Success, Opportunities, and Challenges for Statistics and Biostatistics in the Data Science Era

A Report of the July 2016 NSF-Sponsored Workshop for Chairs of Departments of Biostatistics and Statistics

ASA

In July of 2016, forty statistics and biostatistics department chairs gathered at the headquarters of the American Statistical Association (ASA) for a National Science Foundation (NSF)-sponsored workshop featuring a dozen speakers from fields of great interest to the statistics community. For two days, the panelists and attendees discussed the many opportunities and challenges for the discipline of statistics with the emergence of data science and the associated explosion of interest in statistics and biostatistics. This document articulates the major themes arising from the workshop. While intended to be comprehensive, it is also purposefully brief in order to convey succinctly the many important takeaways to audiences with little available reading time.

We summarize the discussions arising from each of the four panels that filled the workshop's two days: (i) Workforce Demand and Needs; (ii) Research Funding; (iii) Education and Research in the Data Science Era; and (iv) Support and Mentoring of Junior Faculty.

Throughout this document, we occasionally provide external links to relevant resources that were not explicitly cited during the panel discussions.

The organizing committee has identified important takeaways for a variety of audiences:

"Our computer science students all take the stats 101 courses…Someone who isn't completely comfortable talking about distributions, inference, maximum estimation, all of these things, is not useful to me as an employer."

—Andrew Moore, Dean, School of Computer Science, Carnegie Mellon University

Department Chairs and Other Academic Leaders: Panelists from the workshop panels had important takeaways for department chairs and other academic leaders. The panelists highlighted the unique skills of statistics students and emphasized the importance of technically well-trained students who also had excellent communication and interpersonal skills. The funding research panel emphasized the many opportunities for statisticians to apply for federal research support, the importance of interdisciplinary research, and the need for statisticians to serve on funding review panels. The panel assumed statistics to be an integral part of data science and encouraged administrators to remove interdisciplinary walls as much as possible. For data science and analytics education programs (e.g., master's or bachelor's degrees in data science and analytics), the importance of designing courses from scratch was stressed, urging programs to avoid simply cobbling together courses from computer science, statistics, information science, and mathematics.

"For the statistics community, I think the job market is great. I can't find enough people who are qualified to do the kind of things that we do."

—Chris Peterson, Vice President, Model Risk Management, Capital One

Students: The panelists listed many qualities and skills desired by employers, from technical and coding skills to writing, speaking, and interpersonal skills. We encourage students to take an active role in ensuring they have the technical statistical and communication skills in demand by employers. The amount of material that can be provided in courses is limited by course time, and students should seek out and take advantage of co-curricular opportunities to supplement their education. Such opportunities include, but are not limited to, internships, research projects, datafests or hackathons, meet-ups and other networking opportunities, and conferences. In a 2015 JSM Encore webinar (*http://bit.ly/2tgoms5*), AT&T's Chris Volinsky advises students on the

To access the online version, visit *www.amstat.org/ASA/Meetings/Department-Chairs-Workshop.aspx*.

PLANNING COMMITTEE MEMBERS

David Hunter (Committee Chair)
Head, Department of Statistics
Pennsylvania State University

Varghese George
Chair, Department of Biostatistics
Augusta University

Xihong Lin
Chair, Department of Biostatistics
Harvard School of Public Health

Sally Morton
Dean, College of Science
Virginia Tech

Sarah Nusser
Vice President for Research
Iowa State University

Jean Opsomer
past Chair, Department of Statistics
Colorado State University

Bruno Sanso
past Chair, Department of Applied Mathematics and Statistics
University of California, Santa Cruz

Arnold Stromberg
Chair, Department of Statistics
University of Kentucky

Steve Pierson
Director, Science Policy
American Statistical Association

Donna LaLonde
Director, Strategic Initiatives and Outreach
American Statistical Association

Members of the planning committee applied for and received funding to support the workshop from the Division of Mathematical Sciences of the National Science Foundation under grant #1628941.

job market to develop a wide range of tools and problem-solving skills so that one isn't forced to try to apply one or two tools to every problem. We encourage all viewers to view the workforce panel videos as well as those of Rappa and Malone to hear what employers are looking for in new hires.

Junior Faculty: For junior faculty, we highlight recommendations from the research funding panel and the supporting and mentoring panel. Statisticians, including junior faculty, are strongly encouraged to participate in funding review panels at NSF, NIH, and other funding agencies when the opportunity to do so arises. Such service benefits reviewers as well as the statistics profession and science generally. The mentoring panel encouraged junior faculty to adopt a holistic view of mentoring in which, for example, tenure is merely one step in a lifelong career, not an end in itself, and mentoring does not end with the award of tenure. Junior faculty were encouraged to cultivate both intradepartmental and interdepartmental relationships.

The Statistics Community at Large: As advocates for the practice and profession of statistics, members of the broader community will find the recommendations of the workforce panel useful in terms of the unique skills that statisticians bring to any enterprise. The data science panel provides context to understand the skills and knowledge early-career statisticians are acquiring.

Planning for the Workshop

Statistics and biostatistics departments across the country are being asked to deal with enormous growth in statistics course enrollments and degrees. They are also working to respond to both employer demands, for graduates with domain knowledge and collaboration skills, and institutional pressure, to develop new undergraduate and graduate degree programs in data science and analytics. Their faculty members are highly sought after for scientific collaboration and occasionally private sector consulting, so chairs must be able to provide mentoring to ensure departmental stability. The American Statistical Association's Caucus of Academic Representatives (CAR) organized a July 2016 workshop to address these challenges for department chairs, who must provide leadership on research, educational, and personnel issues on their campuses.

The CAR is composed of department chairs, departmental representatives, and other academic administrators. The purpose of this caucus is to promote the statistics discipline within the academic community, to provide resources for academic biostatisticians and statisticians to successfully advocate for the discipline, and to be a conduit between the academy and other sectors of the profession. In support of this mission, CAR has sponsored a half-day workshop for department chairs at the Joint Statistical Meetings since 2006. Given this history and mission, the July 2016 Workshop Planning Committee was chaired by the 2016 Chair of CAR; other members of CAR's executive committee also served on the planning committee. Additional planning committee members were invited to serve to ensure diverse representation in terms of size and type of academic department.

Panel Topics and Panelists for the Workshop

The planning committee chose to organize the two-day workshop around four panel discussions, two on each day of the workshop. Department chairs serve multiple constituencies, and it was in large part the concerns of these constituencies that informed the panel themes. Once the four themes were decided based on the organizers' sense of the most pressing issues facing departments of statistics and biostatistics, the organizers identified and recruited a set of experts on each of the four themes. The themes and the panelists who attended the workshop are as follows:

Panel 1: Workforce Demand and Needs

- **Erik Andrejko***, Vice President, Science, The Climate Corporation [Video†]
- **Erica Groshen***, Commissioner, Bureau of Labor Statistics [Video, Slides†]
- **David Morganstein**, Vice President and Director of the Statistical Staff, Westat [Video, Slides]
- **Chris Peterson**, Vice President, Model Risk Management, Capital One
- **Aarti Shah**, Senior Vice President and Chief Information Officer, Eli Lilly and Company [Video]

Panel 2: Research Funding

- **Chaitan Baru**, Senior Advisor for Data Science, NSF Computer and Information Science and Engineering (CISE) Directorate [Slides]
- **Michelle Dunn***, Senior Advisor for Data Science Training, Diversity, and Outreach, NIH Office of the Associate Director for Data Science (ADDS) [Slides]
- **Michael Lauer**, Director, Extramural Research, NIH [Slides]
- **Michael Vogelius***, Director, NSF Division of Mathematical Sciences (DMS), Mathematical and Physical Sciences (MPS) Directorate

Panel 3: Education and Research in the Data Science Era

- **Chris Malone**, Professor, Department of Mathematics and Statistics, Winona State University [Video, Slides]
- **Andrew Moore**, Dean, School of Computer Science, Carnegie Mellon University [Video, Slides]
- **Sarah Nusser**, Vice President for Research, Iowa State University [Video, Slides]
- **Michael Rappa**, Founding Director, Institute for Advanced Analytics, North Carolina State University [Video, Instead of using slides, Rappa spoke while scrolling through the program homepage]
- **Chris Wiggins**, Chief Data Scientist, *New York Times* and Associate Professor, Department of Applied Physics and Applied Mathematics, Columbia University [Slides]

Panel 4: Support and Mentoring of Junior Faculty

- **Genevera Allen***, Assistant Professor of Statistics, Computer Science, and Electrical and Computer Engineering, Rice University [Slides]
- **Sally C. Morton**, Dean, College of Science, Virginia Tech [Slides]

*Titles and affiliation are listed for panelists based on their position as of the workshop in July 2016. Four asterisked names have since changed affiliation as of the publishing of August, 2017. Allen's affiliation has not changed but she has been promoted to associate professor.

†Links to slides and videos are available at *www.amstat.org/ASA/Meetings/Department-Chairs-Workshop.aspx*.

Summaries of Panel Discussions

Editor's note: The pull-out quotes that follow were drawn from transcripts of videos taken during the morning sessions. Due to limited staff resources, video was not taken in the afternoon sessions, so no quotes are provided for afternoon speakers.

Workforce Demand and Needs
(Morning of Day 1)

"All these occupations [mathematicians, actuaries, statisticians, mathematical science teachers, and survey researchers] are expected to grow quite substantially, faster than the average for all occupations. The factors driving this growth are increased use of data to make informed decisions across all sectors and also of course the increase in the data that are available to be analyzed. More and more, throughout our economy, people are making evidence-based decisions, and the folks that you are training are the ones who are going to be providing that evidence."

—Erica Groshen, Commissioner, Bureau of Labor Statistics

"It's given [statistics graduates] will come in with deep technical expertise… In addition to that, what is important? First and foremost I'm going to talk about communication. Communication, communication, communication."

—Aarti Shah, Senior Vice President and Chief Information Officer, Eli Lilly and Company

The intent of this panel was to spur a conversation between statisticians who work in industry and government and the academic community as represented by chairs of departments. Moderated by Bruno Sanso, the panel included Erica Groshen, the commissioner of the Bureau of Labor Statistics, as well as four statisticians working in various industries: Erik Andrejko from The Climate Corporation, David Morganstein from Westat, Chris Peterson from Capital One, and Aarti Shah from Eli Lilly. Dr. Groshen summarized past and predicted trends in demand for statisticians in the U.S. workforce, while the four representatives from industry discussed their experiences as statisticians along with the types of skills their companies valued in the statisticians they hire.

Following is a brief summary of some salient points raised during the panel discussion:

1. **Statisticians are and will be highly employable.** According to BLS projections, the demand for statisticians in the U.S. will grow 33% from 2014 to 2024. Indeed, BLS itself is a major employer of statisticians. Furthermore, jobs in statistics tend to pay relatively well. The BLS projections are based on job title, so they likely underestimate the true demand since statisticians are a part of the data science workforce. A 2017 report of the Business Higher Education Forum and PricewaterhouseCoopers, "Investing in America's Data Science and Analytics Talent (*http://bit.ly/2of2PSt*)," is the latest report further supporting the demand for statistical skills projecting the data science and analytics-related job postings to grow from 2.35 million in 2015 to 2.72 million by 2020. The report updates the oft-quoted 2011 McKinsey Big Data report (*http://bit.ly/2dDfLuS*), "A significant constraint on realizing value from big data will be a shortage of talent, particularly of people with deep expertise in statistics and machine learning, and the managers and analysts who know how to operate companies by using insights from big data…we project that demand for deep analytical positions in a big data world could exceed the supply being produced on current trends by 140,000 to 190,000 positions." For more on this theme, see *bit.ly/ResourcesForChairs*.

2. **Statisticians' specific skills are in high demand.** Multiple panelists emphasized that statistical techniques cannot effectively be treated as mere "black boxes"; understanding of underlying fundamentals is vital. Quality control, study design, inferential techniques, and understanding of data quality are important in multiple organizations throughout whole processes, not merely

design and analysis phases. Indeed, in 2014, LinkedIn analyzed the skills and experience data of its 330 million member profiles and found "Statistical Analysis and Data Mining" to be the hottest skill. More about the utility and demand for, and high rankings of, statistics degrees can be found at *bit.ly/ResourcesForChairs.*

3. **Industry needs well-rounded employees who also have sound statistical training.** Each of the industry representatives strongly emphasized the importance of communication skills among its employees. Several used the shape of the capital "T" to illustrate the ideal of an employee who has deep knowledge in one area—statistics, in this case—while also able to communicate broadly with colleagues with different areas of expertise.

"When you look at what employers want, they actually want a bundle of things. In fact, they're not even looking for someone who knows everything technically, in fact that might be a red flag for an employer. They want someone who knows something but also knows how to continue to learn and that they don't know everything that they need to know to sort of engage in a particular job."

—Michael Rappa, Founding Director, Institute for Advanced Analytics, North Carolina State University

4. **Beyond communication skills, statisticians in industry must master numerous non-technical skills**. These include organizational and leadership skills, curiosity and openness to lifelong learning, ability to work collaboratively and in a diverse team, commitment to acting ethically, and business savvy.

5. **Computing is increasingly important for statisticians.** Employers prefer, and in some cases require, experience with coding, relevant software, and version control among even its statisticians. Sufficient computing background to enable one to learn new computing techniques, an inevitable part of a modern job as a statistician, is strongly recommended.

6. **Degree programs may not provide all of the skills needed by successful statisticians.** Given all of the non-technical skills required (the horizontal bar on the "T"), along with the importance of an attitude of self-learning, students should be encouraged to develop these skills beyond mere degree program requirements. Internships and other co-curricular experiences can be useful for exposing students to the type of initiative they will need to show as individuals.

"The last thing is…these employers are really looking for someone who isn't obsessed with data but interested in moving their organization forward, that they have contextual understanding."

—Michael Rappa

7. **Following the panel's presentations, audience members and panelists discussed multiple issues.**

    a. There is a tension between the two parts of the "T," that is, preparing students deeply in statistical methods and theory on one hand, and exposing them to a broad range of software and non-statistical topics such as communication skills on the other.

    b. Many agreed that teaching students how to learn is more important than trying to predict each individual skill that will ultimately be useful in the workplace.

    c. The question was raised as to whether more statistics instruction could be pushed to the pre-college years, particularly high school. If this happens, what role should (university) departments of statistics and biostatistics play?

"The other thing that I would mention is coding ability: having students coming out of school who have the ability to actually write quality code, know what encapsulation means, know how to comment code, can use modern version control…"

—Chris Peterson

d. Scientific communication and business communication often have different styles and different aims; neither is necessarily more important than the other. The former is often deliberative and rigorous, whereas the latter can be more effective when combined with narratives.

Research Funding Panel
(Afternoon of Day 1)

"Teamwork is prized by employers, absolutely prized. They'll use a kind of language for it, they'll say, "We're looking for a cultural fit. We're looking for fit in the position…"

—Michael Rappa

"We find the magic happens when you connect discipline A, to discipline B, to discipline C and get those individuals in a small room, give them an interesting problem, and see what comes out of that."

—Erik Andrejko, Vice President, Science, The Climate Corporation

"Organizational skills [are important]. How do you organize the work?…Few if any of [our new hires] are trained in ways to organize their work. They know how to solve the problem. They might know what is the right analytic model. They might be able to do that in the order that it's asked, but that's not the same thing as say, how long will it take you [to complete the project]…"

—David Morganstein, Vice President and Director of the Statistical Staff, Westat

Given the importance of external funding in scientific research generally and also, increasingly, in statistical research in particular, we felt a conversation between department heads and representatives of governmental funding agencies would aid communication in both directions. That is, we wanted to give statisticians a chance to hear about the vision and concerns of the funding agencies while affording them the chance to provide feedback to these same agencies. Moderated by Varghese George and Arny Stromberg, the panel included Michael Lauer and Michelle Dunn, along with Chaitan Baru and Michael Vogelius. Dr. Lauer, NIH's director of extramural research, discussed problems of reproducibility in science and how statisticians might help, while the other three panelists offered viewpoints on statistical contributions based on their particular positions within their agencies.

Following is a brief summary of some salient points raised during the panel discussion:

1. **The reproducibility crisis in science has raised awareness about the importance of engaging with statisticians.** Relatedly, statistical training for non-statistician researchers should be improved. The engagement of statisticians in scientific studies varies across disciplines. For example, it is currently better for clinical than for pre-clinical research, which points to opportunities for statisticians in clinical research—as exemplified, for instance, by the introduction of rigorous clinical design concepts by the Howard Hughes Medical Institute—as well as areas in which more work may be needed to increase statistical engagement.

2. **Research funding agencies are actively seeking to harness and influence data science for their missions**, providing many opportunities for statisticians. Statisticians should proactively take advantage of these opportunities, especially as other disciplines have taken a lead in data science. The many NSF and NIH programs to support big data and data science are outlined in the slides of Dunn, Baru, and Vogelius; interested readers should consult the agency and program websites for up-to-date specifics. NSF strongly encourages collaboration of mathematicians, computer scientists, and statisticians on data science.

3. **Multiple challenges hamper the engagement of statisticians in broader scientific review, particularly at NIH.** It has proven difficult for NIH's scientific review officers (SROs) to convince statisticians to serve on NIH study sections. In turn, the lack of statisticians on study sections leads to poor visibility of, and lack

of respect for, statistical contributions. This issue is exacerbated by the fact that study sections are usually domain-specific, whereas statisticians are often generalists.

4. **It is not clear how much and what type of statistical training a non-statistician should ideally have**. In other words, when should a researcher visit a statistician as opposed to possessing the requisite statistical expertise him or herself? The answer to this question has ramifications both because academic statisticians are often the ones who teach the statistics courses taken by non-statisticians and because we want non-statisticians to possess enough training to recognize the importance of statistics in the scientific endeavor.

5. **Junior faculty members do not necessarily receive high-quality mentoring about participation on review panels.** The chairs and panelists discussed the challenges of how to advise junior faculty on collaborating on proposals, noting that it would be nice to advise early-career faculty not to join proposals for less than, say, 10% of effort. However, they also acknowledged that arbitrary minimums are too rigid and not realistic. Thus, department chairs should play a proactive role in assuring that junior faculty members receive good advice so that they can be involved in the funding process without being taken advantage of.

6. **The chairs and speakers made many recommendations to address the issues raised during this panel.**

    a. The statistical community should engage federal agencies and other scientific leaders to convey the cultural message to scientists and engineers that they should think ahead to engage statistical scientists where appropriate and that to do so is beneficial in all phases of the scientific enterprise.

    b. It is important for statisticians serving on study sections to understand how to be effective study section members. The ASA Committee on Funded Research is working on a document on best practices that is expected to be ready in the latter half of 2017.

    c. The ASA should work with research funding agencies to help disseminate best practices regarding panel officers' engagement with statisticians. The ASA could also help to connect funding panel officers with department chairs to help them identify statisticians and biostatisticians for their panels.

    d. The ASA could work with research funding agencies to ensure that the next generations of scientists, especially biomedical researchers, have sufficient statistical training. One recommendation in this regard was for NIH non-statistics T32 trainees to receive a week of statistics training.

"Communicating with the client [is another important skill]: Giving them an outline and saying, is this the right thing? Is this what you're looking for? Sending them early drafts. These are all things that make us as a company more effective for sure, and we can use more of those skills from students and new employees. "

—David Morganstein

e. The ASA could issue best-practice documents to address the issues raised here. For example, the Clinical and Translational Science Awards (CTSAs) program (*https://ctsacentral.org/consortium/best-practices*) issues many best-practice documents, the following possibly being of particular interest: Evaluation Metrics for Biostatistical and Epidemiological Collaborations (*www.ncbi.nlm.nih.gov/pubmed/21284015*) and Principles of Community Engagement 2nd Edition (*www.atsdr.cdc.gov/communityengagement*).

Education and Research in the Data Science Era
(Morning of Day 2)

The rapid increase in the demand for data science training is profoundly affecting statistics education. Moderated by Dave Hunter, our data science panel consisted of academicians at a variety of institutions and in a variety of jobs discussing their thoughts on how we train our students in the age of data science. Chris Wiggins, associate professor of applied physics and applied math at Columbia University and also the chief data scientist for the *New York Times*, discussed the history of data science and offered thoughts on where it goes from here. Andrew Moore, dean of computer science at Carnegie Mellon University, talked about the effective symbiosis among computer scientists, statisticians, and others at data science nexus. Chris Malone and Michael Rappa described two academic programs that successfully incorporate data science, the undergraduate statistics major at Winona State University and the Master of Science in Analytics at North Carolina State University, respectively. Finally, Sarah Nusser offered her perspective, as Vice President for Research at Iowa State University, on open and reproducible science.

"One of our absolute requirements in all machine-learning-related curriculum and our computational-related computer science curriculum is that you begin with probability and statistics."

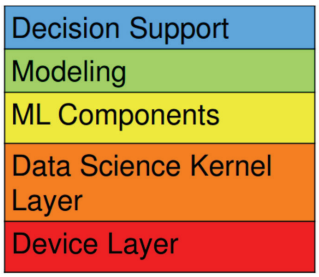
—Andrew Moore

"The whole point of university systems is so that people can become world class in some areas and, while they do it, they have to have respect for the other areas."

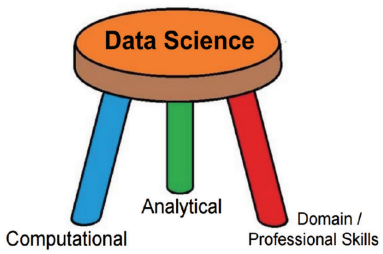
—Andrew Moore

Here are some of the issues raised during this discussion:

1. **The emerging field of data science continues to evolve. It must incorporate statistical thinking and it is not a subset of any other field.** Multiple speakers presented graphics capturing elements of the interdisciplinary nature of current data science programs. In explaining the data stack shown below, Andrew Moore said people should concentrate on becoming an expert on one or two layers of the stack and then learn to rely on and work with experts from other layers of the stack. Chris Malone presented the data science stool shown below to represent the three core areas of the discipline. Chris Wiggins presented a 2010 Venn diagram due to Drew Conway that was an early attempt to characterize the intersecting disciplines that contribute to data science. During the first day's afternoon session, Michael Vogelius listed mathematics, statistics, and computer science as the three foundational disciplines of data science.



Andrew Moore: The Data Stack©



Chris Malone: The Data Science Stool

2. **There can be a distinction between academic training in data science and data science skills required by industry.** Wiggins pointed out that data science in academia tends to be viewed via the broad fields it encompasses, whereas data science in industry tends to be viewed as a list of skills. In this sense, data science resembles many long-established fields, in which education strengthens students' technical background and mental flexibility rather than providing direct vocational training. Indeed, as Malone repeatedly stated, a viable academic data science program need not be staffed by data scientists—thus, the reality that most faculty lack experience as data scientists is not necessarily problematic.

3. **Both undergraduate and graduate data science programs are viable, important, and in demand.** Programs such as the masters of analytics at North Carolina State, as described by Rappa, have enjoyed tremendous success by, among other things, engaging employers to make sure that these programs meet employer demands. Malone reinforced the importance of engagement with industry representatives in developing Winona State University's undergraduate program. Another common thread among successful data science programs discussed by the panelists is that they begin from desired learning outcomes of the program rather than simply amalgamating existing courses in statistics, math, and computer science.

"We took those five things [that employers want] and then we literally built a curriculum from scratch, lecture by lecture, sort of threading these lecture streams together.

—Michael Rappa

4. **Interdisciplinary focus is essential for successful research programs.** Moore emphasized that statistics and computer science need each other and that success comes from focusing on solving real problems and ignoring disciplinary boundaries. This fact jibes with an existing culture within the statistics community to engage in meaningful collaborative work with scientists in diverse fields. Speakers in multiple panels highlighted the increasing rate of publication by statisticians in machine learning and CS journals, a trend viewed positively by everyone who commented on it.

"Undergraduates have to convince employers of what they can do. Employers aren't ready for undergraduate data science students…So that professional skills development class that we were talking about, that's what we help them with. Marketing themselves. Branding themselves. Telling them that they can't look for statistician in a job title when they're looking. They don't know that."

—Chris Malone, Professor, Department of Mathematics and Statistics, Winona State University

"We have this culture…you're not here to do stuff on behalf of your department or your college. You're trying to build something for the rest of the world. So you've got to pick something big that is not just statistics for its own sake or computer science for its own sake, you take some of the big challenges of the world and you get the students, those young idealistic crazy people, excited about those and then the faculty help them. That's when, if you've got a discipline which is worth anything, you find that your discipline can help in solving the problem."

—Andrew Moore

"If you have to share data, you have to think about sharing that data early on. Here's a fantastic opportunity to really think about design proactively."

—Sarah Nusser, Vice President for Research, Iowa State University

5. **The best traditions of open science, in the age of data science, require novel ways of thinking about how data are curated and shared.** Nusser discussed multiple issues involved in developing principles and ethics for open data. Appropriately sharing data can increase public trust in scholarship, among other benefits to science, and there is a strong role for statisticians to play in this development.

6. **Existing statements on data science by professional associations reinforce the ideas discussed in this panel.** Many of the above characterizations are consistent with the ASA Statement on the Role of Statistics in Data Science (*http://bit.ly/1nI0Cdu*) and the Computer Research Association statement entitled Computing Research and the Emerging Field of Data Science (*http://cra.org/data-science*). The former lists statistics as one of three foundational areas of data science along with computing and data management, while the latter says the "emerging discipline relies on a novel mix of mathematical and statistical modeling, computational thinking and methods, data representation and management, and domain expertise," and that "Data science starts with a strong set of foundations adapted from several fields including statistics, mathematics, social science, natural sciences, and computer science."

Support and Mentoring of Junior Faculty
(Afternoon of Day 2)

The focus of this panel was to address the question, "who are junior faculty?" with the goal of identifying strategies to support effective mentoring programs. Jean Opsomer served as moderator for this panel. Presenters were Sally Morton, Dean of the College of Science at Virginia Tech, and Genevera Allen, assistant professor in the Department of Statistics at Rice University.

1. **Newer generations of faculty are motivated by many factors, including a desire to make a difference and a focus on open and reproducible research.** The research of the new generation of young faculty is becoming ever more interdisciplinary, resulting in myriad rippling practices. For example, interdisciplinary research papers tend to be published in domain-specific journals and/or conference proceedings rather than statistics journals. Workshop participants learned that young statistics and biostatistics faculty members are pursuing increasingly diverse and interdisciplinary research and that they are motivated more than ever to make a difference on scientific or societal problems. In this context, "interdisciplinary" research is more than mere statistical consulting and involves a lengthy process of scientists from multiple disciplines truly collaborating to solve scientific problems.

2. **The tenure consideration criteria are changing to reflect the increasingly interdisciplinary research of faculty.** Departments seem to be in various stages of modernizing their criteria for promotion and tenure in order to retain excellent faculty members whose work reflects the trends described above. It may even be the case that non-traditional means of disseminating information,

such as blogs or software repositories, are becoming as important as published papers in terms of their scientific impact. In addition to the listserv of the Caucus of Academic Representatives, there are many resources available to departments as they consider modernizing their metrics of success. For example, Sally Morton is a co-author on a paper titled Evaluating Academic Scientists Collaborating in Team-Based Research: A Proposed Framework. The ASA may also want to develop guidance documents like those of the American Mathematical Society, the titles of which include Citation and Impact in Mathematical Publications, The Culture of Federal Support for Academic Research in Mathematics, and Rates of Publication. For combined mathematics and statistics departments, we refer readers to the American Statistical Association (ASA) Endorsement of the Mathematical Association of America (MAA) "Guidelines for Programs and Departments in Undergraduate Mathematical Sciences" (*http://bit.ly/2ubfbOd*).

3. **Work/life balance is a critical issue.** Mechanisms for reaching the appropriate equilibrium should be addressed directly, not left to informal conversation. It is essential that departmental and university expectations are clearly communicated. To this end, junior faculty members should be encouraged to seek advice outside their departments as appropriate, for instance, from human resources offices that can provide information about personnel policies.

4. **Mentoring should and does occur both formally and informally.** Mentoring is distinct from performance evaluation. Offering junior faculty members multiple opportunities to receive mentoring both within and outside their departments will allow them to find the best fit. Often multiple sources of advice are more helpful than one; department chairs should avoid taking on the full responsibility for mentoring junior faculty members, particularly since they (chairs) have a formal evaluative role to play. Furthermore, some institutions have explicitly established a right of every junior faculty member to a mentor outside her department who is forbidden from discussion with that department prior to tenure. Where possible, official credit for service should be awarded to those who formally serve as mentors.

5. **Networking is critical.** Departments should provide tangible support for networking with colleagues, for instance by supporting travel to conferences, facilitating invitations to present to other research groups, and assisting in the development of connections with funding agencies.

6. **Mentoring is a long-term proposition.** Effective mentoring takes one's entire 30- to 40-year career into account. To this end, a mentor can help a junior faculty member clearly articulate her goals. In particular, the attainment of tenure, if applicable, is merely one step in an entire career and should not be seen as the ultimate goal of an effective mentoring program.

7. **Following the panel's presentations, audience members and panelists discussed multiple issues.**

    a. Post-tenure faculty members face a slightly different set of issues than pre-tenure faculty. For instance, lack of startup funding for post-tenure faculty might require occasional financial commitments from departments, if possible, to support networking opportunities.

    b. The security afforded by a tenured position may be seen as less important to the current generation of young scientists than it was to previous generations. Indeed, many seemed to feel that the abolishment of tenure altogether would not be a bad thing. However, it clearly remains a perk and while many administrators may feel ambivalent about abolishing it, they presumably don't want to be the only university to do so or they would be at a competitive disadvantage.

    c. In mentoring junior faculty, it is important to convey a sense of the community and its expectations. There is an American Mathematical Society online document (*www.ams.org/profession/leaders/culture/culture*) that explains its culture. Producing such a document might be a useful activity for the CAR and/or the ASA.

"We mentor faculty who seem to be trying to be a lone hero. In the complex world that we live in there are very few fields where you can actually get anything done if you don't collaborate. We coach for that. We, the leaders, talk about it all the time."

—Andrew Moore

Summary

Participants and presenters in the workshop's four panels developed consensus on numerous important themes. The workforce panel and subsequent discussion reinforced the current and future demand for individuals with substantial statistical education and who have the ability to communicate and collaborate. The data science panel recognized that an explicit definition of data science is evolving but affirmed the essential role of statistics, mathematics, and computer science in both the theoretical foundations and applications of data science. From the discussion, it is clear that successful programs will exploit the interdisciplinary nature of the domain and will form substantive partnerships both with multiple academic departments and with industry. The research funding panel highlighted the need for statisticians to be involved in the proposal review process to ensure the quality of the science that is funded. The discussion also highlighted the need for interdisciplinary research to take on the challenging problems in data science. The mentoring panel acknowledged the changing nature of the dissemination of scientific accomplishments and the impact these changes will have on academic career progress. The importance of career-long mentoring was also highlighted. Finally, the benefit of providing chairs opportunities to discuss issues and concerns with each other became apparent as perhaps the most important outcome of the workshop.

Related Resources and Dissemination

This document represents one of several efforts to disseminate the outcomes of the workshop as proposed to NSF. The organizing committee believes such a document is important not only as a reference for chairs who attended or who wanted to attend the workshop but also as a means to reach the diverse audiences listed at the beginning of the document. Additional dissemination efforts include a summary and discussion presented by the organizing committee during the annual half-day workshop, organized by CAR, preceding the Joint Statistical Meetings in August 2016; an article in the November 2016 issue of *Amstat News (http://bit.ly/2sSr07M)*; the online agenda with video and slides posted on the ASA meetings page for the workshop (*http://bit.ly/2exKlav*); and a final report to NSF, to be prepared by the organizing committee by August 31, 2017. The full video playlist is available at *http://bit.ly/2ukJtOL*.

Department chairs may also find helpful resources for advocating for statistics and biostatistics to university administrators, students, and the public at large listed in the draft ASA document entitled Statistics and Biostatistics Department Chair Resources. Comments on this document are welcome and should be sent to Steve Pierson (*pierson@amstat.org*).

A list of workshop participants can be found at *http://ww2.amstat.org/misc/ParticipantsList.pdf*.

Next Steps

The workshop participants and the organizing committee spoke very positively about the workshop and recommended it be done biennially if not more frequently. They recommended keeping the annual half-day workshop at the JSM location but appreciated the two-day workshop at ASA headquarters as a chance to discuss issues they face in more depth with fellow chairs and invited speakers. They also appreciated the opportunity to learn more about the ASA and its staff that is afforded by holding the workshop at ASA headquarters.

The workshop also illustrated the importance of the CAR and of strong lines of communication among department chairs as well as among chairs, workshop speakers, and ASA staff. Such interaction not only helps the department chairs in their work for their departments but also promotes the discipline of statistics and the growing community of statisticians.