Research Highlights from JSM 2020

Presentations feature research with applications to COVID-19, sports, forensic science, the opioid crisis, climate change, precision medicine, stock market returns, algorithmic fairness, social media and more.

ALEXANDRIA, Va. (July 30, 2020) — The 2020 Joint Statistical Meetings will bring together statisticians and data scientists from around the world from Sunday, August 2, through Thursday, August 6. This year, for the first time, JSM will be held in a virtual format. This tip sheet highlights interesting presentations from the conference. Complementary press registration is open, courtesy of the ASA. Information about how to receive press credentials is at the bottom of this release.

FEATURED RESEARCH (SYNOPSES BELOW)
**ASTERISKS IDENTIFY CORONAVIRUS/COVID-19-RELEVANT RESEARCH

MONDAY HIGHLIGHTS

1. Improving NCAA football rankings with data science
2. Precision medicine for stem cell transplants
3. Statistical analysis of footprints in forensic science
4. The role of uncertain infection status in controlling epidemics **
5. Deep Learning AI neural networks for climate change

TUESDAY HIGHLIGHTS

6. Which emoticon personality are you? ;-)
7. Undercounting invisible immigrant communities in the census
8. How do algorithmic tools affect fairness and quality of decision-making?
9. Mathematical model for reopening businesses during COVID-19 **

WEDNESDAY HIGHLIGHTS

10. Analyzing the fairness of a pre-trial risk algorithm
11. Data science tools for monitoring patient safety during clinical trials
12. "Nowcasting" and forecasting COVID-19 **
13. Text-mining news articles to predict stock returns

THURSDAY HIGHLIGHTS

14. Group testing for COVID-19: What is it, and how could it be used more effectively? **
15. Panel on statistical significance and *p*-values
16. COVID-19 infectious disease modeling and statistics: Myths, maxims and mobilization **
17. New model for contact tracing and disease spread **
18. Who are the scientific grant gatekeepers?
19. Statistical models for comparing state opioid policies

MONDAY SYNOPSES

1) Improving NCAA football rankings with data science

In US college football, declaring a national champion hasn't been easy. Prior to 2014, the statistical rating method used was plagued with criticisms, and currently a 13-member committee selects and seeds teams for playoffs. But some fans still wonder if there's a better way. In this presentation, Shane Reese of Brigham Young University will present a new statistical rating system, called Ratings Using Score Histories, developed with colleagues to help select playoff teams. Its novel feature: it uses data from a game's score process—that is, the score for each point in time throughout a game—for an entire season. Unlike previous methods, this new system treats teams from weaker and stronger conferences more fairly and also makes use of all available data. The presentation will demonstrate how the rating system can be used, including results from the 2019–2020 season.

The presentation, "RUSH: An Evolutionary Approach to Ranking College Football Teams," will take place Monday, August 3, 2020: 10:00 AM to 11:50 AM EDT.
Abstract: https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstractid=309359

2) Precision medicine for stem cell transplants

Treating blood cancers like leukemia sometimes calls for the aggressive approach of transplanting stem cells from a healthy donor into a patient's bone marrow. To help prepare the patient's body for the transplant, a drug called busulfan is injected directly into the veins. It's tricky to get its exact dosage correct, however—too much can lead to toxicity or even death, while too little can make it easier for the cancer to return. In this presentation, Peter Thall from the MD Anderson Cancer Center at the University of Texas will describe the new "precision medicine" statistical model that he and colleagues created to determine the right dosage, resulting in a method that be easily used by any transplant doctor. By switching from the current

"one-size-fits-all" strategy to the new method, the researchers calculate that doctors can extend many patients' lives dramatically—by an average of 10 to 14 months, for example, for 40- to 60-year-olds in complete remission, which is an improvement of up to 290%.

The presentation, "Bayesian Nonparametric Survival Regression for Optimizing Precision Dosing of Intravenous Busulfan in Allogeneic Stem Cell Transplantation," will take place Monday, August 3, 2020: 10:00 AM to 11:50 AM EDT.
Abstract: https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstractid=308042

3) Statistical analysis of footprints in forensic science

Using footprints in forensic science to link a suspect to a crime scene is statistically complicated. All shoe soles have some kind of random markings—stray holes and scratches acquired during normal use—which investigators can use to compare a suspect's shoe with a crime scene print. But first investigators need to mathematically understand how these random markings accumulate in the first place. In this presentation, Naomi Kaplan-Damary at the University of California, Irvine and the Hebrew University of Jerusalem will present work with colleagues that involved the analysis of nearly 400 shoes. Using the same equations that can describe the distribution of trees in a forest or stars in the Milky Way, the researchers pinpointed the areas on the soles that are more likely to pick up distinguishing blemishes, which will help determine their importance as evidence.

The presentation, "A Step Forward in Estimating the Probability of Accidental Mark Locations on a Shoe Sole," will take place Monday, August 3, 2020: 1:00 PM to 2:50 PM EDT.
Abstract: https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstractid=309225

4) The role of uncertain infection status in controlling epidemics **

During an epidemic, public health officials ideally would know exactly who was infected at any moment so they could quickly intervene to stop the spread (by quarantining the infected individuals, for example). In real life, of course, there's usually a lot of uncertainty around who's infected and who's not. This uncertainty means that some uninfected people will be needlessly quarantined, disrupting families and workplaces, while some infected people will be allowed to mix freely and spread the disease faster. Jessica Hoffman from University of Texas at Austin will present theoretical results with colleagues showing that even a tiny uncertainty has a dramatic impact on the amount of time and resources (such as quarantining) needed to contain the epidemic. This work implies that a community should invest in knowing exactly who is infected—through contact tracing, for example—or else it will pay the price tenfold later.

The presentation, "The Cost of Uncertainty in Curing Epidemics," will take place Monday, August 3, 2020: 1:00 PM to 2:50 PM EDT.
Abstract: https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstractid=312849

5) Deep Learning AI neural networks for climate change

Deep Learning is a powerful type of machine learning artificial intelligence that has been used with extraordinary success in fields such as computer vision, speech recognition, and language translation. But its use in climate change science is new. Prabhat from Lawrence Berkeley National Laboratory and colleagues have worked on using deep learning for the problem of detecting extreme weather events such as hurricanes and severe weather fronts. In this presentation, Prabhat will describe their work training a state-of-the-art deep learning network to find extreme weather patterns in complex "ground-truth" climate data sets. He will also show how they can now apply the trained network to new climate data sets and use this to understand how extreme weather patterns will change in the future.

The presentation, "Deep Learning for Extreme Weather Detection," will take place Monday, August 3, 2020: 1:00 PM to 2:50 PM EDT.
Abstract: https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstractid=309281

TUESDAY SYNOPSES

6) Which emoticon personality are you? ;-)

Before emojis there were emoticons—primitive emoji pictographs that conveyed emotion through plain-text characters. Users could choose between different styles—a smiley of :-) or :D, for example, instead of :). Do people choose different emoticon styles at random, or do patterns exist that reveal something about a person? To investigate, Juha Alho from the University of Helsinki analyzed discussions between 2001 and 2015 on suomi24, a large Reddit-style social-networking site in Finland, for a total of 48 million individual posts. Using a statistical method called correspondence analysis, Alho uncovered four distinct emoticon user "personalities": the Classics, the Noses, the D-Grins, and the Multi-Mouths. In this talk, Alho will describe these personality groups, discuss daily emoticon usage patterns, show trends over time, and reveal which sports forums—from golf to ice swimming to parkour—had the largest relative shares of each of the four emoticon personalities.

The presentation, "What Authors Reveal of Themselves in Internet Discussions?," will take place Tuesday, August 4, 2020: 10:00 AM to 11:50 AM EDT.
Abstract: https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstract id=308134

7) Undercounting invisible immigrant communities in the census

It's been estimated that almost 11 million undocumented immigrants currently live in the US. This number, and their potential exclusion on the census, has become a hotly debated political issue. How do researchers even estimate the number of undocumented immigrants, and what happens if they are not counted? In this presentation, Nadia Flores-Yeffal of Texas Tech University will answer these questions and provide context around the counting of invisible communities of immigrants. She estimates that the undercount in the 2020 Census of both undocumented immigrants and their US-born family members could be up to 8% of the entire population, due in part to the prevalence of "mixed-status families."

The presentation, "How Are Invisible Communities of Immigrants in the United States Counted? What Happens If They're Undercounted?," will take place Tuesday, August 4, 2020: 10:00 AM to 11:50 AM EDT.
Abstract: https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstract id=312361

8) How do algorithmic tools affect fairness and quality of decision making?

In today's data-rich society, many of our decisions are guided partly by machine recommendations, from online shopping to movie recommendations. Judges often use algorithm recommendations as well, for example when weighing the risks of releasing an arrestee on bail before a trial. Much of the debate around these pretrial risk-assessment instruments has focused on accuracy and fairness of the algorithms themselves, however, not on how the algorithms influence and shape their users' behavior. Kosuke Imai from Harvard University and colleagues have developed a statistical framework for experimentally evaluating the impact of machine recommendations on human decisions, including whether or not they improve the fairness of decisions or lead to decisions with better outcomes. This presentation will illustrate the new methods with an example from the criminal justice system, showing how the use of a risk-assessment algorithm influenced judges' decisions and whether it resulted in racial or gender bias in results.

The presentation, "Experimental Evaluation of Computer-Assisted Human Decision-Making: Application to Pretrial Risk Assessment Instrument," will take place Tuesday, August 4, 2020: 1:00 PM to 2:50 PM EDT.
Abstract: https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstractid=309356

9) Mathematical model for reopening businesses during COVID-19 **

How can leaders decide how to reopen the economy safely while the coronavirus is still circulating in the population? In this presentation, Hongyu Miao of the University of Texas Health Science Center at Houston will describe work with colleagues in developing a mathematical model that considers both profit and infection risk from a business entity's perspective. They propose an algebraic equation that describes the net profit a business can generate by reopening and also shouldering the costs associated with virus suppression and worker protection. The presentation will illustrate the model with case studies, discuss what role personal protective equipment should play in the workplace, and show how a business could control infection rates in a workplace while also generating a positive net profit.

The presentation, "Modeling of Business Reopening When Facing SARS-CoV-2 Pandemic: Protection, Cost, and Risk," will take place Tuesday, August 4, 2020: 1:00 PM to 2:50 PM EDT.
Abstract: https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstractid=313328

WEDNESDAY SYNOPSES

10) Analyzing the fairness of a pre-trial risk algorithm

Judges often use algorithms to help categorize the risk that an arrested person will commit another crime before their trial or fail to appear for court dates. The fairness of these of these pre-trial risk assessment tools has been called into question in recent years, however. In this presentation, Megan Price from the Human Rights Data Analysis Group will present work with colleagues that investigated the fairness of such a tool in San Francisco. Their work looked at the algorithm's sensitivity to "overbooking," where a defendant is booked on more serious charges that are ultimately dropped, sometimes in exchange for a guilty plea to lesser charges. Their results showed that in more than a quarter of the cases, overbooking was associated with the defendants receiving stricter pre-trial recommendations than they would have received otherwise. The researchers say this raises questions about the appropriateness of these tools for high-stakes situations.

The presentation, "Assessing Risk Assessment in San Francisco," will take place Wednesday, August 5, 2020: 10:00 AM to 11:50 AM EDT.
Abstract: https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstractid=309351

11) Data science tools for monitoring patient safety during clinical trials

Clinical trials that study the effectiveness of new drugs—including potential COVID-19 treatments—need to keep careful watch on a multitude of patient health measurements to make sure the regimen is safe. Traditionally, this data has been reported to clinicians in pages of static tables and lists, which makes it hard to spot important patterns. In this presentation, James Buchanan from Covilance and his colleagues from a multidisciplinary working group will present a new, free visualization tool designed to address this problem. The Hepatic Explorer is an interactive open-source web-based data science application for monitoring liver toxicity that allows a researcher to both visualize data as a whole and also explore red-flag areas. This tool could be of particular help during the current pandemic, Buchanan says, because COVID-19 patients often show abnormal liver readings that need to be distinguished from the study drug's effects.

The presentation, "Improved Signal Detection and Evaluation Using New Open-Source Interactive Safety Graphics," will take place Wednesday, August 5, 2020: 10:00 AM to 11:50 AM EDT.
Abstract: https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstractid=309756

12) "Nowcasting" and forecasting COVID-19 **

Modeling the spread of COVID-19 typically takes one of two approaches: mathematical models such as SIR/SEIR that focus on the theoretical mechanisms driving the spread or statistical models more driven by data actually being observed. Lily Wang from Iowa State University and collaborators have developed a modeling approach that combines the advantages of mathematical and statistical models to conduct short-term and long-term forecasts. Their "spatio-temporal epidemic model" (STEM) also allows researchers to take into account the particular characteristics of each county that affect both disease spread and fatalities, such as the mobility, age distributions, health infrastructure, and racial and ethnic demographics. In this presentation, Wang will demonstrate the online dashboard they developed based on STEM, which allows users to visualize, track, and predict COVID-19 infections and deaths. Wang will also reveal the model's latest projections for August through December, and show how they compare with the CDC's reported projections.

The presentation, "Spatiotemporal Dynamics: Nowcasting and Forecasting COVID-19 in the United States," will take place Wednesday, August 5, 2020: 10:00 AM to 11:50 AM EDT. Abstract: https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstractid=309454

13) Text-mining news articles to predict stock returns

Advances in data science and machine learning have made it possible to mine enormous sets of news articles and other text data to capture subtleties and sentiments of the language within. The power of this methodology, however, has yet to be fully applied to the field of finance. Zheng Tracy Ke from Harvard University, Dacheng Xiu from The University of Chicago, and their colleagues have developed a natural language processing method that's specifically designed to mine text documents to predict stock returns. In this talk, the researchers will present results from applying the new methodology to 6.7 million articles from the Dow Jones Newswires, one of the most actively monitored financial news streams. They will show that their approach can be used to investigate how stock prices respond to the news, and also has value for practical asset management.

The presentation, "Predicting Returns with Text Data," will take place Wednesday, August 5, 2020: 10:00 AM to 11:50 AM EDT. Abstract: https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstractid=309567

THURSDAY SYNOPSES

14) Group testing for COVID-19: What is it, and how could it be used more effectively? **

Everyone agrees COVID-19 testing is the key to containing the coronavirus, saving lives, and reopening the economy. The problem: There simply aren't enough tests for all the people we would like to screen. The solution? "Just group it," says Chris Bilder of the University of Nebraska-Lincoln. His work with colleagues on "group testing," also known as "specimen pooling" and "pooled testing," has shown that this clever statistical strategy is technically feasible and could make available testing resources go a lot further. In this introductory overview lecture, Bilder will explain what group testing is, discuss its history and challenges, show how it is being used, and explain how it could be implemented more effectively.

The presentation, "JUST GROUP IT. Group Testing for Identification," will take place Thursday, August 6, 2020: 10:00 AM to 11:50 AM EDT. Abstract: https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstractid=314408

15) Panel on statistical significance and *p*-values

In this session organized by Deborah Mayo of Virginia Tech, four panelists will revisit the debate around statistical significance and *p*-values, including the presuppositions of criticisms that have been raised, the ramifications of reforms that have been proposed, and an appraisal of alternative methods.

The session, "P-Values and 'Statistical Significance': Deconstructing the Arguments," will take place Thursday, August 6, 2020: 10:00 AM to 11:50 AM EDT.
Abstract: https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstractid=309634

16) COVID-19 infectious disease modeling and statistics: Myths, maxims and mobilization **

In this 40-minute ASA Public Lecture, noted epidemiologist Britta Jewell from the MRC Centre for Global Disease Analysis at Imperial College London and noted statistician Nick Jewell from the London School of Hygiene & Tropical Medicine will discuss facts and myths surrounding the COVID-19 pandemic in the US. They will also share insights that can be gleaned from mathematical models and statistical information and will explain what the country needs to do next around collecting and interpreting data.

The presentation, "COVID-19: Infectious Disease Modeling and Statistics—Myths, Maxims, and Mobilization," will take place Thursday, August 6, 2020: 12:00 PM to 1:00 PM EDT.
Details: https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/ActivityDetails.cfm?SessionID=220178

17) New model for contact tracing and disease spread **

Traditional infectious disease models assume that people mix randomly in a population, and that everyone is equally likely to come into contact everyone else. In reality, however, we each have our own network of contacts that we're more likely to mix with, which is why contact tracing is so important during pandemics. Fan Bu at Duke University and colleagues set out to develop a new method combining both approaches and also improving on past models. Their method can account for how our networks evolve during an epidemic as our behavior changes, and how this in turn affects disease spread. The model can also handle real-world situations with only partial data and where uncertainty is important. In this presentation, Bu will use real data from a 2013 flu transmission to show how the new method can incorporate high-tech contact tracing data to improve modeling and forecasting.

The presentation, "Likelihood-Based Inference for Partially Observed Epidemics on Dynamic Networks," will take place Thursday, August 6, 2020: 1:00 PM to 2:50 PM EDT.
Abstract: https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstractid=309898

18) Who are the scientific grant gatekeepers?

The progress of scientific research relies heavily on grant peer review—a process in which scientists agree to spend significant amounts of time anonymously evaluating grant applications submitted by other researchers. Despite its importance, there has been very little data on characteristics of scientists doing the reviews, so a comprehensive survey of scientists was developed and administered to learn more. Stephen Gallo of the American Institute of Biological Sciences will present results showing an uneven distribution of grant peer review participation, with nearly half of all reported reviews done by less than a quarter of all respondents and most reporting they were working at maximum capacity. Implications for the future of science will be discussed.

The presentation, "The Participation and Motivations of Grant Peer Reviewers," will take place Thursday, August 6, 2020: 3:00 PM to 4:50 PM EDT.
Abstract: https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstractid=309505

19) Statistical models for comparing state opioid policies

To address the US opioid crisis, public health experts in different states are trying out a number of policy approaches, such as educating physicians, monitoring prescriptions, and making opioid overdose medication available. Published studies that evaluate these different approaches have increased more than 10-fold in the past 15 years, but it's still difficult to statistically compare the success of different programs across states. In this presentation, Beth Ann Griffin from the RAND Corporation will discuss work with colleagues that examined these statistical methods, especially when a state's particular opioid problems influence the policies it chooses. Simulation results from real-world data will be presented.

The presentation, "Evaluating Methods to Estimate the Effect of State Laws on Opioid-Related Outcomes in the Presence of Selection Bias," will take place Thursday, August 6, 2020: 3:00 PM to 4:50 PM EDT.
Abstract: https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/AbstractDetails.cfm?abstractid=309239

MORE MEETING INFORMATION

USEFUL LINKS
Main meeting page: https://ww2.amstat.org/meetings/jsm/2020/
Online program page: https://ww2.amstat.org/meetings/jsm/2020/onlineprogram/index.cfm

PRESS REGISTRATION
The American Statistical Association will provide free registration to all staff journalists, freelance journalists, and student journalists. Press credentials are approved at the discretion of the ASA. For access, please email *publicrelations@amstat.org*.

ABOUT JSM 2020
Occurring annually since 1974, JSM is a joint effort of the American Statistical Association, Casualty Actuarial Society, Caucus for Women in Statistics, International Biometric Society (ENAR and WNAR), Institute of Mathematical Statistics, Statistical Society of Canada, International Chinese Statistical Association, International Indian Statistical Association, Korean International Statistical Society, International Society for Bayesian Analysis, Royal Statistical Society, and International Statistical Institute. JSM activities include oral presentations, panel sessions, poster presentations, professional development courses, an exhibit hall, a career service, society and section business meetings, committee meetings, social activities and networking opportunities.

ABOUT ASA
The American Statistical Association is the world's largest community of statisticians and the oldest continuously operating professional science society in the United States. Its 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. For additional information, please visit the ASA website at www.amstat.org.

FOR MORE INFORMATION
Regina Nuzzo, Senior Advisor for Statistics Communication
regina@amstat.org

Steve Pierson, Director of Science Policy
(703) 302-1841