

# Nudges, grit and grinding: Navigating a career as an academic biostatistician while working to build a diverse profession

**Scarlett L. Bellamy, ScD**

Professor, Department of Epidemiology and Biostatistics

Director of Graduate Studies in Biostatistics

September 22, 2018



# Preliminaries and disclosures

Perspective based on my experience as:

- Biostatistician
  - ▶ 1995; BA Mathematics – Hampton University
  - ▶ 1997; ScM Biostatistics – Harvard School of Public Health
  - ▶ 2001; ScD Biostatistics – Harvard School of Public Health
- Academician – (1) large, urban medical center; and (2) medium, school of public health
- Generally operate with a Plan B in mind (generally don't think of myself as a typical Type A personality)
- Personal life just as important as my professional life



# What is Biostatistics?

- Science of the design and analysis of biomedical research studies
- The science of statistics applied to the analysis of biological or medical data
- The application of statistics to biology, medicine, nursing, and other health-related professions
- The science of obtaining, analyzing, and interpreting data using statistical theory and methods to address problems in the biological and health sciences.

# Where do we typically work?

How are we typically trained?

- Undergraduate (Bachelor's degree)
- Graduate degrees – both
  - ▶ Master's degrees (MPH, MS, ScM, etc)
  - ▶ Doctoral degrees (PhD, ScD, etc)

Where we typically work?

- Academia – undergraduate vs. graduate; biostatistics vs. statistics vs. math vs. ??; research vs. teaching; etc.
- Government – federal, state, local
- Industry
  - ▶ Pharmaceutical companies: (e.g., GSK, Merck, Pfizer, Johnson & Johnson)
  - ▶ Non-governmental/contract research organizations (e.g., RAND, Quintiles, Rho)
  - ▶ Software companies (e.g., SAS, STATA)
  - ▶ Big data technology companies (e.g., Google, Amazon, Lyft, Uber)
- Entrepreneur/Consulting

# What Do Biostatisticians do?

In a nutshell...

- Often work collaboratively to design scientific studies and analyze data generated as part of 'biomedical' research (e.g., public health, medicine, pharmaceuticals, medical informatics)
  - ▶ e.g clinical trial, observational study, large database study (informatics?), basic science experiments
- Develop better ways to design studies and analyze data
- Teach
- Train and mentor students, collaborators, etc
- Communicate (verbally and in writing) with collaborators and other scientists in the field, quantitative aspects of a project.

# My Journey: Nudges, Grit and Grinding I



- Born to teenage parents in rural NC
- Spent time living between grandparents (maternal grandmother + paternal grandmother and grandfather)
- 7-8yo grandma diagnosed with breast cancer; passed away when I was 10; promised her I would become a doctor
- Moved to MD w. dad + step-mom

# My Journey: Nudges, Grit and Grinding II



- History of mathematicians/math teachers in the family (dad); great-great grandmother, grandfather's brother, aunt
- Aunt taught at NCCU for many years; spent summers with her and sat in her classes; Introduced me to a physician who encouraged me to study something that I liked; learn about being a doctor in med school
- HS loved math, but was not the best student; worked really hard
- Math major at Hampton (Student Enhancement in Mathematics and Science (SEMS) scholar)
- Biostatistics major at Harvard

# My Journey: Nudges, Grit and Grinding III



Graduate school...

- Choosing a graduate program
- Challenging
- General Format: Coursework (2 yrs, 3 per term) + Exam + Dissertation writing + Exam
- Me:
  - ▶ Coursework (average)
  - ▶ Qualifying Exam (2x, prize for best score)
  - ▶ Dissertation (chose an advisor that I liked)
  - ▶ Oral Exam – conditional pass (mandatory public speaking class)
  - ▶ Finished in 2001!!



# My Journey: Nudges, Grit and Grinding IV



- 17 years in academia; professional currency some mix of: research/funding, teaching and publishing
  - ▶ Penn (2001-2016) – Started as Assistant Professor and promoted to Professor
    - ★ From an applications perspective.. often sense of explaining why the public health problems that you are working on are important
    - ★ Can be viewed as a service profession(al), rather than a true colleague
    - ★ Perceived MD/PhD hierarchy
    - ★ Clinical/medical culture differences
  - ▶ Drexel (2016 - )
    - ★ Develop new PhD program in biostatistics (leadership)
    - ★ Return to public health

- Now... I have lots more flexibility and autonomy regarding how I spend my time
  - ▶ Professional Leadership
    - ★ ENAR, ASA
    - ★ Graduate Program at Drexel
    - ★ Outreach



- Professional organizations
  - ▶ Help to stay sharp!
  - ▶ Stay connected
  - ▶ Growth opportunities
  - ▶ Leadership opportunities
  - ▶ Diversity outreach



# Who \*IS\* the sexiest?

Who \*IS\* the sexiest?

Statistics:

- NYTimes 2009: Title: "**For Today's Graduate, Just One Word: Statistics**"; "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."
- WIRED 2009: 'What's ubiquitous and cheap?' Varian asks. 'Data.' And what is scarce? The analytic ability to utilize that data.

Data Science:

- Harvard Business Review 2012: sexiest job of the 21st Century
- Glassdoor 2016: the best job of the year

“*I keep saying the sexy job in the next ten years will be statisticians. People think I'm joking, but who would've guessed that computer engineers would've been the sexy job of the 1990s?*

Hal Varian, *The McKinsey Quarterly*, January 2009



DATA

## Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

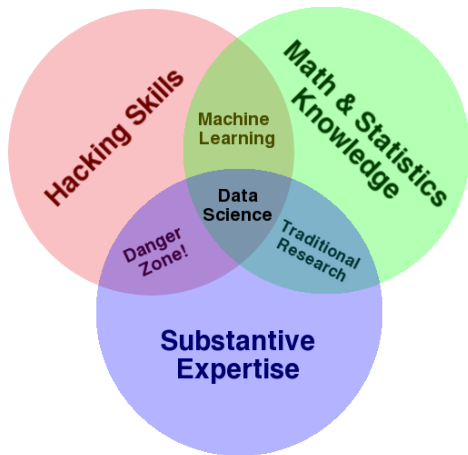
# Is Data Science as a 'disruptive technology' ?

- "The statistics profession is caught at a confusing moment: the activities which preoccupied it over centuries are now in the limelight, but those activities are claimed to be bright shiny new, and carried out by (although not actually invented by) upstarts and strangers." 50 Years of Data Science, Donoho 2015
- Joint Statistical Meetings (2013) – Nate Silver "I think data-scientist is a sexed up term for a statistician....Statistics is a branch of science. Data scientist is slightly redundant in some way and people shouldn't berate the term statistician."
- The discussion (in print and at large professional gatherings)
  - ▶ "Aren't we Data Science?" – Column of ASA President Marie Davidian in AmStat News, July, 2013
  - ▶ "Let us own Data Science" – IMS Presidential address of Bin Yu, reprinted in IMS bulletin October 2014
  - ▶ "A grand debate: is data science just a rebranding of statistics?" Martin Goodson, co-organizer of the Royal Statistical Society meeting May 11, 2015 on the relation of Statistics and Data Science, in internet postings promoting that event.
  - ▶ "But I am a data scientist too, aren't I?" – IBS President-Elect Louise Ryan, 2017 ENAR Presidential Invited Address

# Why do I care?

- Unique opportunity to 'modernize' (bio)statistics
  - ▶ ReThink how we train future graduate students
  - ▶ Blur the lines between statistics and data science
  - ▶ Focus on the 'science' of data science – using statistical science to answer important questions
  - ▶ Potential to entice a broader pool of trainees to the field

Figure: The Data Science Venn Diagram



- **Hacking:** Being able to manipulate text files at the command line in order to construct analyzable databases; not necessarily computer scientists
- **Math/Stat knowledge:** Not necessarily PhD in statistics
- **Substantive Expertise:** Where the 'science' comes in; discovery and building knowledge, which requires some motivating questions about the world and hypotheses that can be brought to data and tested with statistical methods

# Prototypical biostatistics graduate training sequence

- Theory Sequence: Probability, Inference I, Inference II
- Methods Sequence: Methods I and II, Linear Models, Categorical Data Analysis, Survival Analysis, Longitudinal Data Analysis, Study Design, Statistical Computing
- Practicum: Consulting (e.g., class or students participate in lab), HIPPA, Bioethics
- Applications concentration: Environmental Health, Cancer Epidemiology, etc (Minor Courses)

Relevant questions:

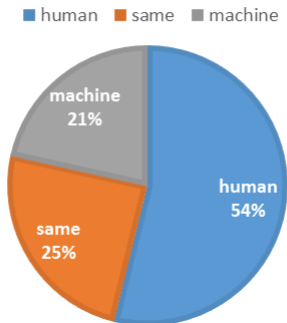
- Is there room in current PhD/MS training sequence to expose students to data science?
- To broaden relevance, statistics as a profession needs to broaden its scope. Invite more people in.



# Case study – Humans vs. Machine: Machine Learning Approach to ENAR Program Creation

- Traditionally, ENAR receives about 400 contributed abstracts which are then collated "by hand" into cohesive sessions with 6-7 presentations.
  - ▶ The 'humans': 3 'Chairs' + 15 reps from various ASA Sections + 4 At-Large members
- This year the program chairs conducted the process in the traditional way and also explored a more modern data science approach
  - ▶ Network analysis: identified large subgroups based on key words
  - ▶ Natural language processing: extracted key information from the title and abstract
  - ▶ Cluster analysis: categorized abstracts into groups of similar topics
- We then sent out surveys to a random selection of authors and asked them to rate sessions organized by hand (human) versus using machine learning (machine).

## HUMAN VS. MACHINES SURVEY RESULTS



- Glass half empty interpretation: Humans 'won' (54%)
- Glass half full interpretation: Combine those who preferred 'machine' made program with those who didn't have a strong preference for one over the other ( $21\% + 25\% = 46\%$ ) – 'machine' nearly there!
  - ▶ We also learned a few things that we will tweak with 2018 program (e.g., offer more key word choices, etc.)
  - ▶ Hard code our 'human' knowledge (e.g., distinguish 'time to event' from 'time series', 'failure time', etc.)

# Clarifications

- Do I think *\*all\** statisticians should become data scientists? No.
- Do I think *\*all\** universities offering (bio)statistics degrees should add some element(s) of data science? Yes.
- Do I think *\*you\** should pursue graduate training in a mix of (bio)statistics + data science? Yes.

# What about that title?

- Nudges:
  - ▶ Responding to (gentle) pushes to go out of my comfort zone
- Grit:
  - ▶ Push through challenges
  - ▶ Call on inner/outer resilience to get you through
  - ▶ Know that difficulties will always come and go – 'adulting'
- Grinding:
  - ▶ Getting stuff done

# Undergraduate grinding

- GRADES, GRADES, GRADES
- Try 'low risk' exposures to graduate training disciplines of interest – better to figure out what you do (and do not!) like before enrolling (e.g., courses, summer programs, workshops, etc.)
- Prepare the strongest application possible
- On campus visits
- Talk to people who work in areas that interest you



# How do I learn more about Biostatistics?

- \*\* ENAR – Fostering Diversity in Biostatistics Workshop
- \*\* ASA – Diversity Workshop, Mentoring Program
  - ▶ Also visit [thisisstatistics.org](http://thisisstatistics.org)
- SIBS – Summer Institute for Training in Biostatistics

2015, the NHLBI supports 8 SIBS programs, with co-sponsorship from the National Center for Advancing Translational Sciences (NCATS): Boston University, Columbia University, Emory University, North Carolina State University, University of Iowa, University of Minnesota, University of Pittsburgh at Pittsburgh, University of Wisconsin at Madison
- Biostatistics Epidemiology Summer Training Diversity Program (BEST) Diversity Program

(<http://www.mailman.columbia.edu/academic-departments/biostatistics/best-diversity-program>)
- Summer Program in Biostatistics & Computational Biology

(<http://www.hsph.harvard.edu/biostatistics/summer-program/>)