguistics, and Literature at the University of Costa Rica, and who is also a final-year Statistics student at the same university. Cluster analysis methods (hierarchical, k-means, and fuzzy clustering) were applied to group participants. The hierarchical analysis revealed two clearly distinct clusters: cluster 2 shows more positive and conscious linguistic attitudes, while cluster 1 demonstrates lower appreciation for linguistic diversity. People from peripheral areas tend to show a more reflective and positive attitude toward linguistic diversity, possibly due to exposure to stigma or greater awareness of their dialectal particularities.

Cluster Analysis of Costa Rican Spanish Speakers According to Their Linguistic Attitudes: A Comparative Approach by Area of Origin

Nellyza Ramirez, University of Costa Rica

The objective of this article is to identify clusters of Costa Rican Spanish speakers based on components of linguistic attitudes, to identify similar characteristics according to sociodemographic variables such as age, gender, educational level, place of origin, perceptions of peculiar speech, and whether they consider themselves to be from an area outside the Greater Metropolitan Area. Three components of linguistic attitudes are analyzed: the cognitive, affective, and behavioral dimensions. The Linguistic Attitudes Scale of Costa Rica (EAL-CR) was used, developed by Kendy Valverde Vargas, a philologist and professor at the School of Philology, Linguistics, and Literature at the University of Costa Rica, and who is also a final-year Statistics student at the same university. Cluster analysis methods (hierarchical, k-means, and fuzzy clustering) were applied to group participants. The hierarchical analysis revealed two clearly distinct clusters: cluster 2 shows more positive and conscious linguistic attitudes, while cluster 1 demonstrates lower appreciation for linguistic diversity. People from peripheral areas tend to show a more reflective and positive attitude toward linguistic diversity, possibly due to exposure to stigma or greater awareness of their dialectal particularities.



Session Info

S317

TECHNICAL STATISTICAL SESSION

October 14th, 21:00 - 22:00 UTC

Statistics and Data Science in Aging: Innovative Methods and Impactful Applications

USE SCHED TO ACCESS THE ZOOM ROOM

This session focuses on novel methods and applications in aging and Alzheimer's disease and is organized by the Statistics and Data Science in Aging Interest Group. Speakers will share their exciting work, interest in aging research, and opportunities for further research in aging. This session includes innovative methodology developments for longitudinal data analysis and impactful applications using publicly available datasets in studies of older adults. Jaime will introduce novel machine learning methods for clustered and longitudinal data and an application for predicting mobility limitation in older adults over time using the Health, Aging, and Body Composition Study dataset. Siting and Alicia will discuss a group-based trajectory modeling approach for analyzing frailty among middle-aged and older adults in the Health and Retirement Study based on the accumulation of weight loss, exhaustion, low activity, slowness, and weakness. Zheyu will introduce a multivariate nonlinear latent variable model for estimating the shape and sequence of Alzheimer's disease biomarker cascade for early diagnosis with noisy, sparse and limited longitudinal data. Julia will discuss a novel longitudinal factor model for prediction of three-year Alzheimer's dementia risk with standard cognitive tests and baseline covariates as predictors, which improves enrollment of high-risk subjects in a hypothetical clinical trial.

ORGANIZER: Jaime Speiser, Wake Forest University School of Medicine

CHAIR: Jaime Speiser, Wake Forest University School of Medicine



Speaker Bios

S317



DR. JAIME SPEISER

Wake Forest University School of Medicine

<https://sites.google.com/view/jaimespeiser/home>

Dr. Speiser is a biostatistician focused on prediction modeling with applications in medicine. Her work involves developing novel machine learning methodology for prediction modeling, providing guidance for best practices for developing prediction models, and collaborating with medical researchers to develop models that inform clinical decision making in the learning health system. She serves on the executive committees of the American Statistical Association's Statistical Learning and Data Science Section and Statistics and Data Science in Aging Interest Group. Dr. Speiser is currently funded by a NIH K25 award and is the Leadership Fellow for Wake Forest's Pepper Center and Clinical and Translational Research Institute. Dr. Speiser received her B.S. in mathematics at Elon University and M.S. in statistics at the Ohio State University. She completed her Ph.D. in biostatistics at the Medical University of South Carolina.



MS. SITING CHEN

Oregon Health and Science University

Siting Chen is a biostatistician with Biostatistics & Design Program (BDP) at Oregon Health and Science University (OHSU). Siting specializes in quantitative methodology for health science research, with expertise in longitudinal modelling, survey data analysis, geriatric and aging research, etc. Her research focuses on socioeconomic determinants of health and related disparities in aging trajectories in gerontological research.



MS. ALICIA FERYN

Oregon Health and Science University

Alicia Feryn is a biostatistician with Biostatistics & Design Program (BDP) at Oregon Health and Science University (OHSU). Alicia's research focuses on mortality, aging, and longevity, with a particular emphasis on proteomic predictors and, more recently, the frailty phenotype. She has harmonized the frailty phenotype definition across several large cohort studies—including the Health and Retirement Study (HRS)—with the goal of enabling future meta-analytic work. Alicia is committed to supporting impactful, methodologically rigorous research in the field of aging.



DR. ZHEYU WANG

Johns Hopkins University

<https://profiles.hopkinsmedicine.org/provider/zheyu-wang/2777733>

Dr. Zheyu Wang is an Associate Professor in the Division of Quantitative Sciences at the Sidney Kimmel Comprehensive Cancer Center, with joint appointments in the Departments of Biostatistics and Applied Mathematics and Statistics at Johns Hopkins University. Her research focuses on developing and applying statistical methods to improve diagnostic accuracy and reduce diagnostic errors in healthcare. She has expertise in evaluating diagnostic tests and biomarkers when a gold standard is unavailable or imperfect, with a particular interest in early detection of Alzheimer's disease. Dr. Wang also leads efforts to leverage electronic health record data to identify and reduce diagnostic errors, working collaboratively across disciplines to develop statistical models and automated algorithms informed by real-world clinical data.



MS. JULIA GALLINI

Boston University

<https://www.linkedin.com/in/julia-gallini-366b81b4/>

Julia Gallini is a PhD candidate in biostatistics at Boston University. Her research is focused primarily on longitudinal modeling in neurodegenerative diseases including Alzheimer's dementia and Parkinson's disease. Her specific methodological interests include state-space models and joint longitudinal-survival models. Prior to beginning her PhD at BU, she completed a master's degree in biostatistics at Emory University and worked for three years at the Atlanta VA Medical Center as a research biostatistician.



Predicting Mobility Limitation in Older Adults With Novel Machine Learning Approaches for Longitudinal Data Analysis

Jaime Speiser, Wake Forest University School of Medicine

Prediction models aim to help medical providers, individuals and caretakers make informed, data-driven decisions about risk of developing poor health outcomes, such as fall injury or mobility limitation in older adults. Most models for outcomes in older adults use cross-sectional data, although leveraging repeated measurements of predictors and outcomes over time may result in higher prediction accuracy. This presentation will focus on novel methodology for longitudinal, repeated measures prediction models. This includes extensions of generalized linear mixed models that pair with machine learning methods such as decision tree, random forest, and neural networks. We apply novel machine learning methods, as well as traditional statistical approaches, to predict mobility limitation in older adults using the Health, Aging, and Body Composition dataset.

Group-Based Trajectories of Frailty Among US Older Adults in the Health and Retirement Study

Siting Chen and Alicia Feryn, Oregon Health & Science University

The frailty phenotype reflects diminished reserve across multiple physiological systems and is linked to greater vulnerability for adverse health outcomes. This study aims to identify distinct trajectories of frailty among US older adults and examine risk factors for rapid frailty progression. We used an adapted version of the Fried frailty phenotype, which defines frailty based on five criteria using 0–5-point scale: weight loss, exhaustion, weakness, slowness, and low physical activity. We analyzed data from the Health and Retirement Study (HRS), a nationally representative longitudinal survey of non-institutionalized Americans aged 50 and older, restricting to participants aged 65 and above who were surveyed in wave 8 or later due to availability of frailty measurements. We applied group-based trajectory modeling (GBTM) to identify distinct aging-related frailty phenotype trajectories. We then used multinomial logistic regression to assess sociodemographic and health-related predictors of trajectory group membership. Our analysis identified four distinct frailty trajectories with varying rates of progression with age, along with predictors that differentiated these groups. These findings provide insights into frailty progression and can inform targeted prevention and early intervention strategies. Future work will account for time-varying covariates in GBTM and apply multi-state modeling to better capture transitions between frailty states and address survivor bias.

Joint Modeling of Multivariate Biomarker Trajectories Under Sparse and Limited Longitudinal Data

Zheyu Wang, Johns Hopkins University

Alzheimer's Disease (AD) research increasingly recognizes the importance of characterizing dynamic biomarker trajectories to better understand the underlying pathological processes. Key challenges in this effort include the fact that the disease process is not directly observable and needs to be differentiated from general aging-related changes. Meanwhile, biomarker measurements are often noisy, influenced by individual characteristics, and often exhibit complex temporal and cross-

sectional correlations. These complexities are further compounded by the common data realities in most longitudinal AD studies: relatively small sample sizes, infrequent measurements, and limited longitudinal coverage that may not capture the full course of biomarker progression. In this work, we propose a nonlinear latent variable model to address these challenges. Our model characterizes the trajectories of multi-domain AD biomarkers trajectories as a function of an unobservable, continuous latent variable representing disease progression. This approach enables the separation of disease-related biomarker changes from those due to general aging, and accounts for complex correlation structures and individual factors. We also proposed practical and computational strategies to enable fast and reliable computation and inference under practical constraints. We will end our talk with an application in ADNI.

Cognitive Factor-Based Selection Increases Power in Alzheimer's Dementia Randomized Clinical Trials

Julia Gallini, Boston University

Alzheimer's dementia (AD) is of increasing concern as populations achieve longer lifespans. Many of the recent failed AD clinical trials had a low number of AD events and were thus underpowered. Previous trials have attempted to address this issue by requiring signs of cognitive decline in brain imaging for trial enrollment. However, this method systematically excludes populations without access to healthcare and results in overwhelmingly wealthy and white trial participants. We therefore propose the use of a predictive model based on cognitive test scores to enroll cognitively normal yet high risk participants in a hypothetical clinical trial. Cognitive test scores are a widely accessible tool so their use in enrollment would be less likely to exclude marginalized populations than imaging data. We developed a novel longitudinal factor model to predict AD conversion within a 3-year window based on data from the National Alzheimer's Coordinating Center. Through simulation we demonstrate that when compared to other methods of subject selection our predictive model provides substantial improvements in statistical power in hypothetical clinical trials across a range of drug effects.



Session Info

S318

TECHNICAL STATISTICAL SESSION

October 14th, 22:00 - 23:00 UTC

Statistical Methods for Impactful Research on Health and the Food Environment

USE SCHED TO ACCESS THE ZOOM ROOM

Accurately assessing the relationship between food access and health is essential for understanding and addressing nutrition-related disparities, yet researchers face substantial challenges in data accuracy, availability, and methodological rigor. This session brings together statistical perspectives on food access and health from multiple angles, including methods to correct bias from misclassified or error-prone proximity measures, evaluation of the impact of measurement error on health equity analyses, and methodological considerations for assessing community-based interventions. Through applications involving electronic health records, geospatial analysis, and program evaluation, the talks demonstrate innovative approaches to producing reliable, actionable insights to inform policies and programs aimed at improving access to healthy foods and reducing disease burden.

ORGANIZER: Sarah Lotspeich, Wake Forest University**CHAIR:** Sarah Lotspeich, Wake Forest University

Speaker Bios

S318



MS. ASHLEY MULLAN

Vanderbilt University

<https://ashleymullan.github.io/>

Ashley Mullan is a second year Ph.D. student in biostatistics at Vanderbilt University. She earned her master's degree in Statistics and graduate certificate in Data Science from Wake Forest University, and she earned her bachelor's degrees in Applied Mathematics and Philosophy from the University of Scranton. Ashley's main (bio)statistical interests include developing and applying methods for mismeasured or missing data, collaborating with child welfare policy makers to enhance the quality of services provided to at-risk children, and supporting the next generation of (bio)statistics superstars. In her spare time, she collects and analyzes data about her own pop culture consumption habits. She's an avid fan of (sit/rom)com, (data) comm, and (oxford) commas!



MS. CASSANDRA HUNG

Wake Forest University

<https://www.linkedin.com/in/cassandra-hung-035a89241/>

Cassandra Hung is a fourth-year undergraduate student at Wake Forest University, where she is studying Applied Mathematics and Statistics. Previously, she attended the Big Data Summer Institute at the University of Michigan and an NSF REU at the University of Iowa. She is currently involved in research on measurement error and health disparities, and is excited to pursue a future career in (bio)statistics. Outside of her classes and research, Cassandra enjoys spending time outside, reading, and trying new foods.



DR. ANDREA LANE

Duke University

<https://datascience.duke.edu/people/andrea-lane/>

Andrea Lane, PhD, is an Assistant Professor of the Practice in the Master in Interdisciplinary Data Science program at Duke University in North Carolina, US. Andrea joined Duke in 2022 after completing her PhD in Biostatistics at Emory University. Her work as a collaborative biostatistician focuses on assessing and addressing health inequities, and her work in statistics education focuses on incorporating equity principles and critical pedagogy into statistics and data science coursework.



MS. SIZHE CHEN

Duke University

Size Chen is a second-year student in the Master in Interdisciplinary Data Science program at Duke University in North Carolina, US. She earned her bachelor's degree in psychology & statistics from Boston University in 2024.



Modeling Misclassification: Exploring the Relationship between Diabetes and Access to Healthy Foods*Ashley Mullan, Vanderbilt University*

Access to healthy food is key to maintaining a healthy lifestyle and can be quantified by the distance to the nearest grocery store. However, calculating this distance forces a trade-off between cost and correctness. Accurate route-based distances following passable roads are cost-prohibitive, while simple straight-line distances ignoring infrastructure and natural barriers are accessible yet error-prone. Categorizing low-access neighborhoods based on these straight-line distances induces misclassification and introduces bias into standard regression models estimating the relationship between disease prevalence and access. Yet, fully observing the more accurate, route-based food access measure is often impossible, which induces a missing data problem. We combat bias and address missingness with a new maximum likelihood estimator for Poisson regression with a binary, misclassified exposure (access to healthy food within some threshold), where the misclassification may depend on additional error-free covariates. In simulations, we show the consequence of ignoring the misclassification (bias) and how the proposed estimator corrects for bias while preserving more statistical efficiency than the complete case analysis (i.e., deleting observations with missing data). Finally, we apply our estimator to model the relationship between census tract diabetes prevalence and access to healthy food in northwestern North Carolina.

Exploring the Influence of Error on Measures of Health Equity: A Case Study*Cassandra Hung, Wake Forest University*

Healthy eating is an important part of living a healthy life, but disparities in food access may result in disproportionately high rates of disease in communities with fewer healthy food options. To explore the connection between food access and nutrition-related disease outcomes, we investigate the relationship between individual patients' access (proximity) to their nearest healthy food retailer and the prevalence of type 2 diabetes. However, measuring proximity raises additional questions, as distance measures are often either computationally simple and inaccurate (straight-line) or computationally complex and accurate (map-based). Using the simpler but less accurate straight-line distance may introduce measurement error that misrepresents a patient's access to healthy foods, which raises the question of whether this error needs to be accounted for in order to accurately assess the relationship between access and type 2 diabetes prevalence. To address these questions, we extract patient information (including diabetes diagnoses) from the electronic health record (EHR), geocode patients' home addresses, and calculate both straight-line and map-based proximity to healthy foods. We then quantify whether patients with farther proximities to healthy foods face a higher burden of prevalent diabetes, and assess any differences that may arise from using straight-line instead of map-based distances in this analysis.

Evaluating a Fresh Produce Delivery Program to Assess Potential Health Benefits*Andrea Lane and Sizhe Chen, Duke University*

In 2017, students at the Duke School of Medicine founded the Fresh Produce Program to provide eligible households in the Durham, NC area with biweekly boxes of fresh produce. In 2020, in response to the COVID-19 pandemic, the program transitioned to a delivery model. Eligibility is based on financial need and health risk factors, but the program has not been formally evaluated to determine potential health benefits from increased access to fresh produce. Our evaluation links program participants to their electronic health records (EHR) within the Duke system, presenting statistical challenges related to extensive data cleaning and manipulation. Additionally, the program's natural experiment design requires careful methodological decisions to ensure valid conclusions. In this talk, we will describe these analytic challenges and share preliminary findings from our evaluation of this community health initiative.



Session Info

S401

CAREER OR PROFESSIONAL DEVELOPMENT SESSION

October 14th, 15:30 - 16:15 UTC

Advanced Prompting Techniques for Data Science Workflows

USE SCHED TO ACCESS THE ZOOM ROOM

Transform your data science practice with strategic prompting techniques in this intensive 60-minute hands-on workshop. Participants will explore how Large Language Models revolutionize data analysis, visualization, and code generation through effective prompt engineering. Learn to craft prompts that generate production-ready code, provide data cleaning strategies, and recommend optimal visualization approaches.

This interactive session guides participants through comparative prompting across three platforms (Poe, Hugging Face, and Groq) using real-world scenarios including healthcare data cleaning, customer churn analysis, and imbalanced classification problems. The second half demonstrates practical LLM integration using Google Colab notebooks with bird migration datasets, showing how well-crafted prompts accelerate exploratory analysis, automate coding tasks, and enhance decision-making.

Designed for data scientists, analysts, and researchers seeking AI assistance integration, this workshop requires no prior AI experience while delivering immediately applicable strategies. Attendees leave with proven prompt templates, platform insights, and a comprehensive resource guide for implementing LLM-assisted workflows, dramatically reducing time spent on repetitive tasks while improving analysis quality.

ORGANIZER: Jeanne McClure, NC State University**CHAIR:** Jeanne McClure, NC State University

Speaker Bios

S401

**DR. JEANNE MCCLURE**

NC State University

<https://datascienceacademy.ncsu.edu/people/jeanne-mcclure/>

Dr. Jeanne McClure is a postdoctoral fellow at NC State University's Data Science and AI Academy and founder of Ars Innovate Technologies and Consulting. She specializes in AI integration, prompt engineering, and workforce development, helping faculty, students, and professionals apply large language models in teaching, research, and daily workflows. A three-time NC State graduate with a Ph.D. in Learning Design and Technology, Jeanne bridges learning theory with practical applications in data science and AI. Her work emphasizes ethical AI use, open-source tools, and reproducibility. She is a former NCES fellow, a 2025 Posit Cloud Opportunity Grant recipient, and serves on the UNC System AI Pedagogy Committee. Jeanne is passionate about expanding access to AI and data science education across sectors, supporting professional growth through hands-on learning and innovative, real-world applications.



Session Info

S402

TECHNICAL STATISTICAL SESSION

October 14th, 16:00 - 17:00 UTC

Bayesian Analytics in the Wild

USE SCHED TO ACCESS THE ZOOM ROOM

Bayesian methods are increasingly being applied beyond theory to tackle complex, real-world problems across science, technology, and even sports. This session showcases three examples of “Bayesian analytics in the wild.” The first talk explores how Bayesian analyses can address preferential sampling in environmental monitoring, providing more reliable inference when data collection is uneven or biased toward areas of concern. The second highlights a new R package for examining space-time data using a Bayesian spatio-temporal factor analysis model, offering tools for dimension reduction, visualization, and interpretation of complex environmental and ecological datasets. The third presentation steps into the sports arena, demonstrating how Bayesian models can be used to predict game outcomes in the National Hockey League (NHL), integrating prior knowledge with dynamic data to capture the uncertainty inherent in competitive sports. Together, these talks illustrate the flexibility, creativity, and power of Bayesian analytics in diverse applied settings.

ORGANIZER: Candace Berrett, Brigham Young University

CHAIR: Candace Berrett, Brigham Young University



Speaker Bios

S402

**CAMILLA MCKINNON**

Brigham Young University

Camilla McKinnon is a Statistics Master's student at Brigham Young University. Her research involves statistical modeling to better understand harmful algal blooms, with attention to biases from preferential sampling and different data collection types. She graduated from BYU in 2024 with a B.S. in Statistics: Data Science with minors in Math and Music.

**CANDACE BERRETT**

Brigham Young University

Dr. Candace Berrett is a Mel Carter Professor and Associate Chair in the Department of Statistics at Brigham Young University. Her work focuses on developing Bayesian methods for analyzing complex spatial and spatio-temporal data, with a focus on environmental and public health applications. She is actively involved in the American Statistical Association (currently serving as chair-elect for the Section on Bayesian Statistical Science) and the International Society for Bayesian Analysis (currently serving as past chair for the Section on Environmental Sciences).

**MORGAN KURTH**

University of Waterloo

Morgan Kurth is a Statistics Master's student at the University of Waterloo. Prior to joining Waterloo, she earned a BS in Statistics: Data Science from Brigham Young University with a minor in Business and was Assistant General Manager and Statistician for the UVU Men's Hockey program. She is passionate about using Bayesian methods for sports analytics, and her graduate research focuses on the impact of player interactions on NHL forward line performance.





Preferential Sampling Model for Harmful Algal Blooms in Utah Lake

Camilla McKinnon, Brigham Young University

Harmful Algal Blooms (HABS) can produce toxins that pose serious health risks to humans, pets, and livestock. Data from Utah Lake over the past years have largely been collected in a preferential manner, meaning samples are primarily gathered when HABS are already suspected to be present. This introduces bias, as it fails to capture the full range of bloom conditions, including periods of absence or low risk. A further challenge is distinguishing between presence-only data and systematic presence/absence data, both of which are available and require different modeling strategies. To address these issues, we are developing a Bayesian hierarchical framework that incorporates seasonal patterns and environmental covariates to estimate latent bloom behavior. This ongoing work aims to improve inference on HAB dynamics and support more reliable risk assessment.

BSTFA: An R Package for Spatio-Temporal Data

Candace Berrett, Brigham Young University

Analyzing complex spatio-temporal datasets often requires tools that can capture underlying dependence structures while remaining interpretable and computationally efficient. The BSTFA package implements Bayesian spatio-temporal factor analysis models that reduce dimensionality, identify latent structure, and provide principled uncertainty quantification. In this talk, I will illustrate the package's workflow and capabilities through applications to temperature and PM2.5 data, showing how BSTFA can uncover meaningful spatial and temporal patterns. These examples highlight how the package enables students and researchers to move from raw, high-dimensional data to actionable insights in environmental and ecological applications.

Predicting Wins in the National Hockey League

Morgan Kurth, University of Waterloo

The purpose of this study is to build a win prediction model for National Hockey League teams. We are using data from the NHL collected during the 2018-2019 hockey season. The data consist of 825,000 play-by-play game events including shots, saves, penalties, faceoffs, hits, etc. Variable summaries (e.g. shots per minute) are produced for each game of the given season. We have developed a model which predicts the winning team of a game using a product partition model framework. This prediction is based on metrics summarizing a team's performance at any time in a particular game. These metrics include Scoring Chances, Saves, Takeaways, and Blocked Shots, all on a per-minute of gameplay, standardized, scale. The project also explores other modeling methods and compares the chosen framework's performance with the accuracy of the Bayesian Additive Regression Tree (BART) model.



Session Info

S403

CAREER OR PROFESSIONAL DEVELOPMENT SESSION

October 14th, 18:00 - 18:45 UTC

Portuguese Biometry outside Academia

USE SCHED TO ACCESS THE ZOOM ROOM

This session is dedicated to the contributions of Portuguese women who work in a non-academic setting in Portugal and use Statistics and Data Science to help improve agricultural yield, hospital management, and healthcare decisions for the betterment of society.

ORGANIZER: Nuno Sepulveda, CEAUL**CHAIR:** Clara Cordeiro, Universidade do Algarve, Portugal

Speaker Bios

S403



DR. DIANA TERESO FERREIRA

Frutas Tereso

<https://www.frutastereso.pt/>

Diana Tereso Ferreira is a young farmer, dedicated to the management of Frutas Tereso, a citrus company, located in the Algarve, Portugal. She is also Vice-President of AlgarOrange (Algarve Citrus' Operators Association) and Vice-President of FEDAGRI (Algarve Agriculture Federation).

With a background in Pharmaceutical Sciences and experience in Health Economics and Outcomes Research, Diana is now dedicated to her family business and to making sure that the Agriculture of the Algarve has a prosperous future.



DR. ANA MARREIROS

Unidade Local de Saúde do Algarve, ULS Alg

Marreiros, A (female). Since 1998, she has been a university professor in the fields of mathematics and statistics, and since 2015, she has been a professor in the Master's in Medicine program at the University of Algarve, where she conducts clinical research in the fields of neurology and vascular surgery. She is a member of the National Ethics Committee for Clinical Research, in Portugal; she was vice-president of the Faculty of Medicine and Biomedical Sciences at the University of Algarve; she is coordinator of the faculty's Medical Education Unit; she is Executive Board Member of Algarve Biomedical Center, clinical academic center; and, since November 2024, she was invited to administer the Algarve Local Health Unit (ULS Algarve).



DR. VALESKA ANDREOZZI

Exigo Consultores

<https://exigoconsultores.com/>

Valeska Andreozzi is a Senior Biostatistician at Exigo Consultores with a PhD in Biomedical Engineering from the Federal University of Rio de Janeiro. She has held academic posts at the University of Lisbon and NOVA Medical School, and her research spans biostatistics, epidemiology, and health technology assessment. Dr. Andreozzi is co-author of a book on modelling survival data and has published extensively on longitudinal and survival analysis, health economics, and applied statistical modeling.





Encouraging the Synergy Between Agriculture and Statistics

Diana Ferreira, Frutas Tereso

In my intervention, I will give some examples of how we, at Frutas Tereso, are using statistics: to analyze data and making informed decisions about treatments and harvests; and to analyze the impact of climate and water availability in our fruits and use this information to influence public policies.

I am also taking this opportunity to set some challenges for the future as I truly believe that the synergy between agriculture and statistics is fundamental not only to improve productivity and efficiency but also to ensure a more sustainable future for agriculture.

A Novel Solution for Social Responses in Health Care System

Ana Marreiros, Unidade Local de Saúde do Algarve, ULS Alg

In this talk, I will present a solution for emergency episodes and hospitalizations that, while also having a basis in social responsibility, incur high costs for hospitals. The problem can be summarized in two questions: 1. At what point do we cease to have a responsibility to provide clinical care for these patients? 2. On the other hand, how can we resolve these cases of social response without dehumanizing them? The creation of a 4-dimensional bi-vector (vector 1: time of clinical discharge; vector 2: time k of stay in the system). Based on this configuration, it is possible to answer these questions, and based on clinical care and social response scores, we are providing social referral responses to our citizens.

Statistics at the Core of Health Technology Assessment

Valeska Andreozzi, Exigo Consultores

Health Technology Assessment (HTA) plays a critical role in informing healthcare decision-making by evaluating the clinical, economic, and social value of health technologies. In a HTA consulting company, statistical methods are central to this process, ensuring that evidence is both robust and relevant for decision makers. This talk will provide an overview of the main statistical approaches applied in HTA, including survival analysis for extrapolating clinical trial outcomes, indirect treatment comparisons to bridge gaps in evidence, and regression-based models for cost and quality-of-life data. By illustrating how these methodologies are applied in real-world projects, the presentation will highlight the unique interface between rigorous statistical science and practical decision-making in healthcare.



Session Info

S404

CAREER OR PROFESSIONAL DEVELOPMENT SESSION

October 14th, 18:30 - 18:45 UTC

S404 - 2RAM2: Relevance, Response, Aggregation, Model & Measure

USE SCHED TO ACCESS THE ZOOM ROOM

One of my secrets to thriving in the commercial data science environment is ensuring that the projects we undertake set us up for success. To achieve this, there are at least five critical decisions, agreements, choices, clarifications, or definitions that need to be made before starting a project.

While there are many frameworks for prioritisation and resource allocation, I found none specifically tailored for data science projects, so I created one myself. Popular frameworks like "Low Hanging Fruit," "MoSCoW," and "Eisenhower Matrix" may work in other contexts, but they lack the specificity needed for successful data science projects.

I will delve into the details of the 2RAM2 framework, providing examples and guiding you through the implementation process. Women, especially young women, face additional challenges because defining successful projects requires a partnership mentality. Often, the requesting party approaches us with a buyer/vendor mindset. Having a formal guiding approach helps establish and reinforce that partnership, making it a natural part of the process, and providing a structured way of saying no, if needed be, in any power imbalanced relationship.

ORGANIZER: Helena Baptista, Eli Lilly and Co. and NOVA IMS

CHAIR: Helena Baptista, Eli Lilly and Co. and NOVA IMS



Speaker Bios

S404

**DR. HELENA BAPTISTA**

Eli Lilly & Co.

Helena Baptista holds a PhD in Statistics and Econometrics from NOVA IMS, Universidade Nova de Lisboa. With over 30 years of experience in the pharmaceutical industry, finance, and academia, Helena is a highly experienced statistician, researcher, and educator. Specialised in applied statistics, forecasting, and time series analysis, Helena has a strong background in statistical modelling, data analytics, and methodological research. As a Professor of Practice, Helena has been actively involved in organizing and teaching courses in applied, basic, and advanced statistics, both in traditional and e-learning formats. In the role of Director within the Business Insights and Analytics team, Helena works in the New Generation Data Science group, focusing on improving existing methodologies for impact modelling and exploring how new data sources can better assist patients. Helena is highly motivated to learn and is dedicated to helping others achieve their full potential.



Session Info

S405

CAREER OR PROFESSIONAL DEVELOPMENT SESSION

October 14th, 19:00 - 20:00 UTC

When should I retire? And then what?

USE SCHED TO ACCESS THE ZOOM ROOM

Stages of life and transitions?

How did you (do you) know when it's time to retire? Relate not only your experience but experiences of other female statisticians.

When (age compared to eligibility for retirement) and why did you decide to?

And several more topics!

Panelists:

Elizabeth Margosches (representing a retiree who does things other than statistics)

Nancy Flournoy (representing a retiree who still does statistics)

Nancy Geller, NHLBI/NIH (representing someone who says she's too old to retire!)

Motomi Mori, St Jude's (someone thinking about retiring)

ORGANIZER: Jessica Kohlschmidt, Ohio State University and Caucus for Women in Statistics

CHAIR: Jessica Kohlschmidt, Ohio State University and Caucus for Women in Statistics



Speaker Bios

S405



DR. ELIZABETH MARGOSCHES

Retired

After taking her MS in Statistics from Rutgers - The State University of New Jersey part-time while working at Educational Testing Service and following it with a MPH and PhD in Biostatistics from the University of Michigan, Dr Margosches spent 32+ years at the US Environmental Protection Agency in health risk assessment before retiring in 2012. She is an ASA Fellow and a past president of CWS.



DR. NANCY FLOURNOY

Retired

Dr Flournoy's early training was in Biostatistics from UCLA (BS, MS) and she worked for some years in California before travelling to the University of Washington to be director of Clinical Statistics of the then-newly-formed Fred Hutchinson Cancer Research Center. She completed her doctorate in Biomathematics from the University of Washington and subsequently became the first female director of the National Science Foundation program in statistics. From there she went to American University for 14 years and then became department chair at the University of Missouri from which she retired. She is a Fellow of the ASA, the IMS, the World Academy of Arts and Sciences, and the AAAS, and is a past president of CWS. She has received the COPSS Scott and David Awards, and the Norwood award from the University of Alabama.



DR. NANCY GELLER

NHLBI/NIH

Dr. Nancy Geller is proud to be one of the founding members of the Caucus and for many years has been the Director of the Office of Biostatistical Research at the National Heart, Lung and Blood Institute (NHLBI). She directs a group of 10-12 statisticians who collaborate with each other and with programs supported by NHLBI. The lines under her eyes are mostly covered by her glasses, but when pressed, she says she's too old to retire! In this session she'll discuss her current thoughts on when would be the right time for her to retire. In addition to her roles in the Caucus and her workplace, she is an ASA Fellow and a past president of the ASA.



DR. MOTOMI MORI

St Jude's

Dr. Mori is a collaborative biostatistician in cancer research with research interest in early-phase clinical trials. She is an ASA fellow and an ASA board member, past president of CWS and IBS WNAR.



Session Info

S406

TECHNICAL STATISTICAL SESSION

October 14th, 20:00 - 20:45 UTC

Advanced Statistical Methods for Complex Health Data

USE SCHED TO ACCESS THE ZOOM ROOM

This session features recent methodological advances in the analysis of complex health-related data. Presentations include a latent variable approach for modeling longitudinal microbiome changes, regression techniques for zero-truncated recurrent events in mental health care, and N-of-1 trial designs for evaluating biosimilar drugs. These methods address challenges such as high dimensionality, temporal dynamics, and complex trial structures, offering new tools for clinical and biomedical research.

ORGANIZER: Wendy Lou, University of Toronto**CHAIR:** Wendy Lou, University of Toronto**SPONSOR:** STATA

Speaker Bios

S406

**ANASTASIA TETERINA**

University of Toronto

Anastasia Teterina is a PhD student in Biostatistics at the University of Toronto, specializing in the analysis methodologies for high-dimensional “omics” data in cohort studies, with the focus on biologically meaningful interpretability. Her research is focused on cohort microbiome data, where she explores approaches to tackle the challenges posed by sparse, compositional, and longitudinal structures. Her works includes several areas, such as use of longitudinal regression models for differential abundance analysis, Bayesian hierarchical generative models for sparse count data, and analytical approaches for longitudinal compositional datasets.

**ANQI (ANGELA) CHEN**

University of Alberta

Anqi (Angela) Chen received her PhD in Statistics from Simon Fraser University under the co-supervision of X. Joan Hu and Rhonda J. Rosychuk. Her research interests include methodological development for analyzing recurrent event data, incomplete data, lifetime data, and longitudinal data. She is currently a Postdoctoral Fellow in the Department of Pediatrics at the University of Alberta, co-supervised by Rhonda J. Rosychuk and X. Joan Hu.

**YUQING LIU**

University of Toronto

Yuqing Liu is a PhD candidate in Biostatistics at University of Toronto. Her research focuses on innovative clinical trial designs for bioequivalence and biosimilar drug development, including crossover designs, N-of-1 trials, and mixed effect models. She has also contributed to collaborative research projects in aging, mental health, and neuroscience, applying advanced statistical methods to large-scale datasets and meta-analyses.





Investigating Temporal Changes in Microbiome Cohort Data using Latent Dirichlet Allocation Approach and Its Extensions

Anastasia Teterina, University of Toronto

Collecting microbiome data in cohort studies (e.g. samples collected at baseline and follow-up time points) is becoming a common practice in clinical research; however, appropriate statistical methods for analyzing such datasets, often complex and high-dimensional, are still limited. Motivated by an ongoing research project involving patients with Metabolic dysfunction-associated steatotic liver disease, we propose to use 2-stage latent variable approach to analyse changes in microbiome. At Stage 1, a Latent Dirichlet allocation (LDA) is utilized to characterise each sample as originating from a mixture of bacterial communities, which can be considered as unobservable underlying features in lower-dimensional space determined by posterior vectors of mixing probabilities. At Stage 2, these mixing probabilities are used to characterize changes of within-subject mixtures of communities over time. In order to take into account compositional nature of the resulting latent variables, two approaches are considered: Dirichlet regression and compositional data analysis methods based on principles of Aitchinson geometry. The proposed approach is illustrated using cohort data from a bariatric surgery study.

assessing the biosimilarity of three drugs using a complete N-of-1 trial design, offering valuable insights into its effectiveness and practical applicability. By refining statistical methodologies, this study aims to strengthen regulatory decision-making and facilitate the broader clinical adoption of biosimilars.

Stratified Regression Analysis of Zero-Truncated Recurrent Event Data, with Applications in Pediatric Mental Health Care

Anqi Chen, University of Alberta

This talk presents a strategy for analyzing zero-truncated recurrent events data. Motivated by a pediatric mental health care (PMHC) program, we are particularly concerned with how the event occurrence is influenced by the occurrences in the past. We consider a stratified Cox regression model with time-varying coefficients for the recurrent events and propose a procedure for estimating the model parameters using the zero-truncated data integrated with population census information. The proposed estimator is evaluated both numerically and asymptotically. Data from the PMHC program are used throughout the whole presentation to motivate and to illustrate the proposed approach. This is joint work with Joan Hu, Rhonda Rosychuk, and Leilei Zeng.

Complete N-of-1 Trial Design in Biosimilar Drug Development

Yuqing Liu, University of Toronto

The assessment and use of biosimilars have become an essential component of modern drug development due to their great potential to reduce healthcare costs and expand access to biological therapies. For biosimilarity comparisons, the standard two-drug, two-sequence crossover trials may be inadequate for increasingly complex scenarios. In particular, when multiple reference products, drugs, or dosing regimens need to be bridged (e.g., across different regulatory regions or product lines), higher-order or multi-arm designs are required. Recent literature highlights N-of-1 trial design to meet this need. Such designs allow within-subject comparisons with high efficiency and multi-drug assessment. However, its application in biosimilar drug development remains underexplored due to the potential challenges in complex data analysis and implementation. This research provides a comprehensive evaluation framework for



Session Info

S407

TECHNICAL STATISTICAL SESSION

October 14th, 20:45- 21:30 UTC

Analysis of Data from Complex Systems

USE SCHED TO ACCESS THE ZOOM ROOM

Advancing in information and digital technologies enhances data collection processes. An abundance of data across industries presents the necessity and challenges for statistical methods. The goal of this session is to showcase statistical methodologies for data from complex systems, including models for analyzing incomplete data from complex surveys; a framework for analyzing time-to-event outcomes using data from health data on multiple platforms, with emphasis on nonprobability and probability surveys; and analyses of drafting and performance of using the Women's National Basketball Association (WNBA) data.

ORGANIZER: Zilin Wang, Wilfrid Laurier University**CHAIR:** Wendy Lou, University of Toronto

Speaker Bios

S407

**IPSEETA GARG**

Wilfrid Laurier University

Ipseeta Garg is a third-year PhD candidate in Statistics at Wilfrid Laurier University. Her research focuses on developing survey methodology for applications in biostatistics, with an emphasis on establishing rigorous inferential foundations to improve survival estimation. In addition to her research, she is dedicated to teaching and mentorship, where she supports undergraduate students, particularly those from non-mathematical backgrounds, in developing confidence and appreciation for statistics.

**JULIA MATYS**

Wilfrid Laurier University

Julia Matys is an MSc Mathematics student at Wilfrid Laurier University, with a research focus in women's sports analytics, specifically evaluating NCAA women's basketball talent for professional drafting to the WNBA. Her work applies statistical modelling and data analysis to support talent identification, career prediction, and decision-making in women's basketball. Passionate about equity and inclusion in STEM, she co-founded Laurier Women in Mathematics, a student-led organization dedicated to support and empowering women studying mathematics.

**ZILIN WANG**

Wilfrid Laurier University

Zilin Wang is currently an Associate Professor in the Department of Mathematics at Wilfrid Laurier University. She received her PhD degree in Statistics from the University of Western Ontario in 2004. Her research areas include survey sampling, resampling techniques, sports analytics, and statistical modelling in finance. Dr. Wang is passionate about advocating for female researchers and applying statistics to improve the professional lives of equity-deserving groups in the academic environment.



A Framework to Analyze Time-to-Event Outcomes From Non-Probability Samples*Ipseeta Garg, Wilfrid Laurier University*

We propose a framework for analyzing time-to-event outcomes in non-probability cohorts using external probability samples to adjust for selection bias. The framework includes inverse probability weighting (IPW), outcome regression (OR), and doubly robust (DR) estimation, ensuring consistency under model misspecification. We consider two approaches for estimating survival probabilities: pseudo-observations to reconstruct complete data and direct modification of survival probability estimation from censored data. The methods apply to parametric and semi-parametric models, including Kaplan-Meier and Cox proportional hazards models. Simulations show that IPW and OR estimators perform well when correctly specified but degrade under misspecification, whereas DR estimators are unaffected if at least one model is correct. This framework extends statistical methods for continuous outcomes to survival analysis, providing a robust approach for handling bias in non-probability cohorts.

missingness, which are used to model the propensity scores of the study variable. Using the observed weighted likelihood function and the survey estimating equation methods, we estimate and make inferences on the parameters of the distribution of the study variable, the probability of missingness within each subset, and the propensity scores of the study variables. We applied this method to large complex surveys, such as the National Health and Nutrition Examination Survey (NHANES) and Canadian Longitudinal Studies on Aging (CLSA). (This is joint work with Mary Thompson).

From Draft Day to the Court: Evaluating Player Characteristics and Performance of WNBA Draftees*Julia Matys, Wilfrid Laurier University*

The Women's National Basketball Association (WNBA) is one of the most competitive sports leagues in the world, with limited roster spots and strict salary caps making it challenging for players to secure and maintain their place on a team's roster. Each year, the WNBA draft serves as a pivotal event where teams select top talent from the National Collegiate Athletic Association (NCAA) and international leagues. For teams, making informed draft decisions is crucial to acquiring players who can meet the league's high standards and contribute to the ultimate goal: a championship. The talk aims to apply statistical methodology to identify key characteristics of successful draftees, focusing on variables that influence their ability to sustain their roster spot. Logistic regression models are used to analyze the probability of a successful draft of a player. Mixed effects models are used to account for the clustering effect of college conferences, recognizing that players from the same college conference may share performance characteristics. Cox proportional hazards models are used to provide insight into the longevity of player careers in the WNBA. The combination of the best three model outputs is used to create a draft score for the 2025 senior class of the NCAA DI Women's Basketball program. We revealed that the key factors that contribute most to success and longevity in the WNBA.

Modelling Missing-Not-at-Random for Data From Complex Surveys*Zilin Wang, Wilfrid Laurier University*

In the analysis of incomplete data in complex surveys, the probability of being non-missing (propensity score) could depend on the study variable itself, a missing mechanism called missing-not-at-random (MNAR). In this paper, we model the propensity score using the MNAR mechanism, assuming that the probability of missingness is not homogeneous across the support of the study variable. Conditional on observed values in subsets of support of the study variable, we assign different probabilities of



Session Info

S408

TECHNICAL STATISTICAL SESSION

October 14th, 21:30- 22:00 UTC

Practical Methods for Real-World Research: Balancing Rigor and Simplicity

USE SCHED TO ACCESS THE ZOOM ROOM

This session explores methodological strategies for conducting research using real-world data, with a focus on balancing statistical rigor and practical clarity. The first presentation compares propensity score methods offering insights into how different approaches handle confounding and influence survival estimates. The second presentation advocates for simplicity in data analysis, demonstrating how clear definitions, disciplined data preparation, and transparent modeling can yield more interpretable and trusted results than overly complex techniques. Together, these talks highlight the importance of methodological choices in shaping research outcomes in real-world settings.

ORGANIZER: Wendy Lou, University of Toronto**CHAIR:** Wendy Lou, University of Toronto

Speaker Bios

S408

**XIAOXUAN (ROSE) HAN**

AstraZeneca

Xiaoxuan (Rose) Han recently completed her MSc in Biostatistics at the University of Toronto and joined AstraZeneca as a Statistician. For her master's thesis, she collaborated with Roche to compare propensity score methods for assessing racial disparities in multiple myeloma health outcomes using real-world data from Flatiron Health. She also conducted research with the Interventional Psychiatry Program at St. Michael's Hospital on the impact of life stressors on mental health outcomes during and after the COVID-19 pandemic. Her research interests include propensity scores, survival analysis, real-world evidence, and health disparities.

**QIAOWEI (VICKY) LIN**

Unity Health Toronto

Qiaowei (Vicky) Lin is a Data Analyst in the Interventional Psychiatry Program at St. Michael's Hospital, working under the direction of Dr. Venkat Bhat and supervision of Dr. Wendy Lou. She earned her MSc in Biostatistics from the University of Toronto and a bachelor's degree in Statistics with a minor in Applied Mathematics from the University of Waterloo. Her work focuses on statistical analyses for survey design data, meta-analyses, and longitudinal clinical data related to mental health, contributing to over ten peer-reviewed publications in the past two years.



Comparisons of Propensity Score Methods for Assessing Racial Disparities in Multiple Myeloma Health Outcomes

Xiaoxuan Han, AstraZeneca

Propensity score (PS) methods are widely used to reduce confounding in observational research but have not been systematically compared in the context of assessing racial disparities in multiple myeloma (MM) outcomes using real-world data (RWD). This research applies and evaluates seven PS methods, including nearest neighbor matching with and without a caliper, optimal matching, stabilized and trimmed inverse probability of treatment weighting, PS stratification, and using PS as a covariate, to examine racial disparities in overall survival and time to next treatment among 12,053 MM patients in the United States from the Flatiron Health database. Each PS method is assessed based on its performance in balancing demographic and biomarker covariates, with race/ethnicity effects estimated using Cox proportional hazards models. The findings offer practical guidance on the strengths and limitations of PS methods in addressing confounding and provide methodological insights for future racial disparities research using RWD.

Less Is More: Simple Methods for Impactful Clinical Analysis

Qiaowei Lin, Unity Health Toronto

In clinical research, the most impactful analysis isn't always the most complex—it's the one that everyone understands, trusts, and can replicate. In this talk, the KISS principle is applied to show how clear definitions, disciplined data cleaning, and transparent variable construction often address clinical questions before advanced modeling is even needed. Using real-world examples, we will demonstrate how simple, well-specified methods, such as adjusted regressions or meta-analyses with clearly stated assumptions, can offer more practical and interpretable answers than complex models that are poorly understood or inappropriately applied. While advanced models have their place, relying on them without a solid grasp can undermine clarity and confidence in results. The presentation will introduce a clinician-centered workflow that prioritizes shared understanding—co-defining exposures, outcomes, and covariates before any modeling begins. Through two case studies (a survey dataset and a meta-analysis), we'll explore how "less is more" leads to faster, clearer, and more trusted results in real-world setting.



Session Info

S409

CAREER OR PROFESSIONAL DEVELOPMENT SESSION

October 14th, 22:00- 23:00 UTC

Women in Statistics: Contributions and Challenges in Argentina

USE SCHED TO ACCESS THE ZOOM ROOM

¿Qué importancia tiene la estadística en distintas áreas del conocimiento? ¿Qué rol han tenido y pueden tener las mujeres en este campo?. Estas preguntas serán el punto de partida de una mesa redonda con cuatro investigadoras argentinas que se han destacado en la estadística desde diferentes disciplinas: Hidelgart Ahumada, Mónica Balzarini, María del Pilar Díaz y Marta Quaglino. Moderada por María Inés Stimolo, la sesión propondrá un espacio de diálogo en el que las invitadas compartirán sus trayectorias profesionales, las motivaciones que las impulsaron a dedicarse a la estadística, los desafíos que enfrentaron y las estrategias que les permitieron sobresalir en sus ámbitos. Asimismo, reflexionarán sobre el presente y las oportunidades futuras para las mujeres en estadística y en ciencia de datos. La sesión busca inspirar y animar a más mujeres latinoamericanas a descubrir en la estadística y la ciencia de datos una herramienta poderosa para desarrollarse profesionalmente y abrir nuevas oportunidades.

What is the importance of statistics across different fields of knowledge? What roles have women played and can continue to playing this discipline?. These questions will serve as the starting point for a roundtable discussion with four Argentine researchers who have made significant contributions to statistics from diverse disciplines: Hidelgart Ahumada, Mónica Balzarini, María del Pilar Díaz, and Marta Quaglino. Moderated by María Inés Stimolo, the session will provide a space for dialogue in which the invited speakers will share their professional journeys, the motivations that led them to pursue statistics, the challenges they faced, and the strategies that enabled them to excel in their fields. They will also reflect on the present and future opportunities for women in statistics and data science. The goal of this session is to inspire and encourage more Latin American women to discover statistics and data science as powerful tools for professional growth and for opening new opportunities.

ORGANIZER: María Inés Stimolo, Universidad Nacional de Córdoba

CHAIR: María Inés Stimolo, Universidad Nacional de Córdoba



Speaker Bios

S409



PROF. HILDEGART AHUMADA

National Academy of Economic Sciences of Argentina

<https://anceargentina.org/>

President of the National Academy of Economic Sciences of Argentina. Past President of the Argentine Economic Association. Fellow of the John S. Guggenheim Memorial Foundation. Visiting research fellow of Nuffield College (University of Oxford). Research fellow and former professor of econometrics at the School of Government of the Di Tella University. BA in Economics (with honors) University of La Plata. M Phil in Economics (University of Oxford). Former Head of the Economic Research Department at the Central Bank of Argentina. Her research interest covers: automatic selection methods, time series, panel data, climate econometrics, forecasting and applied economic modelling.



PROF. MARÍA DEL PILAR DÍAZ

Universidad Nacional de Córdoba

https://www.researchgate.net/profile/Maria-Del-Pilar-Diaz?ev=hdr_xprf

María del Pilar Díaz is Professor Emeritus of Statistics at the National University of Córdoba. She holds a PhD in Statistics and Agronomic Experimentation from the University of São Paulo (Brazil) and completed a postdoctoral fellowship at the University of Milan (Italy). She was a principal researcher at CONICET and led the Environmental Epidemiology Group on Chronic Diseases. Her career combines methodological development in statistical models with applications in the epidemiology of non-communicable diseases, particularly cancer. She has published numerous scientific articles and is internationally recognized in the fields of biostatistics and environmental epidemiology.



PROF. MÓNICA BALZARINI

Universidad Nacional de Córdoba

<https://bicyt.conicet.gov.ar/fichas/p/monica-graciela-balzarini>

Mónica Graciela Balzarini is an Agricultural Engineer from the National University of Córdoba (UNC), Argentina, with a Master's in Biometry from the University of Buenos Aires and a Ph.D. in Applied Statistics from Louisiana State University, USA. She is currently Director of the CONICET Scientific and Technological Center (CCT) in Córdoba, Vice-Director of the Unit of Phytopathology and Agricultural Modeling (UFYMA), Principal Researcher at CONICET and Professor of Statistics and Biometry at UNC. She was Director of the Master's Program in Applied Statistics from UNC. She has made significant contributions to statistical software development (InfoStat and InfoGen). She is author of peer-reviewed articles, she has supervised numerous undergraduate and graduate theses. Her research focuses on statistical modeling, multivariate methods, and machine learning applied to genomic and spatial data analysis. and is widely recognized for her contributions to biostatistics and agricultural science.



PROF. MARTA QUAGLINO

Universidad Nacional de Rosario

<https://www.researchgate.net/profile/Quaglino-Marta>

Marta Quaglino is a Statistician, graduate of the National University of Rosario, and holds a PhD in Statistics from the Polytechnic University of Valencia, Spain. She was Full Professor in the Bachelor's Degree in Statistics from 1984 to 2023 and she is a Category I researcher in the National Incentive Program. She directed the Department of Statistics (2003–2011) and has led the PhD Program in Statistics since its creation in 2013, the first in the country accredited by CONEAU. Her research interests include multivariate analysis, biostatistics, and statistical methods for continuous improvement processes. She has published in national and international journals and has actively participated in conferences, research projects, and the training of new researchers.



Sponsors



The 2025 IDWSDS Organizing Committee gratefully acknowledges our sponsors and thanks them for their generous support of this year's program. Their contributions make it possible for us to deliver such an outstanding conference. We also extend our sincere appreciation to the 2025 Friends of IDWSDS for their valuable role in making this year's symposium a success.

Platinum Level



Silver Level



Thank you for joining us at this year's conference! We truly appreciate your contributions in making this a day possible. The organizers, presenters, and volunteers are the heart of IDWSDS! We enjoyed 24 hours filled with inspiration and encouragement. We hope the energy and momentum of our celebration will carry on throughout the year. The accomplishments, passion, and global impact of women in statistics and data science is strong! We look forward to having you with us again next year—mark your calendars for October 13th!

