

JULY 2016 • Issue #469

AMSTATNEWS

The Membership Magazine of the American Statistical Association • <http://magazine.amstat.org>

LeVange Elected President of the ASA

Williamson Elected
Vice President



ALSO:

Reproducible Research
in *JASA*

The Makings of a Survey
Methodologist—and
a Survey Methodology
Conference

ASA LEADERS REMINISCE

Lynne Billard

In the 19th installment of the Amstat News series of interviews with ASA presidents and executive directors, we feature a discussion with 1996 ASA President Lynne Billard.

Q In the 1980s, you worked on research projects designed to increase understanding of the incubation period of the human immunodeficiency virus. This was at a time when AIDS was poorly understood and greatly feared. What misunderstandings about HIV and AIDS did you and your research partners address in this research, and what were the ramifications on public health education in the United States?

A The most important collaboration on HIV/AIDS was our work (Medley et al. in 1987 and 1988; Billard et al. in 1990) on the mean incubation period between becoming infected with HIV and being diagnosed with AIDS. The data set consisted of the entire U.S. data of those who had received infected blood transfusions and been diagnosed. Prior to our work, this incubation period was effectively calculated by averaging the times for known diagnosed individuals. The basic set-up is that we had truncated data. However, unlike then-truncated data sets, we only had those observations that had actually been diagnosed (i.e., the start and end points were known).

There were clearly other observations out there, but we did not know about them yet because they were still undiagnosed. This meant we did not have any truncated times, nor did we know how many there were. Therefore, we had to build a



Billard

model that included distributions of the unknown truncated times and that estimated the number of unknown observations.

Another innovation was to divide the data into age groups— young children, adults, and the elderly. The results differed by age group—the mean incubation period, based on a Weibull distribution, was shorter for the young, about two years with a standard deviation of 1.25, because their immune systems were not fully developed. We saw a similar result for the elderly, with a mean of about 5.6 years and standard deviation of 2.1, because they were receiving blood transfusions reflecting their not-so-healthy condition. More importantly, the average time for adults was 8.1 with a standard deviation of 3.6; this quickly became a 10-year figure.

At the time, 17- to 25-year-olds were viewed as being the

most vulnerable cohort and also a cohort most influenced by their peers. However, when this incubation period was closer to 10 years—instead of 2-3-4 years—rather than seeing their friends diagnosed while still college mates, so to speak, 10 years meant those college friends were no longer close by to influence their behavior. Thus, for the health educators, it was imperative to make radical changes to the way they approached this issue.

The impact of these results are still vivid to me. Let me back up a bit. I had gone on leave to Imperial College to work with David Cox on some other topic. The week before arriving, Roy Anderson, a renowned population biologist who at the time was unfamiliar with epidemics, approached David for help with the data he had obtained from CDC [Centers for Disease Control and Prevention]. David knew I had worked in epidemic theory and so asked if we could change plans to work on this issue. Sure! That was July or maybe early August of 1986. By October, the mathematics was completed; however, I did not know how to run their computer and we did not want to wait until my return to Georgia in January of 1987. Therefore, one of Anderson's doctoral students, Graham Medley, was brought in to assist in the programming on the Imperial computer. By November, we had our results.

They were startlingly different from previous results. David and I knew the mathematics was correct. The only question was whether there were enough data to ensure robustness; the 1990 paper answered that question affirmatively. Armed with the health education knowledge, rather than statistical rigor, we were convinced by the ethical humanitarian arguments that the results had to be announced then and not later. Therefore, I went back to my office and proceeded to write it all up.

By early 1987, I mailed the draft manuscript to David. Soon thereafter, the journal *nature* asked for a summary of the 1987 paper; the rest of my draft contained the details that came out in the *Proceedings of the Royal Society B* in 1988. I always think of those two papers as one.

It was only *yesterday*—or so it seems; actually, it was 1987—when I heard the first public service announcement on NPR whilst driving into work on the education of the public about the ramifications of being infected with HIV. It was stupefying. I just sat there and marveled that statistics could come up with a real-life result so quickly, a result that would alter the way people saw things, at least as they related to this disease.

Years later, in 2014, when working on a medical boat on the Amazon River, a health teacher from Oregon relayed a story explaining how those NPR announcements changed her life. We marveled at the *smallness* of our world. Here we were, an Australian and an American in Brazil discussing work done in England. Somehow, Alaska came into the equation, too. Such is the impact of, or should we say the breadth of, the world of, statistics!

Lynne Billard earned her BS with first class honors from Australia's New South Wales University in 1966. As an undergraduate student, she was employed as a statistician with the Department of Main Roads in Sydney in 1963–1964 and as a statistician with the Colonial Sugar Refinery in 1964–1965. She earned her PhD in 1969, again from Australia's New South Wales University.

Billard then held the position of lecturer at the University of Birmingham through 1970. She has since served on the faculty of the University of Waterloo, Florida State University, and the University of Georgia. She has also been a visiting faculty member with SUNY at Buffalo and Stanford University and a research fellow with the Naval Postgraduate School and University of California at Berkeley. In 2009, Billard was named an honorary professorial fellow by the University of Melbourne. She also held several administrative positions while with the University of Georgia, including head of the department of computer science and statistics from 1980–1984, head of the department of statistics from 1984–1989, and associate to the dean from 1989–1991. She was named a distinguished university professor by the University of Georgia in 1992.

Billard has published more than 250 articles in a wide range of journals, as well as eight books. In 1990, she received the ASA's Outstanding Statistical Application Award for her work with Graham F. Medley, David R. Cox, and Roy M. Anderson on the distribution of the incubation period for AIDS, which was published in the *Proceedings of the Royal Society, London, Series B* 233 in 1988.

In addition to being the 1996 ASA president, Billard was international president of the International Biometric Society from 1994–1995. She has served on countless national and international committees for the ASA, International Biometric Society, International Statistical Institute, Institute of Mathematical Statistics, Committee of Presidents of Statistical Societies, Interface Foundation of North America, and American Association for the Advancement of Science. She says the most interesting of these experiences was the appointment to the secretary of commerce's Census 2000 Advisory Board.

Billard has also received numerous awards, including ASA Fellow in 1980, Samuel S. Wilks Award in 1999, and ASA Founder in 2003. Recently, she was inducted into the Slovenian Statistical Society as an honorary fellow.

Q You established the “Pathways to the Future,” an annual National Science Foundation workshop from 1988 to 2004 that focused on mentoring women who had recently received PhDs in statistics and wanted academic careers. How did the focus of these workshops evolve over its 16-year life?

A Let me first say that, for some of those years, the Office of Naval Research also funded these workshops; however, the original funding did originate with NSF [National Science Foundation].

It would be great to be able to say that inequities had evolved to the disappearing point to the extent that the workshops were no longer necessary. Unfortunately,



Lynne (left) with her mother, Chris Billard, in 2001. The UNSW mathematics cadetship (see Q3) and her mother were key ingredients to Lynne's career.

while perceptions are that these inequities no longer exist, the data suggest otherwise. By and large, initial hiring is not a problem. Problems lie in inequitable promotion and tenure rates and in salaries. Karen Kafadar and I reported on this in 2015 in *Advancing Women in Science*, where we looked at national academic data up to 2014. Both these aspects depend on subjective evaluation of faculty work in varying ways. Therefore, the need for the workshop focus still pertains today.

Given the unfortunate but stark reality, the workshop would open with presentation of the latest data, followed by several discussions and sessions designed to help participants not become victims to those realities. This included addressing issues such as the importance of publishing their research; how to respond to deflating referee reports—deflating at these early stages of a career since most young researchers are so sure “everyone” will acknowledge and appreciate their work—the importance of attending and presenting their work at meetings; and grants, teaching, promotion, and tenure—the usual steps along an academic career. The success rate of pathways alumnae is very gratifying indeed.

Q Prior to earning your PhD from the University of New South Wales, you worked during summers as a statistician for

the Department of Main Roads in Sydney and for the Colonial Sugar Refinery. How did these experiences shape your ambition for a career in statistics, and how did they influence your approach to research?

A Well, not at all. The summer jobs came after I had embarked on statistics and were arranged by my university for statistics students. Over my schooling, any elective chosen was always what was perceived to be the hardest of the available choices. My mathematics cadetship required that I do two honors mathematics programs. Statistics was considered the most difficult of all the (three) mathematics options offered. Pure mathematics was a given, so statistics was by default my other choice. However, while I may have landed in statistics by an unorthodox route, I was very glad to discover this most exciting world. So, back to your question, yes, it was nice to engage in real-life statistical work. Some of those experiences have fueled many illustrations in my teaching over the years.

Q Of the well over 250 publications you have published, which was the most interesting to you? Why?

A Which *one*? There are many interesting and satisfying papers, depending on the context in which the work was undertaken. However, given the

importance of the results, this has to be the two-paper set of HIV/AIDS papers that we discussed earlier. That said, the derivation of the mathematical results was fun to do, too.

Q Your most recent book, *Symbolic Data Analysis: Conceptual Statistics and Data Mining*, takes a very different and interesting approach to data mining. What is the most innovative way you have seen symbolic representation of data implemented into statistical analyses since you and Edwin Diday published this book in 2006?

A First, a brief description of symbolic data is that they are hypercubes in p -dimensional space, instead of points as in classical data. Some data arise naturally, but most will be products of aggregations of larger data sets.

One example is interval-valued observations (e.g., low vs. high stock prices over time, minimum and maximum daily temperatures, etc.). Take two samples of size $n=1$ with interval observations over $[9, 11]$ and $[0, 20]$, respectively. Any analysis based on the midpoints only will give the same results—which are usually incorrect—when clearly the intervals are different and so analyses should give differing results. It is the internal variations that distinguish symbolic analyses from classical ones.

Probably the most exciting analyses so far are the principal component analyses of interval observations with the PCA projections being polytopes; you can read more about this in the *Journal of Computational and Graphical Statistics* (Le-Rademacher and Billard, 2012). Even more interesting is the fact that the output principal components are histogram-valued observations, not points nor intervals. ■

COMING UP
Please return to this column next month, when we will feature an interview with current ASA Executive Director Ron Wasserstein.