Best Practices for Partnering with a Biostatistician in your Research

Scott Gillespie, MS, MSPH
Associate Director and Lead Biostatistician
Emory Pediatric Biostatistics Core

K-Club
September 12th, 2022
Accompanying Panelists

Traci Leong, PhD
Assistant Professor
Rollins School of Public Health
Emory University

C. Christina Mehta, PhD, MSPH
Associate Professor
Director, Infectious Diseases Biostatistics
Emory University School of Medicine
Agenda

• What is a biostatistician? (Brief Introduction)

• Tips for effectively partnering with a biostatistician in your research
  – Panel discussions and audience participation (4-5 minutes per tip)
What is a biostatistician?

• Biostatisticians develop and apply statistical methods to biomedical and health data

• Partner with other health field professionals to advance scientific discovery:
  – Design studies
  – Make data collection decisions
  – Analyze and interpret data
  – Write articles for dissemination
Biostatistician Make-up

• An experienced biostatistician is competent in four areas:

1. Technical and analytical
   • Familiar with modern statistical methods and software coding

2. Broad subject knowledge
   • Working knowledge of the biomedical content

3. Communication
   • Ability to understand and be understood

4. Problem-solving
   • Synthesize critical study components to answer research questions
Tip #1: Involve a biostatistician early!

To call in the statistician after the experiment is done may be no more than asking them to perform a post-mortem examination: they may be able to say what the experiment died of.

-RA Fisher
When to contact your biostatistician

A. Study conceptualization 😊

B. Study is conceptualized, but needs polishing 😴

C. Data collection phase, before study starts 😞

D. Data collected and needs analysis help 😞

E. Performed own analysis and needs checking 😠

F. Manuscript submitted, answering reviewer criticisms 😞
Advantages

• Help you think through technical objectives of research study

• Reconstruct research questions into research hypotheses, and ultimately, into statistical hypotheses to inform analysis

• Help identify variable types and roles

• Ensure available data and planned analysis are appropriate for answering the research question(s)
Tip #1: Involve a biostatistician early

Panel and Audience Discussions
Tip #2: Biostatisticians know more than you think

• Three broad areas where biostatisticians add value to the research enterprise:
  1. Data analysis
     • Interpretation; technical write-ups
  2. Study design
     • Hypothesis refinement; conceptualization of complex relationships between variables
  3. Education/communication
     • Presentations; journal clubs; one-on-one consultations

• Good data analysis should follow a SAP that compliments a strong study design and is understandable to PI(s)
But also...what information we may not know

• A statistician may lack specific subject knowledge for your study
  – You should not assume we are familiar with all acronyms, jargon, or instruments you propose to use (communication is key here)

• We are generally not database experts and may not be familiar with your data collection software
  – At Emory, we do see and can advise on, Redcap, Excel, some SQL databases

• We may not have experience with a niche method or analysis plan that is common for your field
  – Early contact, providing relevant papers, and table/figure mock-ups are helpful to educate your statistician on your data
Tip #2: Biostatisticians know more than you think

Panel and Audience Discussions
Tip #3: Biostatisticians are part of your expert team

• A complex study should bring together people of various expertise:
  – Scientists/clinicians
  – Biostatisticians
  – Data managers
  – Qualitative experts
  – Informaticians

• These individuals should be viewed as co-collaborators and not just as a service
Inappropriate uses of a biostatistician

• Messages that make a biostatistician quiver:

“The grant is done. I just need a couple lines for the stats.”

“Can you quickly look over my manuscript before I submit? I’m happy to make you a co-author.”

“Can you help me with a quick power calculation?”
“Can you help me with a quick power calculation?”

• Statistical power calculations are often recommended when submitting a grant
  – Sample size needed to show a statistically significant effect (i.e., $p<0.05$, if one exists in the population of study)

• A power calculation should:
  – Ideally be based on data from the literature or pilot information
  – Flow naturally (and holistically) with the study aims and SAP

• Requesting a power calculation without knowledge of aims/SAP is akin to asking a doctor to diagnose without knowledge of symptoms or prior medical history
General collaboration flow for a biostatistician

1. Scientific Question
   - Communication to Statistician

2. Statistical Plan
   - Communication to Statistician
   - Statistical Analysis

3. Write-up/Results
   - Communication to Scientist

4. Scientific Interpretation
   - Statistician's Domain
   - Scientist's Domain

Tip #3: Biostatisticians are part of your expert team

Panel and Audience Discussions
Tip #4: Plan appropriately for collaboration

- General timing recommendations:
  - 20 business days for initial report
  - 30 business days for intramural grant preparation
  - 40 business days for single project extramural grant preparation (e.g., R03, R21, R01, K grants, small foundation awards)
  - 60 business days for complex multi-project extramural grant preparation (e.g., P30, P50, P01, U54, large foundation awards)
Collaboration Deliverables: Grants

- Not just about including a SAP and power analysis

- A biostatistician should provide critical review of your full proposal

- We look for clarity and consistency in language around:
  - Aims
  - Exposure-Outcome relationships
  - Study design
  - Data types, variables collected, and data management
Improving a Grant Aim

- **Aim**: Assess changes in redox potential following glucose ingestion

- **Sub-Aim**: Older age, female sex, and lower insulin production will be associated with weaker effect of glucose ingestion on redox potential

- **Statistician Thoughts**
  - No identification of cohort(s) or data sources
  - Needing clarity on temporality of “changes”
  - Are data paired? What are the data types?
  - Sub-aim: What does “weaker effect” mean?
Improving a Grant Aim

• **Aim**
  – Assess for **paired change** in redox potential from 0 to 120 minutes, following glucose ingestion, in a **single cohort of CF patients** aged 3-7 years at risk for developing CFRD

• **Sub-Aim**
  – Older age (6-7 yrs) will be associated with a reduced **paired change** in redox potential, following glucose ingestion, relative to **younger age (3-5 yrs)**

• Clearly defined sample, data types, exposure-outcome relationships, and anticipated relationships
Collaboration Deliverables: Grants

- Contingent on duration, size, and scope of your study, a biostatistician should provide you with feedback on FTE budget
  - Generic R01-type example

<table>
<thead>
<tr>
<th>Grant Year 1</th>
<th>Grant Year 2</th>
<th>Grant Year 3</th>
<th>Grant Year 4</th>
<th>Grant Year 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Biostatistician</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>10%</td>
<td>5%</td>
<td>5%</td>
<td>35%</td>
</tr>
<tr>
<td><strong>Data Manager</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30%</td>
<td>15%</td>
<td>10%</td>
<td>10%</td>
<td>15%</td>
</tr>
</tbody>
</table>

- Speak with your expert team for specific FTE numbers appropriate for your study
Collaboration Deliverables: Analysis

• The 80/20 rule for analysis:
  – 80% of analysis time is spent on data wrangling (cleaning, standardizing, general processing)
  – 20% on actual analysis

• Analysis deliverables are standardized based on study design:
  – CONSORT (randomized clinical trials)
  – STROBE (observational studies)
CONSORT (Randomized Clinical Trials)
STROBE (Observational Studies)

- Common designs:
  - Cohort study
  - Case-control
  - Cross-sectional
Tip #4: Plan appropriately for collaboration

Panel and Audience Discussions
Tip #5: Presentation of conclusions is a team effort

• Clinical significance versus statistical significance
  – Moving away from sole reliance on p-values (with sufficient sample size, any effect can be < 0.05)
  – Increasingly prefer measures of effect size (e.g., standardized differences, risk/odds ratios) and 95% CI
  – Ask researchers, “What difference would lead you to rethink your practice?”

• Adds validity to the research and improves chances of other researchers replicating the findings
Tip #5: Presentation of conclusions is a team effort

• Biostatisticians should review a full product
  – Ensure conclusions do not go beyond the results
  – Check estimates are interpreted correctly (risk v. odds ratios)
  – Evaluate if bias/confounding were appropriately handled

• Researchers should know what and why certain statistics were used
  – Results should not be a black box
  – Researcher interprets in public or clinical health context
Tip #5: Presentation of conclusions is a team effort

Panel and Audience Discussions
Best Practices for Partnering with a Biostatistician in your Research

Scott Gillespie, MS, MSPH
Associate Director and Lead Biostatistician
Emory Pediatric Biostatistics Core

K-Club
September 12th, 2022