Data, Data, Everywhere!
or
How Can I Get Biostatistical Support for my Research?

Lance A. Waller (lwaller@emory.edu)
Department of Biostatistics and Bioinformatics
Rollins School of Public Health
Emory University
Motivation

- “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.” (NY Times, 8/5/09)
Where I’m coming from…

EIGHTY-TWO PERCENT OF THE POPULATION DOESN’T GET ENOUGH FIBRE IN ITS DIET!

TWO OUT OF THREE KIDS IN AMERICA CAN’T SPELL “CHOLESTEROL”!

NINETY-NINE PERCENT OF THOSE INTERVIEWED SAID THEY HATED CARROTS!

STATISTICS ARE TOTALLY WRONG FIFTY-SIX PER CENT OF THE TIME!

THE STAT FAMILY

- 157 of these compared changes observed in two groups, 78 used the correct procedure, 79 used the incorrect procedure.

“If all of your friends jumped off of a cliff, I suppose you would too.”

Data safety monitoring board interim review revealed alarming increase in stroke in the stent arm of study.

Trial stopped.

Positive press: NY Times editorial:

“It clearly shows the value of conducting rigorous controlled studies with enough patients to provide meaningful results. This is just the kind of ‘comparative effectiveness’ research that the national health care reforms seek to promote.”
Goal of Statistical Analysis

- Every study has a story to tell.
- Every data set has a story to tell.
- Statistical analyses provide quantitative answers to specific questions.
- Every statistical method answers a very specific question.
The whirling vortex of analysis

The question you want to answer

The data you need to answer that question

The question you can answer with those data

The data you can get

Original source: John Richardson, EPA Region IV
“Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise.” Tukey (1962, *Annals of Mathematical Statistics*).

What question you can answer depends on the data you have AND the method you use!
Questions and Answers

Questions answered by many common statistical methods:

- **Z-test:** Are *means* the same?
- **ANOVA:** How do *means* change between treatment groups (adjusting for *variance*)?
- **Chi-square test:** Are *proportions* the same?
- **Regression:** How are covariates *associated* with continuous outcomes?
- **Logistic Regression:** How are covariates *associated* with *proportions* of Yes/No outcomes?
- **Poisson Regression:** How are covariates *associated* with *counts* of outcomes?
Statistics is not a static toolbox.

Every data set has some peculiarity that doesn’t quite match the “perfect setting” for the method.

Statisticians are trained to customize analytic methods to address the problems in your data.

A good design solves many, many analysis problems, but the converse is not always true.

What question you can answer depends on the data you have AND the method you use!
The whirling vortex of analysis

The question you want to answer

The data you need to answer that question

The data you can get

The question you can answer with those data

Original source: John Richardson, EPA Region IV
Types of design

- **Clinical Trial:** An experiment with human subjects.
  - Tight regulations
  - Randomization
  - Careful recruitment
  - Design is *critical*

- **Observational Study:** Observe outcomes and variables
  - Many epidemiologic studies
  - Observe outcome and exposure
  - Less control, no randomization
  - Concern about confounding
<table>
<thead>
<tr>
<th></th>
<th>Exposed</th>
<th>Not Exposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Disease</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Types of Studies

- **Prospective**
  - Recruit Exposed/Not Exposed
  - Observe Disease/No Disease
  - Often more expensive, especially for rare disease.

- **Retrospective**
  - Recruit Disease/No Disease
  - Observe Exposed/Not Exposed
  - Often more efficient
  - Odd ratio estimates *relative risk* for rare disease.
The whirling vortex of analysis

The question you want to answer

The data you need to answer that question

The question you can answer with those data

The data you can get

Original source: John Richardson, EPA Region IV
“The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data.” Tukey (*American Statistician*)

What question you can answer depends on the data you have AND the method you use!
George Box

- “…all models are wrong. The practical question is how wrong do they have to be to not be useful.”
  Box and Draper (1987, p. 74).

- What question you can answer depends on the data you have AND the method you use!
When to Involve a Biostatistician

• There are 2 main stages we (the biostatisticians) “like” to be involved:
  • 3-6 months out during the initial development of your research hypotheses and study design.
  • During/after the data have been collected to work on analysis for abstract and manuscript submissions

3-6 MONTHS PRIOR

Development of research hypotheses and design of study. Sample size calculations, database design, etc.

Following Data Collection

Analysis of data for abstracts and manuscripts. About 1-2 months before target submission date.
3-6 Months Prior

- Narrow down ideas and develop concrete hypotheses
- Determine your primary hypothesis (and a few secondary)
- Discuss the outcome variable(s) are associated with your primary hypothesis
- Determine sample size for your study
- Discuss data collection and management
- Help with IRB submission
When to Involve a Biostatistician

Following Data Collection (1-2 months prior to target date)

- All data collection should be completed or near completion
  - Final analysis will not be performed until data collection is completed.
- List of variables and questions you are interested in looking at
  - Any subgroup analysis?
  - Important interactions?
  - Variable Code Sheet
- Determine what kinds of tables and figures you want for your poster or manuscript.
Options for support

- Biostatistician supported in project personnel
- Departmental support for Biostatistics
- Project cores
- Biostatistical Consulting Center
Who do we have?

- 18 tenured or tenure-track faculty
  - Combination of individual and collaborative research.
  - PIs on R01s, R03s, K, T32s, T15
- 7 research track faculty
  - Primarily in cores or funded on collaborative projects.
- 8 MS-level associate faculty
  - Combination of teaching and collaboration.
- 9 research staff
  - MS-level biostatisticians, data analysts, database developers
  - Project coordinators
- All supported on combinations of collaborative projects and collaborative arrangements.
Key Personnel or Biostatistician?

Advantages
  • Dedicated percent effort
  • Involvement from design to analysis

How to find help?
  • Check with project PI.
  • Ask about biostatistical support.
Departmental Biostatistician

- Effort dedicated to Department/Unit (joint hire)
- Examples: Radiology, CHOA, Surgery, Pediatrics
- Advantages
  - Block of time available for departmental investigators
  - Consistent support across similar projects

- How to find help?
  - Check with Department Chair.
  - Ask about biostatistical support from dedicated biostatistician (CHOA/Pediatrics: Leong and McCracken).
Program project cores

- **Examples:**
  - ACTSI BERD, WCI Shared Resource, CFAR

- **Advantage:**
  - Consistent support across projects

- **Challenge:**
  - Support provides infrastructure, not “free support for everything”.

- **How to find help?**
  - Typically request support through Center website
  - Ask about biostatistical support from Core Director (ACTSI BERD: Robert Lyles, rlyles@emory.edu)
Biostatistical Consulting Center (BCC)

- For dedicated, short- to medium-term support for clearly delineated tasks.
- Task-based fee structure.
  - Hourly equivalent: $80-110/hour for projected effort.
  - Rates based on experience, revised each year.
- Clear statement of work, clear deliverables, assignment of personnel.
- Dedicated support from core staff but access to all faculty/staff in Department.

- How to find help?
  - Check with BCC Director (Waller)
  - Outline objectives, draft statement of work, identify deliverables and timeline.
General framework

- Project support for statistician? Start there.
- Departmental support for statistician? Start there. (Check with chair).
- ACTSI investigator without funds? Contact BERD (Bob Lyles).
  - Primarily for short-term or early start-up projects.
- Bigger project needing dedicated support for a clear task without a grant? Contact BCC (Lance Waller).
  - Primarily for short-term projects with clear deliverables.
What is the ACTSI?

- Atlanta Clinical & Translational Science Institute (ACTSI) is one of 62 current NIH (NCATS)-funded Clinical and Translational Sciences Awards (CTSAs)

- ACTSI is a large-scale Atlanta-based consortium (PI: David S. Stephens, M.D.) with 3 primary partner institutions: Emory, Morehouse School of Medicine, and Georgia Tech
ACTSI’s Mission:

- Through focused **education and training**, innovative **support of discovery**, and **ethical community engagement**, the **collaborative partners** of the ACTSI rapidly and efficiently translate scientific discoveries to impact all populations of the Atlanta community and beyond.

(from ACTSI website:  [www.actsi.org](http://www.actsi.org))

- Biostatistics, Epidemiology, & Research Design (BERD) is one of many ACTSI programs
General framework

Seeking ACTSI-BERD Support

- Emory investigators may visit [www.actsi.org](http://www.actsi.org) and select ‘Submit a Request’ to enter info about affiliation and project.

- Can request a BERD consultation, or a ‘Studio’ (larger-scale meeting with reps from CRN, BERD, and BIP programs).

- **Note:** Most protocols developed with ACTSI support will go through CRN Scientific Advisory Committee (SAC) review (includes biostatistical review).
General framework

- BERD’s primary mission is to improve proposals at developmental stages (e.g., statistical analysis and data management plans, power/sample size considerations)

- Another major goal: Connect researchers with biostatistical faculty expertise for longer-term collaborations

- Example “success story”: Long-term connection between Dr. Jonathan Glass (Emory ALS Clinic) and Biostatistics and Bioinformatics faculty developed from an initial ‘Studio’ meeting

  ⇒ Ongoing doctoral dissertation research w/ Dr. Glass as a committee member
Questions?