Partnerships to Engage and Train the Next Generation: Challenges and Opportunities

Marie Davidian
Department of Statistics
North Carolina State University
Topics

- The Challenge
- The SIBS Experience
- Challenges and Opportunities for the Future
In the news

New York Times, August 6, 2009
“For Today’s Graduate, Just One Word: Statistics”

August 6, 2009
For Today’s Graduate, Just One Word: Statistics
By STEVE LOHR

MOUNTAIN VIEW, Calif. — At Harvard, Carrie Grimes majored in anthropology and archaeology and ventured to places like Honduras, where she studied Mayan settlement patterns by mapping where artifacts were found. But she was drawn to what she calls “all the computer and math stuff” that was part of the job.

"People think of field archaeology as Indiana Jones, but much of what you really do is data analysis,” she said.

Now Ms. Grimes does a different kind of digging. She works at Google, where she uses statistical analysis of mounds of data to come up with ways to improve its search engine.

Ms. Grimes is an Internet-age statistician, one of many who are changing the image of the profession as a place for dourish number nerds. They are finding themselves increasingly in demand — and even cool.

"I keep saying that the sexy job in the next 10 years will be statisticians,” said Hal Varian, chief economist at Google. “And I’m not kidding.”

“I keep saying that the sexy job in the next 10 years will be statisticians,” said Hal Varian, chief economist at Google. “And I’m not kidding.”
“Now they’re all excited to meet me.”
The Age of Big Data
By STEVE LOHR

GOOD with numbers? Fascinated by data? The sound you hear is opportunity knocking.

Mo Zhou was snapped up by I.B.M. last summer, as a freshly minted Yale M.B.A., to join the technology company’s fast-growing ranks of data consultants. They help businesses make sense of an explosion of data — Web traffic and social network comments, as well as software and sensors that monitor shipments, suppliers and customers — to guide decisions, trim costs and lift sales. “I’ve always had a love of numbers,” says Ms. Zhou, whose job as a data analyst suits her skills.

To exploit the data flood, America will need many more like her. A report last year by the McKinsey Global Institute, the research arm of the consulting firm, projected that the United States needs 140,000 to 190,000 more workers with “deep analytical” expertise and 1.5 million more data-literate managers, whether retrained or hired.

The impact of data abundance extends well beyond business. Justin Grimmer, for example, is one of the new breed of political scientists. A 26-year-old assistant professor at Stanford, he combined math with political science in his undergraduate and graduate studies, seeing “an opportunity because the discipline is becoming increasingly data-intensive.” His research involves the computer-automated analysis of blog postings, Congressional speeches and press releases, and news articles, looking for insights into how political ideas spread.

“Statistics are interesting and fun. It’s cool now.”
FOR IMMEDIATE RELEASE
March 29, 2012

Contact: Rick Weiss 202 456-6037 rweiss@ostp.eop.gov
Lisa-Joy Zgorski 703 292-8311 lisajoy@nsf.gov

OBAMA ADMINISTRATION UNVEILS “BIG DATA” INITIATIVE:
ANNOUNCES $200 MILLION IN NEW R&D INVESTMENTS

Aiming to make the most of the fast-growing volume of digital data, the Obama Administration today announced a “Big Data Research and Development Initiative.” By improving our ability to extract knowledge and insights from large and complex collections of digital data, the initiative promises to help solve some the Nation’s most pressing challenges.
Demand

2011 McKinsey Global Institute report:

Big data: The next frontier for innovation, competition, and productivity

“A significant constraint...will be a shortage of...people with deep expertise in statistics and data mining...a talent gap of 140K-190K positions in 2018”

http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation

NC STATE UNIVERSITY
AP Statistics enrollment
(Source: Wikipedia)
So why aren’t more US students going to grad school in Statistics?

- Competition from other STEM disciplines
- Applied Mathematics, Computer Science, Engineering...
- Lack of interest (bad AP/undergrad experience)
- Lack of knowledge of background required
- Lack of knowledge of career opportunities
Missing data

*Science*, February 11, 2011
A long-standing issue...

- The looming shortage of *biostatisticians* in the US was highlighted over a decade ago
- Two NIH workshops in 2001 and 2003 to “examine the need to train more biostatisticians in the US to meet the increasing opportunities” (and needs) in health sciences research
Training of the next generation of biostatisticians:
A call to action in the U.S.

David L. DeMets¹, *, †, Gary Stormo², Michael Boehnke³, Thomas A. Louis⁴,
Jeremy Taylor³ and Dennis Dixon⁵

¹Department of Biostatistics and Medical Informatics, University of Wisconsin, School of Medicine & Public
Health, 600 Highland Avenue, Box 4675, Madison, WI 53792, U.S.A.
²Department of Genetics, Washington University School of Medicine, St. Louis, MO, U.S.A.
³Department of Biostatistics, University of Michigan, School of Public Health, Ann Arbor, MI 48109, U.S.A.
⁴Biostatistics Department, Bloomberg School of Public Health, Johns Hopkins University, Baltimore,
MD 21205, U.S.A.
⁵National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda,
MD 20892, U.S.A.
Main points

- Health sciences research (omics, new technologies, “personalized medicine,”...)
- Cohort of current senior biostatisticians is approaching retirement...
- ...but PhD production has been flat
- Proportion of US PhD students low
- International students have opportunities in home countries (China, India,...)
Main points

- Interdisciplinary team approach is the new model for research; communication/leadership skills essential
- Next generation will need both traditional theory/methods training and facility with emerging methodological areas, massive data, computation, biology,...
Some recommendations

- Stimulate undergraduate awareness/interest
- Establish a general NIH biostatistics training grant program and improve peer review
- Increase industry participation
- Government, private sector must reinvest in training of biostatisticians
The issue persists...

- The gap between supply of and demand is of great concern to biostatistics core directors for the Clinical and Translational Science Award sites (CTSAs).
- Business leaders report being unable to hire PhD statisticians with the requisite skills.
- Big Data will require Big workforce!
Summer Institute for Training in Biostatistics (SIBS)

- NHLBI issued RFA in January 2003
- “Develop, conduct, and evaluate a summer course in the basic principles and methods of biostatistics”
- “...(to attract) new students into the field as a response to the dwindling national pool of trained biostatisticians”
- Target “undergraduate students majoring in mathematics or another quantitative area”
Original SIBS programs, 2004-2009

- Boston University Department of Biostatistics (PI: Lisa Sullivan)
- University of Wisconsin – Madison Department of Biostatistics and Medical Informatics (PI: Dave DeMets)
- North Carolina State University Department of Statistics – Duke Clinical Research Institute (Co-PIs: Dennis Boos, Marie Davidian)
Features

- 6 weeks, tuition for 4-6 hours course credit
- Room, board, transportation
- **Not** a traditional statistics course
- Lectures by biostatisticians, clinicians, epidemiologists, statistical geneticists,…
- Field trips (e.g., Framingham Heart Study site, SAS Institute, GlaxoSmithKline, DCRI)
- Statistical software (SAS/R/etc), data analysis projects using data from real studies
Goal

Through these and other elements, pique the interest of participants and inspire them to pursue graduate training in biostatistics/statistics.
SIBS 2004-2009

- Joint advertising and recruitment
- Numbers of applications (> 100/site)
- Word of mouth, inclusion in lists of internships and summer programs, *Amstat News* articles, sessions at JSM,…
- 409 participants 2004-2009, > 65% went on to graduate school in biostatistics/statistics!
SIBS II, 2010-2012

- Success of “SIBS I” led to “SIBS II”
- Original 3 programs + 5 more
- Emory University (PI: Lance Waller)
- University of Iowa (PI: Kathryn Chaloner)
- University of Pittsburgh (PI: Roslyn Stone)
- University of South Florida (PI: Yiliang Zhu)
- Washington University in St. Louis (PI: G. Charles Gu)
SIBS II, 2010-2012

- 6-7 weeks, 15-25 participants, course credit
- Continued joint advertising and recruitment
- 607 unique applicants in 2010, 163 participants
- 518 unique applicants in 2011, 166 participants
- Of those eligible, ~60% went on to graduate school in biostatistics/statistics
- RFA for 2013-2015, up to 7 programs funded

http://www.nhlbi.nih.gov/funding/training/redbook/sibsweb.htm
The Summer Institutes for Training in Biostatistics are sponsored by the National Heart, Lung, and Blood Institute and the National Center for Research Resources.

A summer program that explores how to apply quantitative methods to investigate important health issues.

The opportunity to learn the principles of applied biostatistics from recognized experts in the field and meet practicing biostatisticians, epidemiologists, and statistical geneticists.

The chance to gain real-world experience working with data from internationally recognized studies funded by the NHLBI.

The chance to learn about opportunities for graduate study, additional training, and career opportunities.

Eligibility:
- Undergraduates majoring in mathematics, statistics, biology, or other science who have an interest in quantitative methods.
- Those with a baccalaureate degree are eligible to apply but priority is given to undergraduates at the time of application.
- Applicants must be American citizens or permanent residents of the United States.

Educational and Financial Benefits:
- No fees or tuition costs (housing, meals, and many extracurricular activities covered).
- Earn college credits.
- Hands-on training with top researchers.
- Access to top Universities’ facilities.

Additional Program Information:
Eight SIBS programs will be held in the summer of 2012. Links to information on the programs including program dates, application procedures, and deadlines, will be available by Fall 2011 at the host department websites.
Participant comments

“I wasn’t sure of my future plans before this program, and now I definitely want to go to grad school in biostatistics”

“Before coming to this program, I had no idea what biostatistics was. Now, I’m definitely thinking of going to grad school in biostatistics”

“After participating in SIBS, I am undoubtedly going to pursue a PhD in biostatistics”

“I would highly recommend this program to anyone not sure of what to do with a math degree”

“Overall, this program was fabulous and I’ll recommend it to other students in the future!”

“I’m leaving the program with a whole new appreciation for statistics...I never realized there was so much out there!”
Participant comments

“I attribute the SIBS program to finalizing my decision in pursuing the field as a career”

“Without the introduction ... that SIBS has provided me, I certainly would not be on the path I am now”

“The SIBS program (was) instrumental not only to the development of my interest in biostatistics but also to my preparation for graduate training in the field.”
The Future

- SIBS has been a tremendous success and demonstrates that such programs work!
- Three more years (2013-2015) but then what happens?
- Additional such programs to inspire interest in statistics more generally are warranted now
Challenges

- SIBS is outstanding at inspiring interest – but how do we sustain it?
- Few follow-up opportunities exist for SIBS alumni who have a few more summers before grad school to cement interest
- Summer programs in competing fields siphon them away.....
Challenges

* We must adapt our curricula – more and more material, Big Data and computing, communication/leadership skills,…

* Statistics graduate students need more opportunities for training while they are in school to prepare them to be effective members of interdisciplinary research teams
Challenges

* In biostatistics, **interdisciplinary training opportunities** come mainly through NIH T32 training grants

* But there has been no ↑ in the number of NIH biostatistics training grants over the last decade (~ 40)

* There is no NIH-wide strategy for training of biostatisticians, despite our broad relevance
How?

- We must take the lead as a profession
- We cannot assume that NIH, NSF, and other agencies can or will fund such programs alone
- All of the academic, government, and industry sectors are beneficiaries and hence are stakeholders
- As are professional societies and agencies
- Professional societies can partner, advocate, promote, facilitate
Opportunities

- Lobby NIH for agency-wide strategy for biostatistics training support
- Form consortia (academic-industry-government-professional society) stakeholders to promote and spearhead
- Adapt successful models where possible; e.g., NCSU Graduate Industrial Traineeships

  www.connectcanadainternships.ca,
  www.masslifesciences.com/grants/challenge.html
WH Big Data initiative

- “Expand the workforce need to develop and use Big Data technologies.”
- “NSF is .... Encouraging research universities to develop interdisciplinary graduate programs to prepare the next generation of data scientists...”

http://www.whitehouse.gov/administration/eop/ostp
Call to action

2013 is the International Year of Statistics – and Big Data are upon us. There’s no better time to come together and conceive of and implement bold new strategies to engage, inspire, and train the next generation!