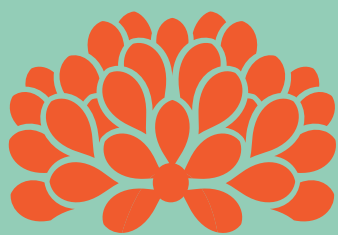# *JASA*
Partners with Authors to Enhance Reproducibility

Also:

New Member Spotlight: Stephen Lewis

Reflections on Volunteering with the Biopharmaceutical Section

PORTLAND
OREGON
JSM
AUGUST 3–8, 2024

**STATISTICS AND DATA SCIENCE:
INFORMING POLICY** *AND*
COUNTERING MISINFORMATION

Register Today!

August 3-8
**2024 Joint Statistical
Meetings**
Portland, Oregon

ww2.amstat.org/meetings/jsm/2024

KEY DATES FOR ATTENDEES

June 4 – July 1, 2024
Regular Registration

July 2 – August 8, 2024
Late Registration

July 5, 2024
Housing Deadline

AMSTATNEWS

JUNE 2024 • ISSUE #564

The American Statistical Association is the world's largest community of statisticians. The ASA supports excellence in the development, application, and dissemination of statistical science through meetings, publications, membership services, education, accreditation, and advocacy. Our members serve in industry, government, and academia in more than 90 countries, advancing research and promoting sound statistical practice to inform public policy and improve human welfare.

features

columns

This column is written for those interested in learning about the world of Data for Good, where statistical analysis is dedicated to good causes that benefit our lives, our communities, and our world. If you would like to know more or have ideas for articles, contact David Corliss at *davidjcorliss@peace-work.org*.

STAT*tr@k* is a column in *Amstat News* and a website geared toward people who are in a statistics program, recently graduated from a statistics program, or recently entered the job world. To read more articles like this one, visit the website at *http://stattrak.amstat.org*. If you have suggestions for future articles, or would like to submit an article, please email Megan Murphy, *Amstat News* managing editor, at *megan@amstat.org*.

ASA Community

New ASA Community

Make sure to log in and check out the new features on the ASA Community. The ASA's members-only online network is the place to go for news, opportunities, and—of course—community. Get involved, start a conversation, share your expertise, or ask a question. Start today at *https:// community.amstat.org.*

Women in Statistics and Data Science

Women in Statistics and Data Science Conference

The 2024 Women in Statistics and Data Science Conference—to be held in Reston, Virginia, October 16–18—will draw hundreds of attendees to celebrate the achievements and career interests of women in statistics and data science. Senior, mid-level, and junior stars representing industrial, academic, and government communities will unite to present their life's work and share their perspectives on the role of women in today's statistics and data science fields. WSDS will also mark its 10th anniversary of empowering, challenging, and inspiring this vibrant community. Learn more at *ww2.amstat.org/ meetings/wsds/2024/index.cfm.*

member news

This month's Top 10 is the 'Top Ten Statistics and Data Science Terms That Sound Like They Could Be in the Olympics'... Page 40

Connect with us on LinkedIn
www.linkedin.com/company/american-statistical-association---asa

Join the ASA Community
http://community.amstat.org

Like us on Facebook
www.facebook.com/AmstatNews

Follow us on Instagram
www.instagram.com/AmstatNews

Subscribe to our YouTube channel
www.youtube.com/user/AmstatVideos

View *Amstat News* online

Telling Our Stories of Innovation and Impact

Friday afternoon of the April ASA Board of Directors meeting culminated in a LinkedIn Live event. Joined by International Prize in Statistics winner Nan Laird, the board premiered the first video from the Telling Our Stories series. Statistician and data scientist Maria Cuellar of the University of Pennsylvania Department of Criminology shared how—in partnership with the Innocence Project—she applies statistical thinking on errors and uncertainty to forensic science with the goal of reducing false convictions.

The video project—launched with Nan's financial support—aims to highlight the profound impact of statistics on daily life, inspire a greater appreciation for its contribution to advancing science and enriching society, and raise awareness of the *transformative power* of our profession.

This innovative project is exciting on multiple fronts. First, it's well aligned with the ASA's mission to promote the practice and profession of statistics. Second, it exemplifies the partnership model between a sponsor (Nan) and the ASA for a common purpose. Third, it addresses an aspiration I have shared before: our responsibility to communicate our IMPACT.

In "Questions Answered: Bonnie Ghosh-Dastidar," which appeared in the February 2023 issue of *Significance* magazine, I noted I have often found fellow statisticians to be humble, conscientious, and happy to work in the shadows. They devote little effort to highlighting their important contributions.

Another major barrier to increasing visibility is the nature of what we do: Our work is often complex, and it can be hard for nonstatisticians to understand or appreciate. *We need to communicate our impact in ways nonstatisticians can appreciate—to tell the stories of how statistics affects the lives of real people every day.*

Redefining Impact: Why Traditional Metrics Don't Tell the Whole Story

Even as we find ways to increase our visibility by communicating our impact, we need to consider how we measure and maximize that impact. We can start by examining current definitions of and assumptions about what impact means. *As a discipline, what do we perceive as impactful? What is incentivized or rewarded?*

Traditionally, we have measured impact with quantifiable metrics such as grant funding and number of peer-reviewed publications. Obviously, these remain important. However, in recent conversations with statisticians and data scientists, I have often heard traditional measures might not fully capture the breadth and depth of an individual's influence. In an era of mushrooming data, nuanced collaboration across disciplines is needed to address society's greatest challenges. Statisticians and data scientists are positioned to play a central and essential role in this collaboration. Our work contributes to analysis in nearly every scientific field.

A major disadvantage of traditional metrics is they often favor individual achievement over team-based work; they may also fail to fully acknowledge the contributions of those who collaborate behind the scenes. A junior researcher who contributes significantly to a groundbreaking project might see their impact diminished by an authorship hierarchy. Educators who inspire their students and create positive learning environments might struggle to quantify these invaluable contributions. In addition, the contributions of those in nontraditional academic roles or industry scientists may not be sufficiently recognized within the statistical community. Redefining impact will require a shift in perspective. *Recognizing the value of collaboration is an essential step toward developing metrics that provide a more holistic picture of an individual's contribution.*

Madhumita (Bonnie) Ghosh-Dastidar

Stories of Exceptional Innovation and Impact

Former ASA president Karen Kafadar mentioned she chose statistics "because it deals with real problems where statistics can make a difference" in her 2019 address at the Joint Statistical Meetings. In fact, this is why I chose to work in public policy. A consistent focus of my work has been to understand the mechanisms that drive health and social inequality. Inspired by the Telling Our Stories series and reflecting that focus, I want to recognize the following inspiring examples of statisticians and data scientists serving society and helping to improve lives.

Measuring Inequality and Discrimination

Why? The Affordable Care Act requires all federal data collection efforts to include information about race, ethnicity, sex, primary language, and disability status. The gold standard for race and ethnicity measurement is self-report, but missing or poorly measured race and ethnicity data can create barriers to monitoring and improving quality, coverage, cost, and access.

How? A statistical model to measure racial and ethnic disparities was developed by RAND statistician Marc Elliott and colleagues. The Bayesian Improved Surname and Geocoding (BISG) family of algorithms, which use data from the US Census Bureau and other sources, can be used in any data set with name and address to measure and compare racial and ethnic groups. These algorithms incorporate census surname and residential address data—and first name data and Medicare administrative information in some versions—to produce a vector of six probabilities of being Hispanic; non-Hispanic White; Black; Asian American, Native Hawaiian, and Pacific Islander; American Indian and Alaska Native; and Multiracial. The BISG achieves concordance of 92–98% for Asian American, Native Hawaiian, and Pacific Islander; Black; Hispanic; and White people and does a little better with first names data.

The Medicare BISG (MBISG) is a specialized version that incorporates additional Medicare data to improve administrative race and ethnicity measurement for Medicare beneficiaries. It requires Medicare data and has concordance of 96–99% for Asian American, Native Hawaiian, and Pacific Islander; Black; Hispanic; and White people.

Impact: These methods are widely used in a variety of settings by federal and state governments, researchers, and commercial interests. For example, they are used by the Centers for Medicare and Medicaid Services to identify and address racial and ethnic disparities in the clinical quality of health care in the Medicare population. They also help ensure equitable algorithms for health care decision-making.

Although this algorithm was devised for health care research, the Consumer Financial Protection Bureau used it to underpin racial discrimination allegations against auto lending companies (e.g., the former General Motors lending arm Ally Financial, which paid $80 million to settle in 2013). "There's inherently some creativity involved," Elliott said. "The challenge is to take a complex problem in the real world and figure out the parts you can translate into the realm of numbers."

Algorithms for Human Rights

Why? Each year, an estimated 27–46 million individuals worldwide are held in modern slavery, generating annual profits of $30–$50 billion. "Trafficking is one of the most pernicious human rights and global health problems," said Victoria Ward of the Stanford Human Data Trafficking Lab. Yet little is known about how the market for human trafficking works—and thus how policies might prevent it.

How? An interdisciplinary team at the Stanford Human Data Trafficking Lab—in partnership with Brazilian federal prosecutors—has developed a human trafficking data repository that integrates existing, disparate administrative data sets to gain an understanding of human trafficking markets and the impact of policies.

For example, Stanford statistician Mike Baiocchi and his team of analysts have developed a decision-support tool for targeting labor trafficking. The architecture is based on a data-processing pipeline that transforms a constant flow of satellite imagery, along with incoming trafficking clues from varied sources, to produce predictions of actionable trafficking risk. Methods include machine learning from artificial intelligence, statistics, and algorithms that try to identify and disrupt the corporate structures of those profiting from the exploitation of human beings.

Impact: The problem is massive. According to Walk Free Global Slavery Index, more than 1 million people are held in conditions of modern slavery on any given day in Brazil. Using advanced AI methods, Baiocchi and his team have helped locate labor camps in the Brazilian arc of deforestation and worked with law enforcement to

safely remove exploited workers. This work also requires engagement with policymakers and frontline actors to identify the most effective solutions on a global scale.

After two rounds of successful piloting—during which the algorithms were used to locate and aid in field operations to rescue exploited workers—the lab is applying rigorous causal methods to assess the real-world impact of their tools in high-stakes investigations by Brazilian federal prosecutors. "… [W]e're listening and learning from how the prosecutors do their jobs," Baiocchi said, "and we're enabling machines to figure out a way to comb these huge data sets and automate and prepare insights for the prosecutors, to help prosecutors get into the field faster."

Improving Public Trust in the US Census

Why? Mandated by Article 1, Section 2 of the *US Constitution,* the decennial census is how we count every resident in the United States. This is extraordinarily important because census results are used for apportionment, redistricting, and distribution of billions of dollars to states, counties, and communities. The 2020 census seriously undercounted the number of Hispanic, Black, and Native American residents while overcounting White and Asian American residents. Such inaccuracies have reduced public trust in census data.

How? In a blog post, Census Bureau Director and statistician Rob Santos described bureau-wide efforts to "include culturally diverse training and educational materials, and tailored, unique cultural approaches with historically undercounted populations" in the next census. The bureau is investigating how training helps field staff develop cultural competencies. The new outreach efforts for undercounted populations are to address issues encountered in the 2020 census. The undercounting of minority populations resonates with Rob, a Latino who is the first person of color to hold the top Census Bureau post. His task is to rebuild battered public trust in the census and achieve a more accurate count.

Impact: To articulate the impact of Rob's work, I went back to a President's Corner column in which he wrote, "I have often used a strategy I call 'growing leaders.' … I partner with a junior staff person to author a blog or bid on a project where we are co-principal investigators. I provide technical/statistical expertise, and my junior colleague provides substantive expertise. Over the

Stanford statistician Mike Baiocchi at the ovens (kiln) in Brazil rainforest where trees are burned (for deforestation)
Photo by Luis Fabiano de Assis, Brazilian federal prosecutor

course of the project, I create a 'space' and provide the nurturing that allows my colleagues to take the lead in every sense of the word (budget, schedule, project team supervision, client contact, analysis, critical thinking, report writing). My professional reward is seeing those with whom I work blossom into leaders and scientists." This legacy of helping is truly a measure of impact.

These are only three examples of individuals who are using statistics and data science to make the world a better and more equitable place. But what these profiles clearly show is—if we want to demonstrate our impact—we need to broaden our definition of it by developing metrics to recognize our contributions to team science and real-world applications, incentivizing creativity and innovation to adapt methods to solve real problems, highlighting leadership as impactful, rewarding service on panels/committees, recognizing mentoring efforts to build the pipeline, and encouraging communications targeted to a nonstatistical audience.

Let's tell our stories, so others will know the value of calling on professional statisticians and data scientists to drive discovery and inform decisions.

M. Ghosh Dastidar

Highlights of the April 19–20, 2024, ASA Board of Directors Meeting

On Friday, April 19, ASA President Bonnie Ghosh-Dastidar gaveled to order the first 2024 meeting of the ASA Board of Directors at the ASA headquarters in Alexandria, Virginia. Highlights of the meeting follow.

Actions

• The board approved the following editor appointments for three-year terms beginning in 2025:

- *Journal of the American Statistical Association Applications* and Case Studies and Coordinating Editor: Hongtu Zhu, University of North Carolina

- *Journal of Statistics and Data Science Education* Editor: Juana Sanchez, University of California at Los Angeles

- *Journal of Computational and Graphical Statistics* Co-Editors: Laura Marie Sangalli, Politecnico di Milano, Italy, and Yuguo Chen, University of Illinois at Urbana-Champaign

- *Journal of Business & Economic Statistics* Co-Editors: Dacheng Xiu, University of Chicago; Michal Kolesár, Princeton; and Yingying Fan, University of Southern California

- *Technometrics* Editor: Kamran Paynabar, Georgia Institute of Technology

- *Journal on Uncertainty Quantification* ASA Editor: Bani Mallick, Texas A&M University

• The board recommended a change to the ASA bylaws for situations in which the past president position is vacant. See sidebar.

• The board selected Chicago as the site for JSM 2027, subject to final negotiations with the city.

• The board approved the creation of an online-only, open-access journal to be titled *ASA Discoveries.*

Telling Our Stories Premiere

Nicole Méndez, criminology student at the University of Pennsylvania, helps Maria Cuellar—assistant professor of criminology, statistics, and data science—decipher forensic data in the Telling Our Stories premiere.

The first video in the Telling Our Stories series premiered April 19 on LinkedIn and highlighted the profound impact statistics has on forensic science. The video features Maria Cuellar, assistant professor of criminology, statistics, and data science at the University of Pennsylvania.

Each video in the series will tell the story of how statisticians and data scientists are advancing science, informing public policy, and contributing to a world in which decisions are data-driven.

Watch the premiere on YouTube at *bit.ly/4dmf7Og*.

Bylaw Changes Recommended

The ASA Board of Directors recommends a change based on its experiences filling vacancies in the past president position.

Three years ago, 2020 ASA President Rob Santos was unable to serve as past president because he was appointed director of the US Census Bureau. And early in 2024, 2023 ASA President Dionne Price died. The board observed that, in the event of a vacancy in the past president position, it would be helpful to be able to call on recent former presidents to assist.

The proposed change is to a portion of Article V, Section 3, of the bylaws. Specifically, the proposal is to add the sentence in italics:

> If a vacancy occurs in the office of Past President, the office will remain vacant for the remainder of that term and the Executive Committee will designate board members to fulfill the ex officio responsibilities of the Past President. *At its discretion, the Executive Committee may designate a former ASA president to fulfill some or all of the ex officio responsibilities of the Past President, including serving on the Board of Directors.*

If you have comments about the proposed change, please email ASA Executive Director Ron Wasserstein at *ron@amstat.org.* The comment period will be open through August 15.

• The board made minor changes to the naming and renaming policies for ASA awards.

• The board replaced a 2016 ASA resolution regarding meeting locations with a new resolution. Details can be found at *https://community.amstat.org/deiba/meetings-information.*

• The board endorsed a statement on data science in pre-K–12 with associations representing teachers in computer science, mathematics, physical sciences, and social sciences.

• The board endorsed the recommendations of the ASA-commissioned report *The Nation's Data at Risk: Assessing the Health of the Federal Statistical Agencies.* The report—funded by the Alfred P. Sloan Foundation, ASA, and George Mason University—is to be released this month.

Reports and Discussions

• The board hosted a LinkedIn Live event launching the Telling Our Stories series, premiering the first video. Nan Laird, whose donation made the series possible, joined the board for the launch. View the video at *bit.ly/4dmf7Og.*

• The board heard and discussed the report of the Membership Council. The committee councils keep the board and its committees connected with each other. The report reflected tremendous activity across the committees of the council over the last 12 months.

• Derek Curtis II, ASA director of finance and administration, updated the board on the status of the ASA's 2023 year-end financials. The year ended in the black due to a good recovery of our investment portfolio. Net expenses and revenues followed close to the budget.

2024 Board of Directors

Bonnie Ghosh-Dastidar
President

Ji-Hyun Lee
President-Elect

Nick Horton
Third-Year Vice President

Jenny Thompson
Second-Year Vice President

Susan Paddock
First-Year Vice President

Kendra Schmid
Third-Year Council of Chapters Representative

Melinda Holt
Second-Year Council of Chapters Representative

Tomi Mori
First-Year Council of Chapters Representative

Michelle Shardell
Third-Year Council of Sections Representative

Jana Asher
Second-Year Council of Sections Representative

Jennifer Parker
First-Year Council of Sections Representative

Ingrid Van Keilegom
International Representative

Antje Hoering
Publications Representative

Jean Opsomer
Treasurer

Ron Wasserstein
ASA Executive Director and Board Secretary

• ASA Treasurer Jean Opsomer reported on the ASA's investments. The committee reviewed and updated its modeling of long-term results from investment strategies based on various spending levels. The analysis suggested the ASA's investment portfolio should retain sufficient funds to support its programs, members, and mission in the next 10 years, should spending be managed well.

• The presidents updated the board on their strategic initiatives. Bonnie Ghosh-Dastidar, 2024 president, outlined her plans to date, focusing on artificial intelligence and data science, community, and mentoring. President-Elect Ji-Hyun Lee's work will center on "building strong bridges," which she hopes to do by fostering opportunities, growing visibility, and diversifying membership.

• Amanda Malloy, ASA director of development, announced the launch of the "Honor a Colleague" program, providing a simple way to publicly acknowledge individuals through donations in tribute to or in memory of the person.

• The board reviewed its JSM site rotation and selection policy with an eye to expanding the number of potential JSM cities.

• The board heard a report from the ASA's commercial realtors about the status of the real estate market in Old Town Alexandria. The ASA headquarters is on the market with the idea of moving to a smaller space, but there is no hurry to sell.

• The board conducted a morning-long discussion of the association's goals in membership retention and recruitment.

The board will meet next on June 14 in a virtual meeting to discuss the ASA budget. The board will then meet in person August 2–3 in Portland, Oregon, before the start of JSM. ∎

Caucus to Host Third International Day of Women

Dong-Yun Kim and Jessica Kohlschmidt,
Conference Co-Chairs

The Caucus for Women in Statistics and Data Science will host the third annual International Day of Women in Statistics and Data Science on October 8. The theme this year is "Empowering the Next Generation of Women Statisticians and Data Scientists."

The event highlights the need for more opportunities, mentorship, and recognition for women. The aims of the virtual conference continue to be the following:

- Showcasing women and their contributions to the statistics and data science fields
- Connecting female statisticians and data scientists around the world
- Encouraging collaborations among global statistical societies
- Increasing diversity and inclusivity in statistics and data science
- Bridging the fields of statistics and data science

There will be both live and recorded presentations. Session ideas can be submitted at *www.idwsds.org/join-the-program*.

Additionally, sponsorship opportunities are available at *www.idwsds.org/sponsorship*. Supporting the International Day of Women in Statistics and Data Science helps pave the way for innovation, collaboration, and empowerment within the global community.

Last year, the International Day of Women in Statistics and Data Science celebrated knowledge, diversity, and empowerment, uniting professionals from 22 countries across six continents. Seventy-two speakers delved into topics ranging from the history of women in statistics to the intricacies of classical and Bayesian statistics while 980 registrants from 60 countries exchanged ideas and forged connections.

For more information, visit the conference website at *www.idwsds.org* or send an email to *idwsds@cwstat.org*. Register for the celebration at *www.idwsds.org/registration* and follow updates on Twitter/X at *https://twitter.com/cwstat.* ∎

ASA/AAAS Mass Media Fellow Hopes to Address Disparities in Health Care

ASA/AAAS Mass Media Fellow and Duke computer science major Freya Gulamali will spend 10 weeks this summer as a science journalist with *The News & Observer* in Raleigh, North Carolina. Long interested in science and journalism, Gulamali sees science journalism as a powerful mechanism to promote the perspective of

Gulamali

patients in the age of artificial intelligence, health, and health care, as well as a driver of accountability. She considers the next few years pivotal as policymakers determine how to balance the interests of insurance companies, health systems, physician groups, and drug manufacturers and wants to help ensure the patient perspective is integrated into policymaking.

Gulamali has interned with the US Food and Drug Administration Office of Generic Drugs Topical and Transdermal Team and currently works with the Duke Institute for Health Innovation to, for example, develop an AI tool to determine the risk of post-operative outcomes such as mortality and stroke.

Supporting programs such as the ASA/AAAS Mass Media Fellowship helps the ASA realize its vision of a world that relies on data and statistical thinking to drive discovery and inform decisions. Beginning in 2017, the ASA's sponsorship has allowed the program to expand its efforts to promote statistical capacity in reporting and provide statisticians with more media experience.

Previous fellows include Nick Thieme, Irineo Cabreros, Diana Cai, Jessica Craig, Aparna Nathan, Jayati Sharma, and Lucy Tu. A call for 2025 fellows will be issued this fall. Learn more about the ASA/AAAS Mass Media Fellowship at *www.aaas.org/fellowships/mass-media.* ∎

MORE ONLINE
To learn more about Freya Gulamali, visit *https://bit.ly/3UCGf2Y.*

Welcome TO OUR NEWEST MEMBERS

Aaron Q. Hawley

Abigail Nguyen

Adel Ahmadi Nadi
University of Waterloo

Adewale Folaranmi Lukman,
University of North Dakota

Aditi Chegu

Aidan John Crossler

Akiko Okamoto
Janssen Research & Development

Alaina Fueda

Allison L. Martinie

Amina Ali Hussein

Amira Carmen Colon

Amirul A. Anuar
Connecticut Children's

Amish Mishra
Taylor University

Amy Moore

Anand Paul
Louisiana State University Health Science Center

Ananda Sen
University of Michigan

Anandamayee Majumdar
San Francisco State University

Anas Fathul

Andrea L. Maes
Insmed Incorporated

Andrew J. Sage
Lawrence University

Andrew P. Bray
University of California at Berkeley

Anjana Grandhi
Merck & Co.

Anyesha Ray
Iowa State University

Anzhelika Belozerova

Apoorva Goyal

Apoorva Salvi

Arazi A. Lubis

Arnav K.

Arnold Kamis

Astari Raihanah
Statistics Indonesia

Austin Ellingworth

Azeem Shaikh

Bella Qian
Harvard University

Benjamin Jacobs
Iowa State University

Benjamin Lee Messer
US Energy Information Administration

Bill Rand

Bing Miu

Biviana Marcela Suárez Sierra

Brandon Fletcher

Brandon Roger Grossardt
Mayo Clinic

Brant Imhoff

Brian Liu

Cameron McPhee
SSRS

Carina Pleier
Ludwigs-Maximilians-Universität

Carlene Brown

Carvis C. Durr
University of Central Florida

Catharine B. Stack

Chaeryon Kang
University of Pittsburgh

Charles Reed McGhee
Affirmity

Chelsea K. Pipkin

Chris Patterson
ITW

Christi Lee

Christina M. Ramirez Kitchen
University of California at Los Angeles

Christine Watters
Novartis Institutes for Biomed Research

Codrin Iftode

Colin Mark Miles

Connor Celum

Connor Mizner

Cooper Harris

Courtnay Dollinger

Curtis Tatsuoka

Daisy Young

Daniel Alhassan

Daniel Esteban Torres Espinoza
SF Consulting 33

Daniel Lewis Sussman
Boston University

Daniel Sperber

Danielle DaSilva

Danielle LaVine
University of California at Los Angeles

Dat Huynh
Boston Scientific

David Ayikambey
Iowa State University

David Jerome Stokes
North Carolina State University

Davis Jobe

Debosmita Kundu
Iowa State University

Dihan Su
Iowa State University

Dixin Shen
Gilead Sciences

Don Ramesh Tholkage
University of Louisville

Donnel K. Cuffie

Margo Moreno

Edward Thomas Gunning
University of Pennsylvania

Efrain Garza
The University of Texas at San Antonio

Egypt Frye

Elaine Spiller
Marquette University

Ellis McLarty

Emily Chang

Emily Simons

Emma Burgin

Eric Herrison Gyamfi

Eugene Cheung

Frances Dean
US Army Acquisition Support Center

Fredrick Osei
Iowa State University

Fu-Wen Liang Kaohsiung
Medican University

Gabriel Arthur
Iowa State University

George Rooney

Grant Cabrera
Iowa State University

Greg A. Norman
Westat

Guanhua Chen

Guodong Liu
Penn State University

Gwen Squires

Hamed Sanusi
Atlanta Public Schools

Hannah Burns

Hanxia Li

Hillary Samples

Hiroko H. Dodge
Massachusetts General Hospital

Holiday Grace Pettijohn

Holland A. Bill

Hong Wang
University of Pittsburgh Cancer Institute

Hudson Rha

Huiling Liao

Ian A. Lane

Chamari G. Idamawatta

Ilhan Yeniada

James D. Ashley
Government Accountability Office

James Stamey
Baylor University

Jean Mosquea

Jennifer Gulley

Jessica C. Aldous

Jessica Harriger-Lin
Western Illinois University

Jessica Y. Dressler
Syneos Health

Jiabu Ye

Jianzhong Hu

Jinee Hira
Lodgestar Investments

Jinmeng Bai

Jiyun Zu
Educational Testing Service

Joan Marie Thul

Jocelyn Langdorf

New Member Spotlight:
STEPHEN LEWIS

We're excited to unveil our latest series, which we introduce one of our newest members. This month, we welcome Stephen Lewis, who answered the following questions so we could get to know him better:

How did you become interested in statistics and/or data science?

I became interested in statistics through basketball. Through years of following the game closely, I became intrigued with the impact stats have on the game and teams' decision-making. I have seen many stats that are very useful and insightful but also many

Lewis

that are misleading. I always excelled at math but didn't want to pursue a math degree in college. When I learned about statistics, I immediately knew it was what I wanted to pursue.

What do you consider your dream job?

A basketball statistician for an NBA team.

What do you hope understanding statistics and/or data science helps you accomplish?

I hope to get involved in basketball analytics, which is something that is very competitive to get into. I believe there will never be perfect stats in sports, as many things are hard to quantify, but there is always room to get better. I feel there is a lot I can contribute.

Is there a particular group of statisticians you would like to reach out to you?

I would like to connect with statisticians in the field of sports (mainly basketball).

What is your favorite hobby?

Playing and watching sports—mainly basketball, but also football. I also enjoy running.

What is something you would like people to know about you that we haven't asked?

I am currently a junior at the University of Delaware and am pursuing my master's degree with my school's 4 + 1 program.

NEW MEMBERS

John Andrew Jones
US Census Bureau

John Bannister
Air Sciences

John F. McManamon
Devry University

John Bowman
Walmart

John O'Brien
Bowdoin College

Jonah Matthew Todd Geddes
Gilead Sciences

Jonathan James Lisic

Jonathan McCurdy
Mount St. Mary's University

Jonnagadda Dev Rao

Joseph A. Osafo

Joseph Zou

Joshua Zhong

Juan Diaz

Judith Badiane
Wright State University

Julia Costacurta

Jung-Jin Lee
Merck & Co.

Kai Zhao

Kailey Keegan

Kaizhou Lei

Katharina Kraft

Katharine Doyle
CANA

Katherine Nishimura

Kazeem Busayo Ogunsusi
Iowa State University

Keara Schmitt
Iowa State University

Keaton Merino

Keith Schleicher Zoetis

Kelli Nicole Kasper
University of New Mexico

Kenneth Simmons

Kenzie Hensel

Kevin Douglas Shank

Kevin F. Forbes
*Energy and Environmental
Data Science*

Kirk Hohsfield

Kohei Horinouchi
Waseda University

Koji Shimomura
West Chester University

Ksenia Zagorskaya

Kusha Nezafati
State Farm

Kyle Raymond
Leo Pharma

Laia Giralt Sole

Laleh Jamshidi
*University of
Wisconsin-Milwaukee*

Lawrence Tatum

Leah Isakov Seqirus

Lida Wang

Lin Chen
The University of Chicago

Lindsay Younis

Linh Tran

Liping Sun
Iowa State University

Lizzie Neumann
Helmut Schmidt University

Logan Stephens

Lori Beth Shelby

Madison Grieb

MaHalie Hamilton

Makayla Tinnes

Manan Saxena

Margie Bell

Marie Hardt
Iowa State University

Marieke Sorge
Sandia National Laboratories

Marina Reis Costa

Mark Isak Shteyn

Mark McCormick
*University of New Mexico Health
Sciences Center*

Martial Longla
University of Mississippi

Marvin Javier

Mary Henthorn

Matthew Bray

Matthew Guo
Erasca

Maximilian Matthe
Goethe-University Frankfurt

Maxwell Hope
US Census Bureau

Md Al Masum Bhuiyan

Md Milon Sarker

Megan Heyman
*Rose-Hulman Institute
of Technology*

Meiqi Liu
Michigan State University

Meng Liu

Mercedes Andrade Bejarano
Universidad del Valle

Michael Braun
SMU Cox School of Business

Michael Kojo Abalo

Michael MacKenzie
Virginia Housing

Michael Weylandt
Baruch College

Miguel de Carvalho
University of Edinburgh

Miles S. Chen
*University of California
at Los Angeles*

Millicent Berko

Min Wong
Iowa State University

Mingyue Hu

Mingze Zhang

Miriam Alvarez-Pintor

Monica Anang
Iowa State University

Morgan Mckiddy

Muhammed Musa

Muslihat Adejoke Gaffari

Mutiu Adebayo Adeniran
Nicoprime Feed, Food and Oil Mills

Nadeesri Wijekoon
FDA

Nancy E. Gove
Seattle Children's Research Institute

Natya Hans

Noah Leavitt
Iowa State University

Noha Ahmed Youssef

Nordia Diana
Marie Thomas

Nyo Hayden Serieux
*Sir Arthur Lewis
Community College*

Olivia Fang

Oloruntele Taiwo

Onno Jan Sharp

Owen C. Holliday

Pablo E. Baldivieso
Oregon State University - Cascades

Parker Gauldin

Pathum Randunu Nawarathna
*Missouri University of Science
and Technology*

Patricia Eckardt
*Good Samaritan
University Hospital*

Paul Engkasser
Iowa State University

Paul S. Albert
National Cancer Institute

Peter Solomon

Phil Greer Ariel
Precision Medicine

Phuong Hong Vu

Ping Zhao

Prabin Thapa
Mayo Clinic

Rachel V. Ball

Robert Neal Montgomery

Robert Stewart

Robert Weant

Rong Zhou

Rory Samuels

Ruba Shalhoub
National Institutes of Health

Ruijin Lu

RuoFei Yin
University of Pittsburgh

Ryan Abella

Ryan Berberek

Ryan Marangattu

Ryan Phillip Pollak

Ryann Stricker

Sahifa Siddiqua

Saif Hasan
Iowa State University

Sam O. Taiwo
Texas Tech University

Samipan Majumder
Iowa State University

Samjhana Sapkota

Samuel Tabiri
Iowa State University

Saskia Comess
Stanford University

Sattwik Ghosal
Iowa State University

Scott Yamamoto
University of Oregon

Sebastian McCrimmon
Iowa State University

Sejal Prachand

Shaista Nisar
University of Tampa

Shawn K. McCollum

Shirshendu Mukherjee
Google

Shiying Liu

Shota Furukawa
Keio University

Shuai Sun

Shuheng Kong

Sonia M. Lopez

Spencer Gordon Wadsworth
Iowa State University

Stefanee Tillman

Stephen P. Jones
The Boeing Company

Sucheen Sundaram

Sunday Osiebuni Aghamie
University of Northern Colorado

Susan Glenn

Taniya Sarkar

Tarun Balani
Hinge Health

Teresa Middleton

Tho Nguyen

Tiantian Zeng

Tingting Hu

Tingting Zhan
Thomas Jefferson University

Tingting Zhang

Tram Bao Huynh
North Carolina State University

Tran Quy Thien Pham

Vanessa Richardson

Victoria Okhomina

Wahyu Wibowo

Wan Yee Hing

Wei Jun Tam

Wenting Xu

Wesner Antoine

William C. Gillette
Ursinus College

William Clements

William Labadie

Wing Yan Yuen
*University of Maryland
College Park*

Xiansi Ma
Iowa State University

Xiaohong Huang
AbbVie

Xiaohua Liu

Xilong Chen

Xunhang Gao
Iowa State University

Yan Wang

Yan-Han Chen
Iowa State University

Yanzhao Wang

Yi Gong
SAS

Yi Liu
Boehringer Ingelheim

Yi Tang Chen
The Ohio State University

Yiheng Pan

Yiting Long
Westat

Yixin Chen

Yong Seok Park

Yu-Chih Chen
University of Pittsburgh

Yujun Wu
Morphic Therapeutic

Zac D. Rios
Utah State University

Zach Taylor

Zeynep Demir

Zhaowen Wang

Zhengzheng Tang
University of Wisconsin-Madison

Zhenyang Ren

Zhikang Dong

Zihang Chen

Zitong Wang

Ziyue Zheng

Zongyue Teng ∎

Unlocking Statistical Insights: NSF's Push for Mathematical Foundations of Digital Twins

To strengthen the connection between the statistical community and National Science Foundation, we continue the series introduced in the May 2023 issue that poses questions to NSF program officers and awardees. If you have questions or comments for the program officers, send them to ASA Director of Science Policy Steve Pierson at pierson@amstat.org.

This month's program officer is Yulia Gel from the Division of Mathematical Sciences in the NSF Directorate for Mathematical and Physical Sciences. The awardee responses are from Qing Mai of Florida State University.

Program Director

Yulia Gel is on leave from The University of Texas at Dallas. This is her third year as a rotator program director of the statistics program.

Digital twins research and NSF funding on digital twins—what is there for statisticians?

A digital twin is an emerging technology that allows for a dynamic virtual representation of various real-world physical systems, objects, and processes. One of the essential constituents of digital twins is the bidirectional connection with their physical counterparts. That is, digital twins can learn to synchronize and communicate with their physical counterparts by continuously exchanging information, using, for example, tools of artificial intelligence, or AI.

In particular, a recent report from the National Academies of Sciences, Engineering, and Medicine titled "Foundational Research Gaps and Future Directions for Digital Twins" defines a digital twin as "a set of virtual information constructs that mimics the structure, context, and behavior of a natural, engineered, or social system (or system-of-systems), is dynamically updated with data from its physical twin, has a predictive capability, and informs decisions that realize value."

To enhance adoption of the digital twins in real-world applications—from health care to civil engineering—we first need to ensure digital twins do indeed reliably represent their physical counterparts in a broad range of what-if scenarios.

For example, how can we construct the confidence intervals of digital twin outputs while accounting for uncertainties such as modeling, data, and process? How can we relate those confidence intervals to decision-making? Is it possible to develop verification, validation, and uncertainty quantification techniques adapting to the physical twin evolving over time? How can we detect and mitigate biases that may potentially be introduced in the digital twin models and associated AI-based tools? How robust are the digital twin models to missing data, outliers, irregularly spaced records, and fusion of observations at various scales?

Many, if not all, these questions cannot be addressed without proper statistical inference and sound statistical methodology, with the machinery ranging from causal inference to robust statistics to experimental design to extreme value analysis.

Recognizing these needs, NSF has released several new solicitations on mathematical and statistical foundations of digital twins. The Foundations for Digital Twins as Catalyzers of Biomedical Technological Innovation NSF 24-561 is a joint tri-agency initiative among the NSF, US Food and Drug Administration, and National Institutes of Health. It focuses on mathematical and statistical principles behind digital twins and synthetic data used in the evaluation of medical devices and the relevance of the developed models in addressing current and emerging challenges affecting the development and assessment of biomedical technologies. The deadline for this solicitation is June 21.

In turn, the Mathematical Foundations of Digital Twins NSF 24-559 is a joint program with the Air Force Office of Scientific Research and supports foundational mathematical and statistical research on digital twins in applied sciences, without focusing on a particular application. The deadline for this solicitation is June 20.

It is hard to overstate the role of statistical sciences in digital twins research, and it is an exciting opportunity for statisticians to lead interdisciplinary projects.

Awardee

Qing Mai is a professor in the department of statistics at Florida State University. She earned her PhD from the University of Minnesota in 2013, and her research interests include high-dimensional data analysis, tensor data analysis, and machine learning. She has been the principal investigator or co-principal investigator for two NSF grants from the Division of Computing and Communication Foundations in the Directorate for Computer and Information Science and Engineering. She has also served on multiple NSF panels.

Mai and her Florida State colleague Xin Zhang received $479,000 from the Communications and Information Foundations program for their proposal, "Cluster Analysis for Highly Correlated, Heavy-Tailed, and Higher-Order Data."

Tell us more about your project, including its motivation and goal.

Rapid advances in modern science and technology are resulting in the generation of data sets of unprecedented size and complexity. A common source of complexity in data sets is the presence of subpopulations. For example, a disease may have several subtypes, and customers may be attracted to different features of the same product. Cluster analysis is a popular tool to identify subpopulations, which affords a refined investigation on each of them.

This project aims to generate innovative clustering methods to understand the heterogeneity common in modern data sets. We answer the challenges of heavy tails, high correlations, and the tensor structure.

Such data sets frequently arise in contemporary scientific studies, but severely comprise classical clustering methods.

We propose a family of probabilistic models to accommodate such data sets, under which we develop model-based clustering methods.

In addition to the allocation of subjects, the methods in this research further find the defining features of each subpopulation. We pay special attention to efficient computation and rigorous theoretical study.

The research team will apply these methods to various real-world problems with the potential to affect multiple fields that rely on such data sets. Open source and user-friendly software will also be provided. Moreover, this project will be integrated with educational and outreach activities, including new courses, interdisciplinary training, and mentoring of underrepresented student groups in the mathematical and statistical sciences.

Describe your approach to the Division of Computing and Communication Foundations.

A critical task of applying outside the Division of Mathematical Sciences is to establish relevance to that specific entity. Some of my approaches toward this goal are the following:

1. For each potentially interesting solicitation, I read abstracts of recent awards to determine whether my research overlaps with the funded projects, and, if so, which of my projects are most suitable.

2. The guidelines of each solicitation are different, and I carefully follow them.

3. In writing my proposal, I emphasize the unique contributions of statisticians to the problem of interest, as well as pay special attention to the computational aspects and scientific applications.

What advice do you have for others applying for NSF funding?

In recent years, the connections are significantly strengthened between statistics and many other areas such as computer science and engineering. Consequently, statistical research is increasingly valued by researchers from such fields. I suggest statisticians actively explore funding opportunities outside the Division of Mathematical Sciences, which increases the impact and visibility of statistics, as well. ∎

Reflections on Volunteering with the Biopharmaceutical Section

Richard C. Zink, Principal Research Fellow, JMP Statistical Discovery

Zink

It's hard to believe nearly five years have passed since serving as the 2019 chair of the Biopharmaceutical Section. I'm happy to see the section is thriving and as busy as ever educating our members and providing them opportunities to collaborate with their peers.

Chair was a role I served toward the end of an active 10+ years with the section. Like many journeys, this one began small and grew over time as I took on more responsibility. To begin, I was a frequent member, starting in 2012, of the organizing committee of the then-called FDA-Industry Statistics Workshop. In other words, I submitted session proposals and worked to convince other organizing committee members, including members of the workshop steering committee, that my proposals were worthwhile and should be selected for the final program.

Despite not being a member of the steering committee, these journeys to Rockville, Maryland, to participate in the organizing committee meetings gave me a basic understanding of the inner workings of the workshop. I found it fascinating.

Responding to a call for volunteers at one of these meetings, I began my involvement with the section podcast that lasted until 2022. These initial footholds with the workshop organizing committee meetings and podcast led to other opportunities,

and these opportunities generally started off with "we need a volunteer to …."

Eventually, participating in organizing committee meetings allowed me to meet numerous steering committee members, which gave me the opportunity to join the workshop steering committee in 2014. I guess I must have done a reasonable job on the steering committee, as I was recommended to serve as industry co-chair for the 2015 workshop. I continued to serve on the workshop organizing and steering committees until 2018, though I took more of a supporting role over time.

Podcasting was a wonderful opportunity to engage the larger biopharmaceutical community and allowed me to engage a significant portion of the biopharmaceutical steering committee: members of the Publications Committee. This committee, run by the publications officer, helped organize and lead activities surrounding the *Biopharmaceutical Report*, webinars, podcasts, webpage, and social media. Because I co-led a key effort of the publications committee, I was able to attend Biopharmaceutical Section steering committee meetings to provide updates on podcasts.

I enjoyed steering committee meetings, as I had an opportunity to hear about ongoing and upcoming activities within the section. For example, there was a push in 2013–2014 to

organize scientific working groups. Given my research areas at the time, I became one of the charter members of the Safety Scientific Working Group in 2014, the first in a long line of scientific working groups for the section.

As time passed, I started thinking about potential opportunities to serve as a section officer. In 2016, I volunteered to become one of two candidates for publications officer. When the call came again two years later, I offered to become a candidate for section chair (and was somehow elected twice).

Section chair is an enormous three-year role shared by three individuals. Each year, the nature of the role changes, with the second year being the most visible. During the first year as chair-elect, you learn the responsibilities of becoming chair in the following year. In the second year as chair, you are responsible for organizing and leading the three section steering committee meetings. In the final year as past chair, your focus shifts to cleanup and documentation, most notably reviewing the manual of operations for accuracy and identifying a list of candidates for the upcoming year. The details for chair and other section roles are listed in the manual, which is posted on the section website.

My time on the steering committee as chair and publications officer lasted from 2016–2020, during which time I witnessed

many important initiatives and much growth within the section. The steering committee continued to review and approve requests for new scientific working groups, which quickly grew in number. In turn, these scientific working groups generated valuable outputs such as manuscripts, webinars, short courses, and books. Notably, the section extended its reach by producing content for YouTube.

A workshop taskforce was convened to address many of the operational challenges that began to plague the workshop, now called the ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop. I commend the steering committee for making this difficult decision, since a lot of thought and investigation were undertaken by the taskforce and there was insufficient time for thorough discussion during steering committee meetings.

We also formed a 40th Anniversary Committee to begin preparing to celebrate our longevity in 2021 and established the Leadership in Practice Committee, which aims to offer training in practical leadership to members of the biopharmaceutical community.

While all these initiatives were important, I think the most notable undertaking during those years was instituting the scholarship award. To date, more than 24 students have been awarded for their research and leadership. The cash award allows students to attend the Joint Statistical Meetings so they can engage with other statisticians and data scientists about their research and pick up new ideas.

As my term as chair came to a close, the world underwent drastic changes due to

COVID-19. While we had preliminary preparations in place, none of us could have anticipated how extensive or how long our virtual existence would become. For example, in the spring of 2019, we held the first fully virtual meeting of the steering committee. There were hiccups, but no one could have imagined we would be holding all our steering committee meetings—as well as entire conferences—virtually one year later. Further, due to the economic pressures of COVID, the owners of the Marriott Wardman Park building went bankrupt, robbing the workshop of its home for many years.

In some ways, the section was prepared for the loss of

our meeting place. One of the taskforce's responsibilities was to identify a larger space for the workshop since it had sold out for several years in a row and there was no space to grow.

And slowly, over the next two years, I completed my responsibilities for podcasting and the Safety Scientific Working Group and took a much-needed break. The time I spent contributing to the section and workshop was some of the greatest of my professional career, and I am grateful for the opportunities I had to serve the community and the trust others placed in me to serve it. I'm not completely retired, however. I hope to participate on the Leadership in Practice Committee soon. ∎

Interested in Volunteering?

I am occasionally asked how to get involved with the Biopharmaceutical Section. It may seem challenging. Even hitting the volunteer button on the website can result in a waiting period as steering committee members figure out how you can contribute. However, a few ideas include the following:

- Chairing a session at JSM or the ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop
- Discussing your research for a podcast
- Recording a short course video for the section YouTube channel
- Forming or actively contributing to a scientific working group
- Teaching a webinar
- Writing an article for the *Biopharmaceutical Report*

Visit the section website at *https://community.amstat. org/biop/home* to find contact information for those on the steering committee leading or chairing the activities you're interested in. These activities will provide visibility, which can lead to additional responsibilities.

If anyone asks for a volunteer, be the first to raise your hand and enthusiastically say, "I can do it," even if you have no idea how to do it. There will be time to figure it out. Twelve years ago, I had no idea how to plan or produce a podcast, but what an amazing journey that turned out to be.

HOSTS

Ron Wasserstein

Donna LaLonde

GUESTS

Tian Zheng

Hongtu Zhu

Mark Glickman

PODCAST

Practical Significance Take Two—
How Can I Help You Today?

This interview with Mark Glickman, senior lecturer on statistics at Harvard University; Tian Zheng, statistics professor and statistics department chair at Columbia University; and Hongtu Zhu, biostatistics professor at The University of North Carolina, was conducted by Practical Significance *co-hosts Donna LaLonde and Ron Wasserstein during a recent episode in which they discussed the implications of artificial intelligence, AI, for our profession.*

Wasserstein: How should the field of statistics evolve to meet the challenges of AI development and application?

Glickman: Recently, AI has been an area statisticians have been paying pretty close attention to. Over the last 10 years or so, there have been some incredible successes in AI. The methods that underlie a lot of these procedures, particularly in neural nets and deep learning, are a little mysterious.

On the one hand, there is a solid foundation for many of these methods, but on the other hand, it doesn't seem to correspond to things statisticians have been thinking very much about. One of the burgeoning areas in statistics and how it's connected to AI is understanding the underpinnings of these methods in AI, which just seem to work.

So, there are certainly many statisticians doing amazing work at the forefront of understanding many of the underpinnings of these mostly deep-learning methods that are driving technology in impressive ways. The other area of AI statisticians are involved with, not so much in a direct developmental way, is helping to be able to evaluate AI methods and work with essentially the output of AI methods to develop approaches that have a bit more familiarity among statisticians.

For example, the whole area of conformal inference is a really big area that's attracting attention among statisticians, mainly because conformal inference basically starts with what could be a black box, and then starts making predictive inference

without really even knowing very much about what generates observations. It's a very powerful tool statisticians are jumping onto and a good example of an area in which statisticians continue making some very important progress in parallel with the development of AI.

Zhu: For the field of statistics to move forward in the AI era, there are many things we need to consider. First, the current curriculum needs to be modernized to fulfill the evolving demands of modern data science. Our curriculum was mainly developed 30 or 40 years ago and would need some modification, particularly in areas of engineering capabilities, practical data analysis appearances, and proficiency in data mining techniques.

Also, existing evaluation systems need some changes. Based on my observations, we need to speed up the review process in our field and make our papers more acceptable to data scientists and practitioners. There needs to be a balance between theory and application. We also need to provide more opportunities to young rising stars and active researchers.

Additionally, the systems need to include all the data science–related journals and conferences, not just focus on statistical journals, because there are so many new journals and conference proceedings. We need to be open-minded. It is very important to encourage and promote greater participation by statisticians in various study sections and train them in effective communication and equitable contribution.

Zheng: I want to build on what Mark and Hongtu said. Mark covered the exciting research trajectories we are observing and Hongtu shared needed infrastructure changes. So, I'm going to address a 'cultural shift.' As a discipline, and as a community, we have come a long way from 15 years ago when big data first started drawing global attention. We now recognize and promote computing-intensive, applied, data-intensive applications; machine learning; and research more than we used to. Computing, machine learning, and data-intensive applications are three important pillars of AI.

Statistics had a slow start when big data first happened, but many departments came together and took on initiatives to make change happen. However, over the past few years, there has been a feeling that we are falling behind again. We're being left out of some of the AI conversations. This is because AI is moving at a much faster speed. A few technological breakthroughs in recent years enabled so many more fields to embrace AI, computing, and data, than in the previous big data era.

The wind is blowing stronger, and the statistics community needs to continue at what we do well, but with a more proactive effort to accelerate and encourage more energetic participation. Big data calls for collaboration, and statisticians can contribute by being effective collaborators. AI is different. It is much more interdisciplinary and transdisciplinary than data science. It calls for end-to-end solutions and team science. If no one in statistics is willing to be the pioneer to go into AI and lead AI applications, then our community will not be able to keep pace with AI research.

Wasserstein: What are the biggest challenges for statistics in the future of AI?

Zhu: There are two major challenges for statistics in the future of AI, beginning with a diminishing pool of students in statistics. Another challenge is about new AI-related funding opportunities.

Zheng: I believe the biggest threats to our discipline are the talent pipeline, faculty development, and research resources. As Hongtu said, if we're not proactively revising our curriculum to enable our students and young researchers, statistics will not be part of today's fast-moving national AI research effort. We have students and young researchers but our talent pipeline into AI has been drying up. Because of this, we are not competitive for AI resources. If this cycle continues, we need to worry for the next generation: whether we can continue to have generation and generation of bright and excited young statisticians identify as decision-makers in data-intensive applications, grounded in the central foundational principle of statistics, and at the same time be able to embrace research in AI.

Glickman: One of my concerns has to do with a lot of the tension that has been going on between statisticians and computer scientists in the area of machine learning—deep learning in particular, where all this development has been done in a way in which the statisticians have been struggling to be at the table. It's important to make clear to the computer science community that statisticians add value, so we're not left behind.

> Statisticians can be much more vocal about our expertise on uncertainty propagation, which most leaders in the field of AI may not be paying quite as much attention to.

Statisticians can be much more vocal about our expertise on uncertainty propagation, which most leaders in the field of AI may not be paying quite as much attention to. But it's so ingrained in the statistician mindset, particularly when we're working with enormous data sets being used for very personalized kinds of applications, like personalized medicine or personalized education. You do need to start worrying about the level of uncertainty in your conclusions, and that's something statisticians can very much help with.

LaLonde: What about AI are you most excited about?

Glickman: Probably like many other people, I woke up one morning in late 2022 with the news that ChatGPT was made available to the public. The reason I paid attention was I was reading an article entirely written by ChatGPT, and then I was told, "Yes, this is something you could play around with."

Even though I was certainly paying reasonable attention to what was going on in the world of AI, I was stunned at what some of these LLM-based generative AI algorithms could do. I'll start by just simply saying I'm very excited about the promise of generative AI and a lot of what it can do.

I regularly use generative AI to help me with my writing, at least at minimum as a way of proofreading my writing, because the language models are so good you can pretty much guarantee your writing is going to improve. I also use generative AI for coding tasks. So, if I need to implement something quickly, or I just don't want to spend a half hour writing something up, I'll just, in a sense, explain what I need to have coded up, and then I'll just get my answer immediately. So, it's a huge, huge time saver.

Zheng: I'm a data person, so I'm generally excited about cool and fun data sets people put out there. It used to be very hard to find a collaborator willing to share data. Now I have collaborators reaching out: "I heard AI and ChatGPT can help us analyze the data." So, I got very excited about this willingness to collaborate.

In the past, when I collaborated with a new discipline, I needed to learn how they organize their data and the special format their data was in. Over the past decades of big data research, we have done such a good job of educating people—and our collaborators have also been educating themselves in machine learning—today's data are becoming better formatted and collected using better research design.

In addition, the kind of tools Mark mentioned, such as ChatGPT, also make collaboration easier. I used to ask a lot of questions to my collaborator about simple definitions and jargon. Now I have ChatGPT as a tutor—a very patient tutor. I can simply ask anything and then really develop an appreciation for the background knowledge quickly. I'm most excited about the richness of the opportunity available to us, available to nearly anyone in statistics who wants to embrace collaboration in AI.

Zhu: We can modify and adapt many AI tools to our projects and expand our spectrum in various applications such as NLP [natural language processing] and image analysis. AI is revolutionizing these areas. Also, I do many NLP types of projects by using open AI, so you don't need to collaborate with NLP researchers, providing us with new opportunities.

We can integrate AI with many existing statistical methods for further method development. I combine the neural networks with quantile regression for many projects in the tech company in particular, such as the treatment effects and distribution or reinforced learning type of things. The use of quantile regression allows you to look at the problems much clearer.

We can also prove many more complex theoretical problems such as these pattern recognition problems. There are many challenges behind it. This is an opportunity for us, right? Let's create new models and tools. I'm excited about these new things. I don't feel that threatened. I want to embrace AI tools. ∎

'Arctic Statistician'

Maciej Chudek

An 'Aussie Bush Ballad' by an Aussie statistician living in the Canadian arctic *(best read in a heavy Australian accent)*

Way up past that there tree line
is where I correlate
out on the windy tundra
my identities equate

If I need a transformation
I'll derive it on the land
Those folks down in the cities
They wouldn't understand

Coz I'm not on some cozy tangent
absorbing midnight sun
I make my math the hard way
And I'll tell you how it's done

If I wanna' feed my family
I'd best get off my arse
Head out into the real world
Where the integers are sparse

I'll fuel up my binomial
Till the power's gettin' high
Then diverge across them frozen lakes
That expand right to the sky

My walks seem almost random
Across those snowy plains
Tracking caribou migrations
Like endless Markov chains

When I fin'ly trace their matrix
Their range in my domain
I'm determinant to get them
But I gotta' be restrained

So I'll estimate the distance
Of the group and of the field
I'll compute the discrete factors
That affect my hunting yield

Then I'll sight in my equations
I'll carefully weight my shot
And squeezing the divisor
I'll solve the bloody lot!

I'll pull out their inner products
Skin right round the outside
I'll quarter them and halve them
And make them into Pi

The northern lights above me
The air biting my face
Regressing and predicting
This glorious subspace

I'll tan the complex numbers
I'll scoop out both their Is
I'll dry 'em out and smoke 'em good
'till the real part simplifies

I'll track down hidden data
To burrows where they hide
I'll set my snares around them
Primed, ready to divide

And when they start converging
All mean and imprecise
I'll sum 'em up along that line
And clean 'em up real nice

'till the year runs its sequence
and the snow melts in the sun
Time harmonically progresses
But the math work ain't yet done

I gather up kurtosis
From boggy autumn stats
I cook up an ANOVA
Full of errors, meats, and fats

And in the summer season
I count the bugs that buzz
Their combinatoric nuisance
Won't bother me because

I stay up way past midnight
With the sun still shinning bright
Strung out, sleepless but striving still
For statistical insight

'till I find my polar axis
Spin my numbers round and round
Coordinate my limits
Get my feet back on the ground

I turn my thinking southward
Across the taiga plains
Infinity converges
'till one insight remains

These Arctic folk around me
Who greet me in the street
We're all in this together
And that means everything to me

So as the cold starts to confine me
And my thoughts all turn to goo
My fingers can't work door nobs
And my math starts failing too

But I'm no longer decreasing
Into deep winter despair
I can function in this cold null space
Coz there's folk up here who care

So that's my occupation
And where I make my mark
I don't need no damn chalkboard
In this endless winter dark

I'm no armchair academic
Who'll sit and postulate
Not some idle theoretician
Or their complex conjugate

I'm no second-rate derivative
or even third or fourth
I'm an Arctic Statistician
A math man, of the north

May Issue Includes Insights on Communicating Statistics

Sharing statistical stories with the world at large—through a book, a talk, a social media post, or the pages of your favorite statistics and data science magazine—can be arduous. No matter how riveting and relevant those stories, common parlance and complex mathematics are not easy bedfellows. You can't assume everyone's keeping up, but instinctively feel dumbing down is wrong. How do you strike the right balance, get your story noticed, and make your voice heard?

In this issue, three statistical communicators bring you their experiences and advice for communicating statistics to broad audiences—each in their own way—with the hope they will inspire you to capture your own audiences.

Additionally, a regular series called Bad Stats that includes salutary lessons from the world of statistics premieres in this issue.

May 2024 Highlights

The R Number
The second in our six-part series by Gavin Freeguard on the evolution and usefulness of the pandemic's famous epidemiological tool

What Is Data Science?
The burning question, inspected through a statistical lens

The Thirteenth Floor
Buildings in New York City often don't have a thirteenth floor—superstition, or something else?

The Philosophy of Statistics
Understanding the philosophical foundation of probability and statistics

Read the full issue at *https://academic.oup.com/jrssig/issue.* ∎

Checking In: LGBTQ+ Advocacy Committee

Suzanne Thornton, National Institute of Standards and Technology

Suzanne Thornton, chair of the ASA's LGBTQ+ Advocacy Committee, is a member of the Statistical Engineering Division at the National Institute of Standards and Technology. Her interests include statistical computing, the foundations of statistical inference, and the ethical application and development of statistical methods.

The ASA LGBTQ+ Advocacy Committee is hosting a webinar titled "An Interdisciplinary Look at Trans and Gender-Non-Conforming Inclusion in STEM" June 12 at 6 p.m. ET. It will feature Maggie Delano, associate professor of engineering at Swarthmore College; Sara Lipshutz, an assistant professor of biology at Duke University; and Jason Eridani, acting assistant director of the Australian Bureau of Statistics and chair of the Statistical Society of Australia's Equity, Diversity, and Inclusivity Committee.

The webinar is free and open to the public, but registration is required at *https://bit.ly/ LGBTQIA-Webinar*.

Members of the committee also plan to hold their business meeting during the Joint Statistical Meetings, which are taking place in Portland, Oregon, August 3–8. The meeting, which will be held in person and virtually, is open to anyone interested in the committee's mission. To join remotely, email Suzanne Thornton at *thornton.suzy@gmail. com* by August 1 with "Advocacy Committee Business Meeting" in the subject line. View the conference program at *https://bit. ly/3wx7mV4* for the meeting's time and location.

Additionally, the committee will have a table in the JSM exhibit hall, and members encourage everyone to stop by to meet them and friends, as well as pick up ASA Pride-themed swag.

Also at JSM, the committee will co-host a reception with the Committee on Women in Statistics, Committee on Membership Retention and Recruitment, and ASA book club to allow anyone interested to socialize and network with statisticians and data scientists of varied backgrounds and levels of experience. View the conference program for details.

To learn more and support the committee as it works toward justice, equity, diversity, and inclusion in all professions, visit *https://community.amstat. org/asalgbtq/home* and click the

ASA Pride Scholarship

The 2023 Pride Scholarship winners are **Erin Rose Lipman** from the University of Washington and **Sarah Buckingham** from The Ohio State University. Winners for 2024 will be announced during the awards ceremony at the Joint Statistical Meetings, to be held in Portland, Oregon, August 3–8.

The scholarship was established to raise awareness for and support the success of LGBTQ+ statisticians and data scientists and allies. Nominations (including self-nominations) for the 2025 scholarship can be sent via the form at *https://form.jotform.com/AmStat/asa-pride-scholarship* by March 31, 2025. To be eligible, the nominee must be enrolled in a statistics or data science graduate program or have completed a statistics or data science degree within five years of the award date. Also, the nominee must identify as an LGBTQ+ person or ally.

The Pride Scholarship is possible because of contributions from ASA members. To donate, visit *https://secure.qgiv.com/ for/amestaass* and make sure to direct your support to the ASA Pride Scholarship. Those who donate during Pride month (June) will be entered into a prize drawing.

Click the "Join Friends of the Committee" button.

"Join Friends of the Committee" button. Friends of the LGBTQ+ Advocacy Committee are the first to know about events the committee hosts and opportunities for professional service and collaborative projects. They are also invited to join annual open committee meetings. ■

Applications Sought for Science Policy Fellowships

Applications are being accepted for the ASA Science Policy Fellowship, during which the fellow spends 1–2 years in Washington, DC, working to elevate the profile of statistics in policy-making and advocating on behalf of the profession.

The fellow will be based at the ASA headquarters in Alexandria, Virginia, but will spend the bulk of her/his time in Washington, experiencing first-hand how federal science policy is formed.

In partnership with the ASA director of science policy, the science policy fellow is responsible for a wide range of activities, including the following:

- Initiating and leading new advocacy topics for the ASA (preferably in collaboration with other societies)

- Advocating specific statistics-related issues to Congress, the administration, and others

- Exploring and implementing ways in which the ASA's science policy activities could be expanded to better meet its responsibilities (e.g., congressional briefings, webinars)

- Maintaining relationships with key policy officials related to statistics (e.g., US Census Bureau, Hill staff, Office of Management and Budget, Office of Science and Technology Policy, National Academies)

- Generally supporting the dual responsibilities of ASA science policy work: raising the profile of statistics in policymaking and advocating on behalf of the statistics profession

- Attending policy-related events (e.g., congressional hearings, coalition meetings, conferences, briefings, colloquia)

- Coordinating ASA participation in statistics and science and technology coalition events

- Speaking publicly on behalf of the ASA, especially for colloquia

- Communicating with ASA members about science policy (e.g., through *Amstat News*, *Member News*, social media)

Fellowship candidates are expected to possess an excellent record of accomplishment, strong commitment to public service, superb interpersonal and communication skills, and demonstrated ability to multitask and work on a team.

Interested applicants holding a PhD in statistics—or a closely related science—should submit a letter of application, short résumé, writing samples (nontechnical if available), and contact information for three professional references. Outstanding applicants with an MS in statistics or related science will also be considered. The letter of application should not exceed two pages and should explain the applicant's policy interest and reason for applying, scientific and professional background, and attributes and experience that would make him/her effective in this position.

These materials should be submitted via email to ASA Director of Science Policy Steve Pierson at *pierson@amstat.org*. Questions about this opportunity and application requirements also may be directed to Pierson.

The ASA Science Policy Fellowship is a full-year, postdoctoral-level position, renewable for a second year. The compensation is competitive with that of congressional science fellows and other DC-area policy fellows. ∎

JASA
Partners with Authors to Enhance Reproducibility

Article Authors

Julia Wrobel, Emory University

Emily C. Hector, North Carolina State University

Lorin Crawford, Microsoft Research and Brown University

Lucy D'Agostino McGowan, Wake Forest University

Natalia da Silva, Universidad de la Republica

Jeff Goldsmith, Columbia University

Stephanie Hicks, Johns Hopkins University

Michael Kane, Yale University

Youjin Lee, Brown University

Vinicius Mayrink, Federal University of Minas Gerais

Christopher J. Paciorek, University of California at Berkeley

Therri Usher, US Food and Drug Administration

Julian Wolfson, University of Minnesota

In 2016, the *Journal of the American Statistical Association* Applications and Case Studies introduced a reproducibility initiative to address the lack of standardized practices for reproducibility in scientific research. This initiative established minimum criteria for the inclusion of code, data, and workflow for *JASA* Applications and Case Studies papers and piloted a new editorial role—associate editor of reproducibility—to implement these standards. This initiative has since expanded to all original research manuscripts published in *JASA*.

The team of associate editors of reproducibility has also grown, the process has become more standardized, and the role of an associate editor of reproducibility in guiding reproducibility in statistical research has been refined. Additionally, *JASA* implemented a reproducibility award to recognize papers published in the journal with outstanding reproducibility materials.

The *JASA* initiative is part of a larger movement in the scientific community to address what is widely considered a 'replicability crisis' in science. However, there are many terms used interchangeably—

including reproducibility, replicability, reliability, robustness, and generalizability—with potentially different definitions, depending on the field of research. Broadly, these terms address either *replicability/generalizability,* which relates to whether research findings can "be shown in other data sets or populations," or *reproducibility,* which is defined as the ability of a researcher to "duplicate the results of a prior study using the same materials as were used by the original investigator."

These definitions derive from the experimental sciences. How they apply to the statistical methods–focused research featured in *JASA* is less clear but no less critical. Reproducibility in statistics supports the validity of the research presented in a paper and underpins the role of scientific research in advancing knowledge. This role is shared by three key stakeholders—journal, academic reader, and lay public—that benefit from and contribute to high reproducibility standards.

First, the reproducibility process fosters a culture of transparency and accountability that is critical to nurturing the trust placed by the lay public in scientific research. It is an essential safeguard against the dissemination of misinformation that can damage this trust and jeopardize the positive societal impact of future scientific discoveries.

In addition, the academic reader will frequently use published research to motivate future research. Reproducible research creates trustworthy cumulative knowledge on which researchers can confidently build to further advance their field.

Finally, the journal gains credibility from a system that holds researchers accountable for their work—credibility that can be leveraged to publish controversial and impactful research with the potential to change a field of study. As the mediator between researchers and lay public, the journal has substantial control of, and therefore substantial responsibility for, the dissemination of research results. *JASA* takes an active mediating role in upholding high reproducibility standards.

Here, reproducibility in statistical research is referred to as "methods reproducibility," which involves providing materials with the following two key elements:

1. They allow for the numbers in the paper to be directly reproduced

2. They are user-friendly enough for future readers to easily use and build on the proposed technique

This second element is the most critical of methods reproducibility and what associate editors of reproducibility at *JASA* strive to uphold as they leave the assessment of paper quality and mathematical accuracy to the traditional review process.

Reproducibility Effort

To clarify their mission, associate editors of reproducibility introduce the framework of "partners in" versus "officers of" reproducibility. Reproducibility officers would police the materials published by the author to ensure all manuscript numbers and tables are precisely recapitulated. *JASA* associate editors of reproducibility view their role as reproducibility partners, who work with and guide authors in providing resources that enable readers to test and build on their work.

To this end, and to ensure a baseline minimum set of standardized materials are provided by the authors of each paper, the associate editors of reproducibility developed a reproducibility assessment process centered on the author contributions checklist form, which is the central document of the reproducibility effort. The purpose of the form is to "document the artifacts associated with a manuscript (i.e., the data and code supporting the computational findings) and describe how to reproduce the findings." Its target audience is future readers of an accepted manuscript, and it is intended to be phrased in the present tense. The workflow of a 'prototypical' reproducibility reviewer is the following:

1. Read through the authors' completed author contributions checklist form for completeness and clarity.

2. Open any README documents provided and evaluate if they clarify the workflow of the attached code. READMEs are highly encouraged.

3. Look through the directory structure of the provided materials to understand their organization.

4. Open and read through attached code files to ensure the code is sufficiently documented for the use and understanding of future readers.

5. Depending on judgment of the associate editor of reproducibility, run examples or a few key pieces of code.

6. Write a review that comments on any issues found in the steps above with the intention of helping authors strengthen materials for future readers.

Author Dos and Don'ts

The reproducibility review strives to be specific and provide concrete action items to strengthen the reproducibility of the research. By following these guidelines, authors can avoid the step of having to undergo multiple rounds of reproducibility revisions.

Table 1. Dos and Don'ts Based on Common Reasons Reproducibility Revisions Are Requested

	Do	Don't
ACC Form	✓ Fully complete the author contributions checklist form	✗ Leave template comments in the form
	✓ Thoroughly explain materials included	✗ Provide one-word answers in sections
Code	✓ Add comments explaining the purpose of functions	✗ Provide uncommented code blocks
	✓ Explain inputs and outputs user-facing functions	
Data	✓ Include detailed instructions about data access	✗ Forget to include a data dictionary
	✓ When feasible, provide the data in the reproducibility materials or a public data repository	✗ Forget code for processing data
	✓ When data is not shared, submit a realistic facsimile	
Workflow	✓ Choose an easy-to-navigate directory structure	✗ Forget a comprehensive README file
	✓ Provide a README file	✗ Use uninformative file names
	✓ Include the file sequence to follow in the README file	

Frequently Asked Questions

Q. *Do the associate editors of reproducibility run all the code and check that all figures and tables are numerically reproducible? If you don't run all the code, how are you even assessing reproducibility?*

A. As a rule, associate editors of reproducibility do not run all code provided by the authors. Their role is to ensure the author contributions checklist form and code are sufficiently documented such that future readers of the paper can implement and build on results. Further, they frequently review papers with large data sets and computationally demanding methods that cannot be easily reproduced numerically but still have a rightful place in *JASA;* they do not want to discourage authors from submitting such manuscripts or hold more computationally tractable methods to a higher bar. As partners in reproducibility with the authors, the associate editors of reproducibility trust the authors have put in a good faith effort to ensure the materials they provide will produce results consistent with their published work.

Q. *Making my code reproducible will take substantial effort and time. Do I really have to provide reproducibility materials?*

A. For reasons previously described, reproducibility plays a crucial role in published research. While time-consuming, preparing reproducibility materials can be greatly facilitated by doing so throughout the research process rather than at the time of submission. Authors should be encouraged to carefully prepare reproducibility materials by the prospect of publication in top-tier journals such as *JASA* and increased likelihood others will cite their work if the method is easy to reproduce.

Q. *Will my paper be rejected based on the reproducibility review?*

A. No, papers at *JASA* will not be rejected based on the reproducibility materials. However, if the paper has been accepted but the authors do not comply with the associate editors of reproducibility requests, the paper will be returned to the authors for additional revisions.

Reproducibility

Q. *Do I fill out the author contributions checklist form when I initially submit my paper to* JASA?

A. No, the reproducibility review process starts after a paper has already gone through one round of review and been flagged for major or minor revisions. When revising and resubmitting their paper in this second round, authors must provide reproducibility materials.

Q. *If I have a GitHub repository with my code/software package, do I still need to fill out the author contributions checklist form?*

A. Yes. GitHub repositories and software packages are useful supplements to a paper. However, it is still important to fill out the author contributions checklist form to ensure standard materials are included and to explain the workflow. In addition, while software packages can help others use the method, associate editors of reproducibility ask that authors also contribute code from their manuscript typically not included in a software package. This includes data processing code, code for running simulations, and code for reproducing tables and figures.

Q. *If I provide a link to a GitHub repository, doesn't that unblind the process?*

A. Blinding is intended to keep bias out of the acceptance and rejection decision. As associate editors of reproducibility do not make decisions to reject or accept papers, an unblinded GitHub repository does not affect the reproducibility review process. Authors may alternatively choose to anonymize their GitHub repository (*https:// anonymous.4open.science*).

Q. *How can I strive to make my materials good enough to be considered for the reproducibility award?*

A. Authors should place themselves in a reader's shoes by imagining what they would need when attempting to reproduce a paper's results or use the paper's methods.

Q. *I have a complicated, large data set with code that is too computationally intensive to run on a standard laptop. Can I still be considered for the reproducibility award?*

A. Yes. All papers accepted by *JASA* Theory and Methods or Applications and Case Studies in the previous calendar year will be considered for the reproducibility award.

Final Words

While the authors put forth *JASA's* current best efforts at guiding this type of reproducibility, they acknowledge methods reproducibility is not a static target and this process will likely change and improve over time.

They believe a wider discussion about methods reproducibility—including adoption of common standards—by editors across statistical journals would benefit the field and hope describing their process can facilitate discussion. In addition, they welcome the adoption of the author contributions checklist form as a template or starting point for other statistics and data science journals initiating their own reproducibility review process. ∎

MORE ONLINE
A complete version of this article and author contact information can be found at *www. tandfonline.com/ doi/full/10.1080/ 01621459.2024.2 340557.*

STATS4GOOD

Global Data for Good Data Sources: A World of Opportunity

David Corliss is the principal data scientist at Grafham Analytics. He serves on the steering committee for the Conference on Statistical Practice and is the founder of Peace-Work.

The Data for Good community is seeing unprecedented growth in the amount, quality, detail, and availability of data of all kinds. Whether your work is in academia, government, or industry—whether you are an experienced researcher, early-career professional, or student—these new data sources put the power of data around the world in your hands for the benefit of all.

Leveraging global data to the greatest advantage requires understanding the sources, their strengths and weaknesses, and the context in which they were developed. In January, I highlighted the globalization of data as one of the most important challenges—and opportunities—being addressed by D4G scientist-advocates today.

Many data sources for international D4G projects can be found right here in the United States. Federal agencies collect data on population and demographics, economics and international trade, and a host of other subjects. This makes Data.gov one of the best places to look for data supporting international projects. Learning to navigate the site, finding the best data sources for your work, and reading peer-reviewed articles using them will be helpful.

Of course, much of the best international data will be found internationally. Outstanding government statistical agencies with a wide variety of reliable data include Statistics Canada, the UK Data Service, and the European Union's Eurostat.

National statistical agencies will often provide the most in-depth information about the country's demographics, economics, industry, education, and public health. At the same time, not all government statistical agencies equal the standard of excellence of those mentioned here. The reliability, and especially the *availability,* of data from governmental agencies can vary widely—especially for countries with the most severe problems and acute needs.

When looking at a new international data source, the standard best practices for using any data can be applied. Search to see if the resource is cited by peer-reviewed papers. Read leading papers for the sources they use and look for review articles on data sources in respected academic journals. Test the data found in the source for internal consistency and check against numbers reported independently by corroborating sources whose only vested interest is the accuracy of the numbers. Pay close attention to the accuracy of numbers for marginalized groups or regions, which can provide a window into the reliability of the source in general. Applying the same rules for citing any paper will go a long way toward finding the best data for your project.

Remember Data.gov is a data search service, not the source itself. Be sure to bookmark websites with the data you need most and check in regularly for updates.

For United Nations agencies and their data resources, the iLibrary is the place to go. This search engine is the key to data sets from across the UN ecosystem, including data used by UN agencies, support for sustainable development goals, and a wealth of research and analysis using the data.

Another tremendously valuable international data resource is the Open Data project from the Work Bank. It includes country-level information on economics, demographics, and much more.

As an example of how these different data sources can be used together, let me tell you about a current project at Peace-Work, a Data for Good nongovernmental organization. The research is on how human trafficking and forced labor are being used as a weapon of war in Ukraine. A search on the UN-iLibrary for migration data found the World

Getting Involved

In opportunities this month, the US Census Bureau has released the Justice Outcomes Explorer, which provides statistics and interactive data visualizations on how people have been in contact with the criminal justice system.

Also this month, early-career scientists and their mentors will be interested in the Science & SciLifeLab Prize for Young Scientists, awarded annually to four young scientists for outstanding life science research based on a doctoral degree earned in the previous two years. All the details are available at *https://scienceprize.scilifelab.se.*

Migration Report and a paper offering guidance for using their longitudinal data. A search at Data.gov identified a country-level list of goods where forced labor and child labor are present. This was matched to country-level data from the World Bank to identify risk factors, which can then be applied to countries involved in the conflict in Ukraine. This example demonstrates how synthesizing data from multiple peer-reviewed data sources can provide critical insight into international D4G projects.

With so many new data sources becoming available every day, there are more opportunities for going global with your next D4G project than ever before. The data heroes of this work are the people working at national and international statistical agencies to compile, test, and make available the data needed for projects that make a real difference. With the ever-growing list of reliable international data sources, there is truly a world of opportunity with Data for Good! ∎
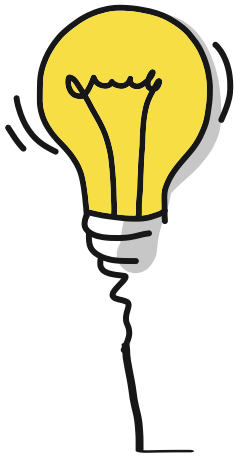
STAT*tr@k*

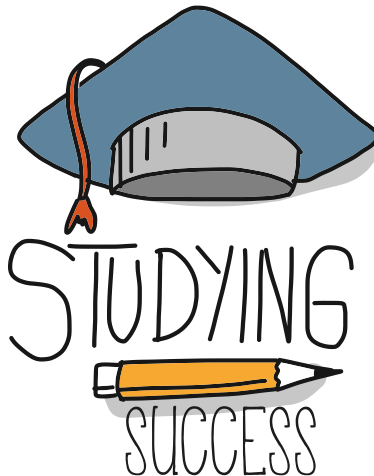# Recommendations for Undergraduate Students Interested in Statistics and Data Science

Mine Dogucu is a tenure-track assistant professor of teaching and vice chair for undergraduate studies in the department of statistics at the University of California at Irvine. She is an educator with an interest in statistics and data science education and an applied statistician with experience in educational research. She is also the co-author of the upcoming book *Bayes Rules! An Introduction to Bayesian Modeling with R.*

College is a busy time. It is busy socially, as it is a great time to build life-long friendships. For some, it is a busy time to make money to live. It is also a busy time for those juggling multiple courses and activities. The most important skill one can build early on is time management. Without intentional time-management methods, demands will be overwhelming. The following three types of apps may be especially helpful for time management:

- Digital calendar

- Task manager

- Social media blocker

For a calendar, it might be practical to use whatever your university is using (e.g., Google, Outlook, etc.) Be mindful of how much time is allocated to classes, work, studying, and social life. You can do this by using different colors for different types of events. Task managers will help you remember assignments and manage repeating tasks (e.g., do laundry every Saturday). Last, many people have a social media addiction. If you have a hard time controlling the time you spend on social media, try a blocker app that can keep you from logging in during certain hours of the day.

There are many opportunities on campus that might also be helpful as you build your career. These can include guest speakers. These speakers are often interested in student engagement. If you can overcome shyness, you can talk to them, as well.

Use the career office and writing offices as much as possible. These offices provide great services and training programs that will help with career readiness. Writing is one of the most important skills you can develop during your studies.

Perhaps the most important service on most campuses is counseling! There is nothing more important than one's health. Mental health is as important as physical health. Seek all the help you need without any hesitation.

Graduate School and the Job Market

It is never too early to start thinking about what comes after college. One of the decisions you have to make is what comes immediately after—graduate school or the job market? Try to gather as much information as possible from everyone about the different paths. Different people with a variety of career paths can help you in the decision-making process. Be mindful of who you

are talking to and their potential biases. For instance, if you get a professor's opinion, they might be more inclined to support the graduate school path, as they are in academia themselves. After all, the most important thing is knowing your own self—your dreams, desires, and life goals.

The Electronic Undergraduate Statistics Research Conference takes place annually online and for free. This conference has two panels—one on graduate school and one on statistics careers. The second panel often has speakers who work in industry and government. These panels and the recordings from previous years may be helpful in demystifying graduate school and the job market. If you attend the live event, you can ask questions.

For those curious about graduate school, Ray Bai's blog post "Demystifying Graduate Admissions for Statistics PhD Programs" might be helpful. In addition, the website A First-Gen's Guide to Grad School: How to Get in, Survive, and Thrive is a great resource.

When it is time to apply for graduate school, you'll need to ask for recommendation letters. Make sure to read "Tips on Recommendation Letters for Students and Instructors" on the Data Pedagogy blog. If you are living in the United States, you might consider reading Simon P. Couch's "Applying to the NSF Graduate Research Fellowship (GRFP)."

Summer programs or internships can also help you decide the path you'd like to take. You can consider one of the Research Experiences for Undergraduates or Summer Institutes in Biostatistics and Data Science programs.

For jobs, if you consider government jobs, make sure to take a look at "Find Your Next Job in Data Science in the Federal Government" and the American Statistical Association's Internships and Fellowships website. You might also like the *Build a Career in Data Science* podcast.
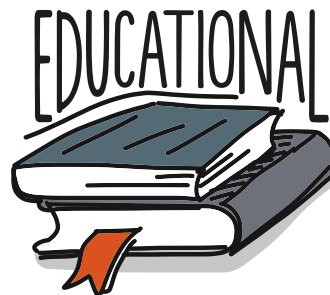
While searching for jobs, make sure to think about your application area of interest. You can widen your search with keywords such as "statistics," "data science," "biostatistics," "bioinformatics," "psychometrics," and "econometrics."

Connecting with People

Human connections are important. That is not because someone will help you find a job. As humans, we learn from each other and how to overcome struggles. When you make connections, you learn some and teach some.

Campus student groups are important for making connections. For instance, on my campus, Data @ UCI is the main group for students interested in data science. This group did not exist three years ago. This is all to say, even if you don't have a group yet, you might consider taking the lead and starting one to build your own data community. You might even consider building a student chapter of the American Statistical Association.

Meetup.com is an online website that has groups of people with similar interests. You can consider joining a local Meetup group and attending their events. R Ladies, PyLadies, and R User Groups are just some of the Meetup groups.

You might also consider becoming a member of a professional organization to connect with other members. Organizations with data science communities include the American Statistical Association, Association for Computing Machinery, and Institute of Electrical and Electronics Engineers.

Conferences and seminars are also great places to connect with others. Keep an eye out for ones in your area.

Portfolio

Regardless of the path you are taking, building a data science portfolio is an important part of learning. Your portfolio can include class projects and projects you do outside of class for fun. These projects do not necessarily need to have groundbreaking scientific knowledge. They can be learning opportunities for you while also showcasing your knowledge to others.

One way of making projects public is using GitHub and/or hosting/linking to them on your website. I cannot emphasize the importance of GitHub (and the like) for showcasing your work and data skills. Having a personal website can really show who you are, your interests, and your projects.

Resources Mentioned in This Article

Electronic Undergraduate Statistics Research Conference
www.causeweb.org/usproc/eusr/2023

"Demystifying Graduate Admissions for Statistics PhD Programs"
https://bit.ly/4bwbpjh

"A First-Gen's Guide to Grad School: How to Get in, Survive, and Thrive"
https://first-gen-guide.com/resources

"Tips on Recommendation Letters for Students and Instructors"
https://bit.ly/3yh4YCi

"Advice on Applying to the NSF Graduate Research Fellowship (GRFP)"
https://bit.ly/44GzpxN

Research Experiences for Undergraduates
www.nsf.gov/crssprgm/reu/reu_search.jsp

Summer Institutes in Biostatistics and Data Science
https://bit.ly/4begQDM

Internships and Fellowships
https://bit.ly/3sBw6JK

Build a Career in Data Science
https://bit.ly/3UVHg7p

ASA Student Chapters
https://bit.ly/4b56Oop

Undergraduate Class Project Competition
www.causeweb.org/usproc/usclap

"Find Your Next Job in Data Science in the Federal Government"
https://usajobs.github.io/microsite-data-science

Data @ UCI
www.dataatuci.com

Association for Computing Machinery
www.acm.org

Institute of Electrical and Electronics Engineers
www.ieee.org

tidytuesday
https://github.com/rfordatascience/tidytuesday

If you are looking for projects outside of class, tidytuesday provides a data set every Tuesday. On social media, especially X, you might find others sharing their work using #TidyTuesday. Community members publicly sharing their work also serve as inspiration.

If working on a class project, you might consider submitting your project to the Undergraduate Class Project Competition. Submission takes place semiannually in December and June. The Electronic Undergraduate Statistics Research Conference is also a great place to submit your projects. If accepted, you can present your work through a video presentation.

Closing Thoughts

You might hear the following:

- You are not a data scientist because you are not strong in math/stats/computer science.

- What you do is not data science because it does not involve advanced computing.

- What you do is not data science because it does not involve advanced probability.

- You cannot be a data scientist because you took courses on X and Y but not Z.

You might hear others say these or you might be saying them to yourself. Even though it is good to know math, stats, and computing, there is no data scientist who knows it all. Focus on your strengths and work to improve your weaknesses.

Last but not least, as you navigate to make a permanent place in the data science world, always remember to consider the impact of your work on humans, other living beings, and the planet. Ethics is perhaps one of the most important parts of data science you'll need to learn. ■

AAAS Welcomes Statistics Section Fellows

The Section on Statistics (Section U) congratulates the following members on their recognition as 2023 fellows of the American Association for the Advancement of Science.

This acknowledges their contributions to advancing science and its applications in service to society. Since 1874, AAAS has annually recognized scientists and others across various disciplines for their accomplishments, including pioneering research, leadership within their field, teaching and mentoring, fostering collaborations, and advancing public understanding of science.

AAAS is the world's largest general scientific society and publisher of the *Science* family of journals. Newly elected fellows are chosen from 24 disciplines aligned with the association's 24 scientific and engineering sections and celebrated for their scientific and socially significant achievements throughout their careers.

Section on Statistics (Section U) members named as 2023 AAAS fellows are as follows:

- **Rebecca A. Betensky,** New York University School of Global Public Health
- **T. Tony Cai,** University of Pennsylvania
- **Ming-Hui Chen,** University of Connecticut
- **Daniel Gillen,** University of California, Irvine
- **Amita Manatunga,** Emory University
- **Len Stefanski**, North Carolina State University
- **Christopher Wikle,** University of Missouri-Columbia
- **Tian Zheng,** Columbia University in the City of New York

To view all the 2023 AAAS fellows, visit the website at *https://bit.ly/3QgFhYU*. ∎

2021 COPSS Presidents' Award Winner Presents 2024 Donna J. Brogan Lecture

Jeff Leek, vice president and chief data officer of the Fred Hutchinson Cancer Center, presented the 2024 Donna J. Brogan Lecture in Biostatistics on April 15.

The winner of the COPSS Presidents' Award in 2021, Leek develops statistical methods, software, and data resources and analyses that help researchers make sense of massive-scale genomic and biomedical data. As an educational leader, he has helped craft online open courses in data science that have enrolled millions of people.

Leek's lecture was titled "Building the Fred Hutch Data Science Lab in the Era of AI."

In his presentation, he discussed the efforts to build the Fred Hutch Data Science Lab to support a healthy data ecosystem for the whole institution. He talked about how data resources, training, statistical software, infrastructure, support, and community are being built to support data-intensive science.

This lecture honors Donna J. Brogan, a former faculty member of Emory University and chair of the Rollins School of Public Health Department of Biostatistics. For details about Brogan and the lecture series, visit *https://bit.ly/4agX3SMl*. ∎

Predict 5 Contest Winners Looked to the Past to Predict the Future

In celebration of Mathematics and Statistics Awareness Month, ASA members Sarah Pagni, Shruti Jain, and Matthew Finkelman put together a contest called Predict 5, in which contestants predicted answers to five questions about different topics and events taking place in April. And the winners are …

High School Category
Prattay Bhattacharya
Anders Peng (honorable mention)

Undergraduate Category
Joshua Castro
Gabriel Forthofer (honorable mention)

Graduate Category
Vincent Onyame
Jonathan Im (honorable mention)

Prediction 1: *In Major League Baseball, how many games will the New York Mets win during the month of April 2024?*

Answer: 15
- Prattay Bhattacharya's Prediction: 13
- Joshua Castro's Prediction: 12
- Vincent Onyame's Prediction: 15

Prediction 2: *How many songs by Taylor Swift will appear on the last Top 40 list (chronologically) to be published on American Top 40 with Ryan Seacrest?*

Answer: 2
- Prattay Bhattacharya's Prediction: 4
- Joshua Castro's Prediction: 2
- Vincent Onyame's Prediction: 1

Prediction 3: *What will the "current national average gas price" (in US dollars and cents) be for regular gas in the United States on April 30, 2024, at 5:00 p.m. ET, according to AAA Gas Prices?*

Answer: $3.657
- Prattay Bhattacharya's Prediction: $3.798
- Joshua Castro's Prediction: $3.60
- Vincent Onyame's Prediction: $3.412

Prediction 4: *What will the high temperature be (in degrees Fahrenheit) in Boston, Massachusetts, on April 30, 2024, according to AccuWeather?*

Answer: 54
- Prattay Bhattacharya's Prediction: 62
- Joshua Castro's Prediction: 54
- Vincent Onyame's Prediction: 62

Prediction 5: *What will the closing stock price (in US dollars and cents) of Intel Corporation be at the end of the day on April 30, 2024, according to Yahoo Finance?*

Answer: $30.47
- Prattay Bhattacharya's Prediction: $37.87
- Joshua Castro's Prediction: $43.43
- Vincent Onyame's Prediction: $34.524 ■

Lumivero Offers $20,000 Research Grant for Early Career Researchers

Lumivero is offering a $20,000 grant to a researcher who shares Lumivero's commitment to improving the world through research excellence. To be eligible, applicants must meet the following qualifications:

- A doctorate earned from 2016–2023

- An academic appointment at an institution of higher education or a contract covering the duration of the project (November 1–October 31, 2026)

- A proposal for a qualitative, statistical, or mixed method research project

- The support of your university

- Involvement of Citavi, NVivo, and NVivo Transcription, NVivo Collaboration, or XLSTAT for at least part of the data analysis in your research

Applications will be accepted until 11:59 p.m. EDT on July 31.

Visit the Lumivero website at *https://bit.ly/4bEiwX3* for details and to apply. ∎

Visit the ASA website to view a comprehensive list of awards and scholarships.

ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop Provides Student Travel Grants

The ASA Biopharmaceutical Section Regulatory-Industry Statistics Workshop has funding for 10 student travel grants. Up to $500 per student, by reimbursement, is available to support students' participation in the workshop. During the abstract submission process, designate that you would like to be considered for the student travel grant.

- To qualify, a student must meet the following requirements:

- Be a full-time student working toward a master's or doctorate degree in statistics, biostatistics, or a related field at a college or university when submitting the application

- Be an author who will present an accepted poster at the 2024 workshop

All recipients of the student travel grants are required to register for the workshop at *https://bit.ly/4bcjpGz* by the advance registration deadline of August 14.

The student travel grant committee is composed of members from industry, the FDA, and academia. Committee members will review all posters in a blinded fashion. ∎

Deadlines and Contact Information for Select ASA National Awards, Special Lectureships, and COPSS Awards

The ASA's extensive awards program (*amstat.org/ASA/Your-Career/Awards-and-Scholarships.aspx*) recognizes statisticians who have made outstanding contributions to the association and statistical profession through research, teaching, consulting, and service.

AWARD	DEADLINE	QUESTIONS & NOMINATIONS
Links Lecture Award	July 1, 2024	*awards@amstat.org*
Dorothy Marie Lamb and Annette Lila Ryne Memorial Scholarship	July 15, 2024	*awards@amstat.org*
Health Policy Statistics Section Achievement Awards	September 15, 2024	Health Policy Statistics Section website (*asahealthpolicy.org/for-students*)
Lester R. Curtin Award	October 15, 2024	*awards@amstat.org*
Deming Lecturer Award	October 15, 2024	*awards@amstat.org*
Lingzi Lu Memorial Award	October 15, 2024	*awards@amstat.org*

JOIN
A SECTION OR CHAPTER

EXPLORE chapter and section leadership opportunities

STAY CURRENT with the different methodologies and applications in your area of expertise

EXPAND your professional network and strengthen your relationships in the community

If you've been thinking about joining an ASA section or regional chapter, we have made it easier than ever. With a few clicks, you can add section and chapter membership and pay online.

Chapter and section membership can greatly enhance the value of your membership.

Add section and chapter membership at
ww2.amstat.org/membersonly/additems.

ASA

Mathematical Statistician

FDA U.S. FOOD & DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH
OFFICE OF TRANSLATIONAL SCIENCES

The Office of Biostatistics is seeking individuals with strong statistical methodology skills and an interest in biomedical applications to serve as mathematical statisticians. Incumbents work with multidisciplinary teams of review scientists in a dynamic, highly challenging, and innovative atmosphere of development, evaluation, and research of drug and therapeutic biologics. The Office of Biostatistics is responsible for reviews in all therapeutic areas CDER supports and can be as diverse as cardio-renal, oncology, rare disease, and antimicrobial products. Incumbents have an opportunity to employ a broad variety of statistical procedures relevant to pre-clinical and clinical evaluation decisions for new and generic drugs as well as new and biosimilar biologics and the emerging field of quantitative risk assessment.

DUTIES AND RESPONSIBILITIES

- Evaluate and advise on protocols for clinical studies and assess the evidence for safety and efficacy from clinical studies submitted in drug and biologics applications.
- Employ a broad variety of statistical procedures relevant to pre-clinical and clinical evaluation decisions for new and generic drugs as well as new and biosimilar biologics and the emerging field of quantitative risk assessment.
- Work with multidisciplinary teams of review scientists in a dynamic, highly challenging, and innovative atmosphere of development, evaluation, and research of drug and therapeutic biologics.
- Refine your consulting, communication, and presentation skills and present at domestic and international professional meetings.
- Engage in an active collaborative regulatory research program which will allow you to advance your skills and professional development.
- Interact with national, international, public, and private organizations on statistical issues, and help develop guidance for the pharmaceutical industry.

QUALIFICATIONS

Applicants should possess an advanced degree with specific coursework in Statistics, Biostatistics or Mathematical Statistics. Applicants with a doctoral degree and associated experience are highly desirable. In addition to a background in statistics, applicants should have an interest in biostatistics, clinical trials, epidemiology, genomics, or risk assessment.

The ability to communicate statistical issues to non-statisticians is vital.

Non-US citizens may apply for term appointments.

BENEFITS

Health and Life Insurance	Paid Holidays	
Long-term Care Insurance	Flexible Spending Accounts (FSA)	
Dental and Vision Insurance	Federal Retirement Plan	
Annual and Sick Leave	Thrift Savings Plan (401k)	

WORK/LIFE BALANCE

Telework & Alternative Work Schedules
Child Care Center | Fitness Center
Employee Assistance Program/Resource Groups
Commuting and Transportation Programs

ARE YOU INTERESTED IN WORKING AT FDA?

SEND YOUR RESUME OR CURRICULUM VITAE TO:
CDEROTSHires@fda.hhs.gov

LOCATIONS

Mathematical Statisticians are located in the Washington, D.C. area.
Remote employment may be available.

FOOD AND DRUG ADMINISTRATION ● CENTER FOR DRUG EVALUATION AND RESEARCH ● OFFICE OF TRANSLATIONAL SCIENCES

Statistical Analyst

FDA | **U.S. FOOD & DRUG ADMINISTRATION**
CENTER FOR DRUG EVALUATION AND RESEARCH
OFFICE OF TRANSLATIONAL SCIENCES

The Office of Biostatistics is recognized for excellence in the application and communication of statistical science in drug regulation and development. We play a central role in promoting innovative, science-based, quantitative decision-making throughout the drug development life-cycle. To support our Center's mission, we provide statistical leadership, expertise, and advice to ensure that safe and effective drugs are available to the American people.

DUTIES AND RESPONSIBILITIES

- Work with a multidisciplinary review team to provide statistical programming and data management support, assess the quality and completeness of submissions, prepare clinical trial analysis datasets, validate sponsor results, assist in modeling and simulation, and suggest possible additional statistical analyses required to fully evaluate the evidence in the submission.
- Collaborate with scientists from the Office of Pharmaceutical Quality, statistical reviewers in OB, and management on a variety of computationally intensive projects to support and improve the efficiency of regulatory product review, evaluation of pharmaceutical quality and applied regulatory research.
- Use machine learning and natural language processing to assess internal and external data sources to support assessment of quality intelligence throughout the product life cycle.
- Develop, validate, implement, document, maintain and support programming tools and software according to standards and accepted validation procedures; Support efforts to develop, document and apply reusable code and/or tools.
- Develop software using the appropriate statistical programming packages for statistical reviewers to support programming-intensive review-related activities such as sensitivity analysis, Bayesian approaches, clinical trials modeling, genomic studies, psychometric Clinical Outcome Assessment (COA) validation, and simulation.
- Promote and improve the Center data standards initiatives mandated by the Prescription Drug User Fee Act; Monitor the quality of the implementation of data standards used in New Drug Application submissions.
- Apply your skills to address unique and precedent-setting problems, while refining your consulting, communication, and presentation skills.

REQUIRED QUALIFICATIONS

Master's degree in statistics or biostatistics.
Familiarity with R, SAS, data science tools, machine learning predictive techniques and natural language processing.

PREFERRED QUALIFICATIONS

Experience in clinical trials, epidemiology, genomics, or risk assessment. Strong skills in multiple programming environments.

Candidates should also have excellent oral and written communication skills.

The ability to communicate statistical issues to non-statisticians is vital.

BENEFITS

Health and Life Insurance	Paid Holidays	
Long-term Care Insurance	Flexible Spending Accounts (FSA)	
Dental and Vision Insurance	Federal Retirement Plan	
Annual and Sick Leave	Thrift Savings Plan (401k)	

WORK/LIFE BALANCE

Telework & Alternative Work Schedules
Child Care Center | Fitness Center
Employee Assistance Program/Resource Groups
Commuting and Transportation Programs

FDA wants to hire you.

ARE YOU INTERESTED IN WORKING AT FDA?

SEND YOUR RESUME OR CURRICULUM VITAE TO:
CDEROTSHires@fda.hhs.gov

LOCATIONS
Statisticians are located in the Washington, D.C. area.
Remote employment may be available.

FOOD AND DRUG ADMINISTRATION ● CENTER FOR DRUG EVALUATION AND RESEARCH ● OFFICE OF TRANSLATIONAL SCIENCES

Come to your Census

Join the U.S. Census Bureau to help produce quality data that enable Americans to better understand our country, its population, resources, economy, and society.

Why the U.S. Census Bureau?

- **Play a pivotal role as an innovator:** We breathe creativity, innovation, agility, and flexible solutions. We value diversity in people and opinions.

- **Grow your career:** We don't just talk about career development—we provide training, mentoring, coaching, and growth opportunities.

- **Enjoy unparalleled benefits and competitive salaries:** Our salaries and incentives are competitive with private industry—and they don't include 60-hour work weeks.

- **Have a real work-life balance:** We offer flexible work schedules, a telework program for eligible positions, and reasonable work hours that show how much we care about you and your family.

Join the U.S. Census Bureau!

Apply for available jobs at census.gov/jobs today.

The U.S. Census Bureau is an equal opportunity employer.

United States® Census Bureau

This month's Top 10 is the **'Top Ten Statistics and Data Science Terms That Sound Like They Could Be in the Olympics'**

Wasserstein

Amstat News continues its entertaining offering by ASA Executive Director Ron Wasserstein, who delivers a special top 10—one that aired during a recent edition of *Practical Significance*. He says, "The summer Olympics are just around the corner. Athletes will compete in more than 30 sports in Paris during July and August. With the Joint Statistical Meetings taking place during the Olympics, it occurred to me many statistical terms sound like they could be events in Paris. So, here are the top 10 statistics and data science terms that sound like they could be in the Olympics."

10
Data wrangling (*we'd pay to watch that*)

09
Lasso (*a good skill for wranglers*)

08
Bootstrapping (*goes with lasso skills, I would think*)

07
Gradient boosting

06
Cross-validation

05
Data mining

04
Variable transformation (*judged on speed and approximate normality*)

03
Extrapolation

02
Adaptive design (*judged on speed and optimality*)

#01

Futility analysis

To listen to the *Practical Significance* podcast, visit *https://magazine.amstat.org/podcast-2*.

CALL FOR PAPERS

JASA Special Issue on Statistical Science in Artificial Intelligence

"Bridging Statistics and AI:
Innovative Approaches to Core Challenges"

Seeking original research papers that address the following key aspects:

Identification of Core AI Problems

Statistical Contributions to AI

Innovative Statistical Theory, Method, and Applications

Submit paper by December 31, 2024
https://bit.ly/3TAlnIS

SDSS
SYMPOSIUM ON
DATA SCIENCE & STATISTICS

RICHMOND, VA • JUNE 4–7, 2024

THANK
you

to our supporting sponsors:

Bristol Myers Squibb™

STATA®

Institute for Data Science & Informatics
University of Missouri

Westat®

Learn more at **ww2.amstat.org/meetings/sdss/2024/sponsors.**