# The Year in Review … And More to Come

It has been a real honor to be ASA president and have the privilege of this forum to communicate with you. Some of you have taken the time to share your thoughts about these columns, which I've greatly appreciated.

Next month, the ASA enters its 181st year, with a new ASA president. It certainly will look different from the ASA of 1839, or 1939, or even 1999. In my last column, I will reflect on accomplishments of the past year, challenges ahead, and—last but most important—my experiences with you, our members, without whom the ASA would not exist.

David Williamson is leading the task group ASA as Enabler of Statistical Impact to create mechanisms to proactively identify areas ripe for statistical involvement. Susan Paddock is organizing a series of articles about inspiring topics the task group considered, including those related to autonomous vehicles, evaluating and comparing "deep learning" algorithms, modeling environmental impacts on physical systems, and reproducibility in science. David is also organizing a statistical impact competition, with entries due December 4. Future plans include a collaborative workshop at the ASA to pursue one or more of the topics that emerge as the most exciting. Some avenues for broadcasting these impacts include John Bailer's Stats + Stories, a special section of an ASA-sponsored journal, and possibly a competitive impact session at future JSMs. Stay tuned!

Julia Sharp and her task group are organizing a diversity and inclusion consortium, a forum for representatives of partner societies to share best practices and engage in joint initiatives for attracting diverse talent to STEM fields—statistics specifically. (Someday, perhaps "STEM" will become "STEMS," with the final S for statistics.)

The ASA has had an active Committee on Minorities since 1978, currently chaired by Dionne Swift. Its charge is to foster greater participation in statistics by the many historically under-represented groups and encourage research in the development, evaluation, and implementation of policies and interventions that improve the condition of minority populations in the US. It organizes two successful events each year: the Diversity Mentoring Program at JSM and StatFest, held in Houston this year.

Julia and her colleagues have been developing a dynamic resource repository for students, faculty, and industry partners to identify grants, conferences, and programs that support and enhance diversity and inclusion. The resource list will be disseminated through the ASA website. If you would like to contribute to the list now or in the future, submit your ideas at *https://forms.gle/dkhvp4ficjyrjTo89*.

The task group held its introductory consortium meeting last month with the National Association of Mathematicians and Math Alliance (which organizes the annual Field of Dreams conference with ASA support) and has contacted the Society for Advancement of Chicanos/Hispanics and Native Americans (SACNAS) to be a possible partner. We look forward to seeing the results of these interactions at future conferences.

Jessica Utts and Jun Yang have started an exciting collaboration between statisticians and computer scientists who wish to address the research and practical challenges associated with disinformation. With help from the ASA's staff, they have developed an impressive website of resources, which you can find at *https://sites.google.com/nonlinearlearning.org/confronting-disinformation/home*. They also have identified several important avenues for research (see *https://drive.google.com/file/d/14DAyh6_hBbwYo2aujj7rLfpwGLo9Vtix/view*)and are planning a workshop titled "Research Challenges in Disinformation."

We had another successful JSM in Denver. Its superb program committee, led by Richard Levine, organized a terrific technical program (see *https://magazine.amstat.org/blog/2019/10/01/jsm-2019-reminiscing-on-the-impact*), with exciting presentations, a public lecture, workshops, and courses.

A significant ASA Board action in July was the adoption of a policy concerning appropriate behavior and a mechanism for addressing violations of it. That policy is in line with those endorsed by societies such as AAAS, the Institute of Mathematical Statistics (IMS), and the Statistical Society of Canada. The particulars may differ slightly, but they all agree on the basic principles.

Most exciting for all of us, 2004 ASA President Bradley Efron received the 2019 International Prize in Statistics from Committee Chair Susan Ellenberg in the presence of the five sponsoring societies' 2019 presidents: Helen MacGillivray (International Statistical Institute), Deborah Ashby (Royal Statistical Society), Louise Ryan (International Biometric

Karen Kafadar

Society), Susan Murphy (IMS), and me. In accepting the award, Brad warmly congratulated the statistics profession on a "clean sweep": all five societies had female presidents.

One final challenge, which I hope to address in my final month as ASA president, concerns issues of significance, multiplicity, and reproducibility. In 2016, the ASA published a statement that simply reiterated what $p$-values are and are not. It did not recommend specific approaches, other than "good statistical practice … principles of good study design and conduct, a variety of numerical and graphical summaries of data, understanding of the phenomenon under study, interpretation of results in context, complete reporting and proper logical and quantitative understanding of what data summaries mean."

The guest editors of the March 2019 supplement to *The American Statistician* went further, writing: "The *ASA Statement on P-Values and Statistical Significance* stopped just short of recommending that declarations of 'statistical significance' be abandoned. We take that step here. … [I]t is time to stop using the term 'statistically significant' entirely."

Many of you have written of instances in which authors and journal editors—and even some ASA members—have mistakenly assumed this editorial represented ASA policy. The mistake is understandable: The editorial was co-authored by an official of the ASA. In fact, the ASA does not endorse *any* article, by *any* author, in *any* journal—even an article written by a member of its own staff in a journal the ASA publishes.

Even our own ASA members are asking each other, "What do we tell our collaborators when they ask us what they should do about statistical hypothesis tests and $p$-values?" Should the ASA have a policy on hypothesis testing or on using "statistical significance"?

Sir David Cox wrote in a 1966 *International Statistical Review* article, "Something like a significance test is needed for the essential task of checking and criticizing models and formulating improved ones, a key aspect of successful applied work."

For analyzing randomized clinical trials, John Tukey advised in a 1991 *Controlled Clinical Trials* article, "Results are reported in terms of *both* amount *and* statistical significance."

Robert Abelson in 1997 noted many things—"oboes, band saws, skis, and college educations"—are misused. He then asked, "Will we want to ban effect sizes too, when their misuse escalates?"

Substituting "significance" for another word (meaningful? important?) is not the solution. Indeed, as Yoav Benjamini noted in his Wald lecture given during this year's JSM, the real issue involves not just words, but rather critical statistical concepts including reproducibility, confidence intervals, and multiplicity.

"Sir Ronald's firm knowledge was *not one extremely significant result, but rather the ability to repeatedly get results significant at 5%*," according to Tukey in a 1969 *American Psychologist* article. Tukey also wrote that a point estimate by itself is useless. And he called our attention to "the problem of multiple comparisons" in 1953 (cited for decades as "unpublished manuscript," until all 300 pages appeared in Volume VIII of *The Collected Works of John W. Tukey* in 1994), which inspired Benjamini and Hochberg's false discovery rate in 1995.

To address these issues, I hope to establish a working group that will prepare a thoughtful and concise piece reflecting "good statistical practice," without leaving the impression that $p$-values and hypothesis tests—and, perhaps by extension as many have inferred, statistical methods generally—have no role in "good statistical practice." As Susan Ellenberg noted in her eloquent presentation at JSM (as the first recipient of the F.N. David award), both have served us well in our history, and many of our illustrious colleagues—past and present, frequentist and Bayesian—have relied on them. The ASA should develop—and publicize—a properly endorsed statement on these issues that will guide good practice.

My late business-minded father used to say, "The easy solution to a problem employee is to fire him." In other words, it takes a wise manager to find the talent in an employee and guide him (or her) to use it accordingly. And so it is with $p$-values. The easy way out is to abolish them. Our collective wisdom will be needed to guide others in using them properly. We as statisticians have work to do, and I hope you will help.

I end this column with grateful thanks to the many people who have left me with a lifetime of memories from 2019. It would be hard for me to overstate the impact statistics has had on me, both personally and professionally, especially this year. ASA staff, the board of directors, our hard-working committee and section officers, and our members from all over the world have been ever so collegial, informative, and gracious. The ASA would not exist without you, our members, and its staff to help us move in directions not foreseen by our predecessors. The research you have published, the projects you have conducted, and the communications you have shared have taught me much and so inspired me in my life. Thank you for welcoming me as your president this year, and may you continue to enrich the lives of others as you have so enriched my own.