Biostatistician’s Perspective:
Roles in Observational Studies,
Electronic Health Records & Big Data

Brad H. Pollock, MPH, PhD

Professor and Chairman,
Dept. of Epidemiology & Biostatistics,
School of Medicine,
University of Texas Health Science Center at
San Antonio
Disclosures

- No financial conflicts of interest
- *Director*, Biostatistics, Epidemiology, Research Design (BERD) Core (UL1 RR025767)
- *Director*, Biomedical Informatics Core, Clinical Translational Science Award (UL1 RR025767)
- *Director*, Biostatistics Shared Resource, Cancer Therapy & Research Center, University of Texas Health Science Center at San Antonio (P30 CA054174)
- *Principal Investigator*, Children’s Oncology Group (COG) Community Clinical Oncology Program (CCOP) Research Base (U10 CA095861)
Disclosures

• I am really an epidemiologist raised by a pack of biostatisticians.
  – Therefore, I am bipolar, so take everything I say with a grain of salt.
Three Domains

- Observational Studies
- Electronic Health Records
- Big Data

- Not orthogonal
- Cut across disciplines
- Cut across units
Observational Research

• Covers research that isn’t interventional

• Standard biostatistics stuff:
  – Hypothesis formation
  – Sample size/feasibility
  – Monitoring (quality)
  – Analysis
  – Reporting
Observational Research (continued)

- Observational designs are deployed when randomized clinical trial (RCT) designs are not possible or not ethical
  - More potential for bias to creep in
  - More complex designs and analysis approaches are needed to minimize bias; i.e., more biostatistical involvement
Observational Research: Team

- Epidemiologists
- Biostatisticians
- Other public health experts
- Clinical and translational scientists
Observational Research: Considerations

+ Omic-specific groups too small to run RCTs
+ “Free” plentiful data from sources like EHRs
+ Pivotal for the design of future RCTs
  - Non-existent or poor quality data
  - Less weight of evidence

• Increasing reliance on observational data for precision medicine

• Need ontologies for observational studies beyond ClinicalTrials.gov
  – Meta-study data repositories
EHR Research

• Building and beginning to exploit EHR repositories

• Research employs mostly observational designs but increasing use of cluster randomized designs (pragmatic trials), and potential for individual RCT designs
  – Being pushed rapidly by PCORnet and the PCORI-CDRNs
EHR Research: Disciplines

- Medical informatics
- Health IT
- Computer science
- Clinical / applied epidemiology
- Biostatistics
- Health services research
- Medical domain expertise
- Bioethics
EHR Research: Trends

• Informatics focus has been on building repository infrastructure, just starting to shift to analytics and development of decision support tools

• Toward a Common Data Model?
  – i2b2/SHRINE vs. Mini-Sentinel Distributed Database

• Realization of the IOM’s Learning Healthcare System
Big Data Research: Subdomains

- -Omics research
- Administrative database research
- Electronic Health Record (EHR) research
Big Data: Biostatistical Considerations

- Often requires complex modeling approaches and validation methods
- Missing data challenges
- Computational platform constraints
- Sample size issues for -omics studies
  - Many variables, few cases
  - Linkage with whole human data
Big Data: Concerns

Paradigm shift with the ever-increasing generation and availability of data, going:

From: **Hypotheses** in search of **Data**

To: **Data** in search of **Hypotheses**

- Missing or inappropriately structured data elements; e.g., the EHR may miss:
  - Dose schedule, dose intensity, AUC, dose modifications, patient PK/PD characteristics
Big Data: Concerns (continued)

- Systematic biases in big data sources only get amplified, not attenuated.

Need extremely tight integration of clinical and translational researchers with bioinformaticians, IT experts, biostatisticians and epidemiologists.
# Transdisciplinary Approach to Big Data

<table>
<thead>
<tr>
<th>Type</th>
<th>Disciplines</th>
<th>Biostatistical Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>-Omics</td>
<td>Bioinformatics</td>
<td>Experimental design, modeling (time series and nested effects); Linkage for clinical annotation</td>
</tr>
<tr>
<td>Administrative</td>
<td>Health Services Research, Applied Epidemiology, Public Health, Medical Informatics</td>
<td>Debiasing observational data; Missing data; Propensity scores; Secular changes in coding for longitudinal datasets</td>
</tr>
<tr>
<td>EHR</td>
<td>Medical Informatics, Health IT, Clinical domain experts, Finance, Psychometrics</td>
<td>Cohort assembly; Deciphering EHR entries (defining incident cases); Missing data; PROs</td>
</tr>
</tbody>
</table>
Examples Illustrating the Need for a Transdisciplinary Approaches

- Oncology of Clinical Research (OCRe)
- IOM’s *Learning Healthcare System*
Ontology of Clinical Research (OCRe)

- Goal is to develop common taxonomy and vocabularies to standardize the storage of study information (meta-study data)

- Elements:
  - Study design type
  - Interventions/exposures
  - Participants
  - Outcomes
  - Statistical analyses
OCRe Study Design Topology

Does the study use primarily qualitative or quantitative methods?
- qualitative
  - Interviews, focus groups, ethnographic studies, etc.
  - Qualitative Studies
- quantitative

Does the investigator assign one or more interventions?
- Yes
- No

Does the investigator have a choice of interventions to assign participants to?
- Yes
  - Is the main comparison within or across participants?
    - Within
      - Does investigator assign interventions to and analyze data only within a single study participant?
        - Yes
          - N-of-1 Crossover
        - No
          - Crossover
    - Across
      - Parallel Group
- No
  - Single Group

Main Control Group Defined By
- Case (outcome)
  - status
  - Exposure (predictor)
  - status
- Are case and control in the same person?
  - Yes
  - No
- Are outcomes measured at same time as predictors?
  - Yes
  - No
- alter measurement of predictors

Observational Studies
  - Case Crossover
  - Case Control
  - Cross-sectional
  - Cohort
The HSDB Team

Jim Brinkley
U Wash

Simona Carini
UCSF

Todd Detwiler
U Wash

Harold Lehmann
Hopkins

Brad Pollock
UTHSC S Ant

Shamim Mollah
Rockefeller

Ida Sim
UCSF

Swati Chakraborty
Duke

Samson Tu
Stanford

Knut Wittkowski
Rockefeller

a CTSA Informatics Consortium project
Learning Healthcare System

The learning healthcare system refers to the cycle of turning health care data into knowledge, translating that knowledge into practice, and creating new data by means of advanced information technology.
Learning Health Care System Process

Clinical trials, comparative effectiveness research, molecular and biologic data

Information-rich, patient-focused data

Evaluation of outcomes

Data aggregation, evidence generation

Transformation of subsequent care delivery
Gathering Data to Make Decisions

Study Designs
- Observations studies
- Quasi-experimental designs
- Cluster randomized clinical trials
- Individual randomized clinical trials

End Products
- Clinical practice guidelines
- Roadmaps / Critical pathways
- Decision support tools
- Automated EHR recommendations
Improved health & health care through research, innovation & dissemination

Primary areas of research focus

- Biostatistics
- Cancer Prevention & Control
- Chronic Illness Care
- Immunization
- Preventive Care & Health Promotion
- Mental Health and Behavioral Medicine
- Health Systems Organization & Finance
- Women’s Health
Summary

- Hard to imagine having research success without close interactions between biostatistics and other disciplines
- Trend is toward more transdisciplinary research
- Developing *Learning Healthcare Systems* will put this to the penultimate test
Transdisciplinary Research
Secrets of Success

• Ability to be able to step outside your skin

• Genuine desire to understand the fundamental principles of other disciplines

• Flexible thinking to reimagine dogmatic or unquestioned concepts within your own discipline

• Desire to have the combined effort be much greater than the sum of the parts
Transdisciplinary Research

Secrets of Success (continued)

• Proximity is 9/10s of the law; i.e., it sure helps to have diverse discipline experts in close proximity
  – Maybe even develop transdisciplinary units/programs that transcend traditional:
    • Divisions
    • Departments
    • Centers / Institutes
    • Universities
    • Hospitals
Thank you