

# Effectively Working with Biostatisticians as Collaborators

ARI Webinar

March 17, 2020

Joanne Beer, PhD and Adam Ciarleglio, PhD



Milken Institute School  
of Public Health  
THE GEORGE WASHINGTON UNIVERSITY

# Outline

- Collaboration
  - What is a biostatistician and how can you collaborate with them?
- Logistics
  - Finding a biostatistician and thinking about effort.
- Additional Considerations
  - Things to think about when working with a biostatistician.

# Collaboration

# What is a Biostatistician?

- A biostatistician is a **person** who applies principles of study design and statistical methodology to address public health/biomedical research questions.
- Typical activities for a biostatistician:
  - Study design
  - Power and sample size determination
  - Generate randomization plans
  - Write statistical analysis sections of grants
  - Data cleaning and wrangling
  - Analysis
  - Presenting results
  - Working with co-investigators to write manuscripts
  - ...
- Many biostatisticians also devote time to developing new statistical methods to address problems that need solutions.

# Emerging Roles for Biostatisticians

Funding Opportunity Title	Computational Approaches for Validating Dimensional Constructs of Relevance to Psychopathology (R01 Clinical Trial Optional)
Activity Code	R01 Research Project Grant
Announcement Type	New
Related Notices	<ul style="list-style-type: none"><li>• <b>August 23, 2019</b> - Clarifying Competing Application Instructions and Notice of Publication of Frequently Asked Questions (FAQs) Regarding Proposed Human Fetal Tissue Research. See Notice <a href="#">NOT-OD-19-137</a>.</li><li>• <b>July 26, 2019</b> - Changes to NIH Requirements Regarding Proposed Human Fetal Tissue Research. See Notice <a href="#">NOT-OD-19-128</a>.</li><li>• <b>November 26, 2018</b> - NIH &amp; AHRQ Announce Upcoming Updates to Application Instructions and Review Criteria for Research Grant Applications. See Notice <a href="#">NOT-OD-18-228</a>.</li></ul>
Funding Opportunity Announcement (FOA) Number	RFA-MH-19-242

Funding Opportunity Title	Explainable Artificial Intelligence for Decoding and Modulating Neural Circuit Activity Linked to Behavior (R01 Clinical Trial Optional)
Activity Code	R01 Research Project Grant
Announcement Type	New
Related Notices	<ul style="list-style-type: none"><li>• <b>September 17, 2019</b> - Notice of Change in Locus of Review in PAR-19-344 "Explainable Artificial Intelligence for Decoding and Modulating Neural Circuit Activity Linked to Behavior (R01 Clinical Trial Optional)". See Notice <a href="#">NOT-MH-19-051</a>.</li><li>• <b>August 23, 2019</b> - Clarifying Competing Application Instructions and Notice of Publication of Frequently Asked Questions (FAQs) Regarding Proposed Human Fetal Tissue Research. See Notice <a href="#">NOT-OD-19-137</a>.</li><li>• <b>July 26, 2019</b> - Changes to NIH Requirements Regarding Proposed Human Fetal Tissue Research. See Notice <a href="#">NOT-OD-19-128</a>.</li></ul>
Funding Opportunity Announcement (FOA) Number	PAR-19-344

Funding Opportunity Title	High-Priority Areas for Research Leveraging EHR and Large-Scale Data (R01 Clinical Trial Not Allowed)
Activity Code	R01 Research Project Grant
Announcement Type	New
Related Notices	<ul style="list-style-type: none"><li>• <b>February 14, 2020</b> - Notice of Special Interest to Highlight Research Priorities for Risk Algorithms Applied in Healthcare Settings to Improve Suicide Prevention. See Notice <a href="#">NOT-MH-20-031</a>.</li><li>• <b>August 23, 2019</b> - Clarifying Competing Application Instructions and Notice of Publication of Frequently Asked Questions (FAQs) Regarding Proposed Human Fetal Tissue Research. See Notice <a href="#">NOT-OD-19-137</a>.</li><li>• <b>July 26, 2019</b> - Changes to NIH Requirements Regarding Proposed Human Fetal Tissue Research. See Notice <a href="#">NOT-OD-19-128</a>.</li><li>• <b>November 26, 2018</b> - NIH &amp; AHRQ Announce Upcoming Updates to Application Instructions and Review Criteria for Research Grant Applications. See Notice <a href="#">NOT-OD-18-228</a>.</li></ul>
Funding Opportunity Announcement (FOA) Number	PAR-18-929

Funding Opportunity Title	Secondary Data Analyses to Explore NIMH Research Domain Criteria (R03)
Activity Code	R03 Small Grant Program
Announcement Type	Reissue of <a href="#">PAR-14-008</a>
Related Notices	<ul style="list-style-type: none"><li>• <b>August 23, 2019</b> - Clarifying Competing Application Instructions and Notice of Publication of Frequently Asked Questions (FAQs) Regarding Proposed Human Fetal Tissue Research. See Notice <a href="#">NOT-OD-19-137</a>.</li><li>• <b>July 26, 2019</b> - Changes to NIH Requirements Regarding Proposed Human Fetal Tissue Research. See Notice <a href="#">NOT-OD-19-128</a>.</li><li>• <b>November 26, 2018</b> - NIH &amp; AHRQ Announce Upcoming Updates to Application Instructions and Review Criteria for Research Grant Applications. See Notice <a href="#">NOT-OD-18-228</a>.</li><li>• <b>NOT-OD-18-009</b> - Reminder: FORMS-E Grant Application Forms and Instructions Must be Used for Due Dates On or After January 25, 2018.</li><li>• <b>September 20, 2017</b> - Updates to Active Funding Opportunity Announcements to Prepare for Policy Changes Impacting Due Dates On or After January 25, 2018. See <a href="#">NOT-OD-17-114</a>.</li><li>• <b>May 10, 2017</b> - New NIH "FORMS-E" Grant Application Forms and Instructions Coming for Due Dates On or After January 25, 2018. See <a href="#">NOT-OD-17-062</a>.</li></ul>
Funding Opportunity Announcement (FOA) Number	PAR-17-158

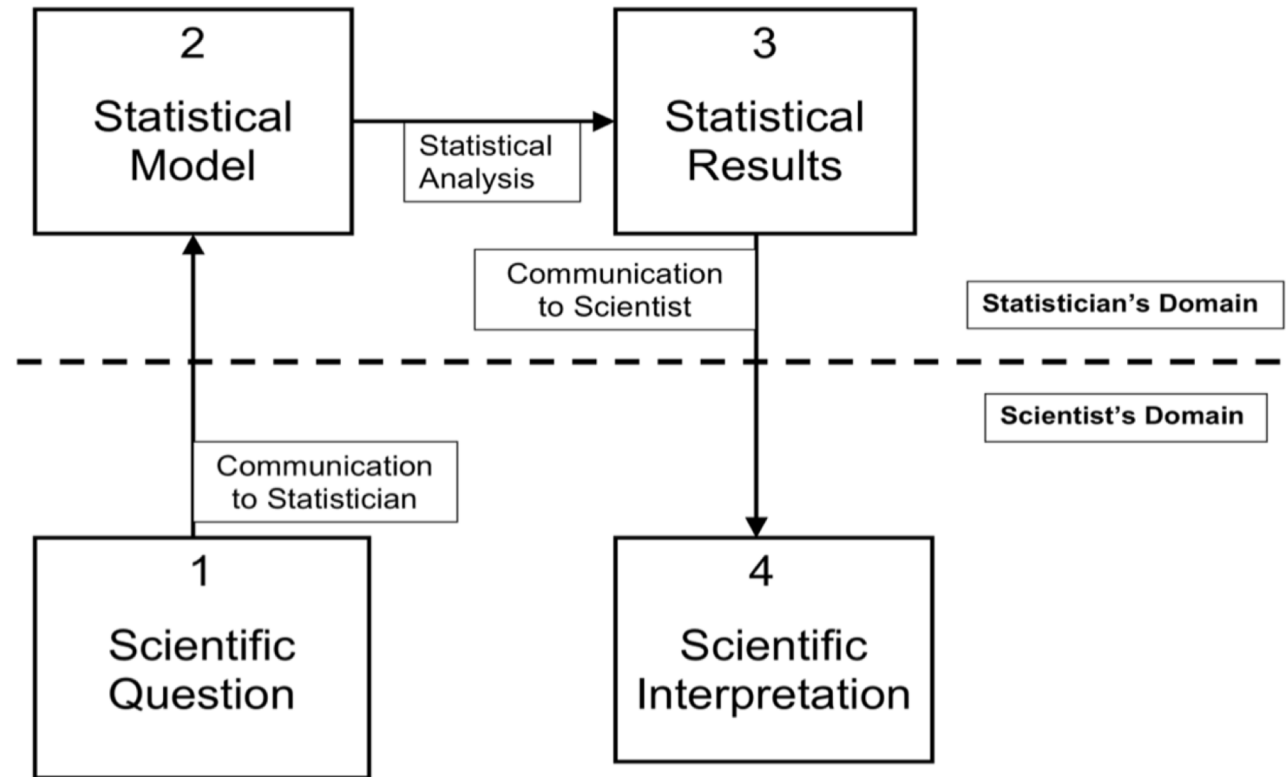
# Pitch for the Collaborative Mindset

- Many biostatisticians are interested in leading projects/papers.
- Many are also interested in developing new statistical methods to address challenging problems that need solutions.
- Many biostatisticians are on the lookout for interesting projects where novel methods can be judiciously applied.
- Develop a symbiotic relationship.
  - You might get to be an author on a statistical methods paper!
- Building a collaboration will pay dividends with respect to future projects.
  - Even if project/grant requires standard analysis (simple power calculations, straightforward application of methods, etc.), be forward thinking.

# Developing Careers Together

- While it may be advantageous to have a senior biostatistician on your grant, there are benefits to finding one who is at a similar stage in their career as you are.
  - Senior biostatisticians may be invested in long-established collaborations.
- Some (hopefully) obvious comments:
  - People like to work with people that they trust.
  - People like to work with people who see their contributions as crucial.
  - People probably work a little bit harder for those who appreciate them.

# More on Collaboration and Developing Careers Together



**Figure 1.** A Simplified Model of Collaborative Statistical Practice

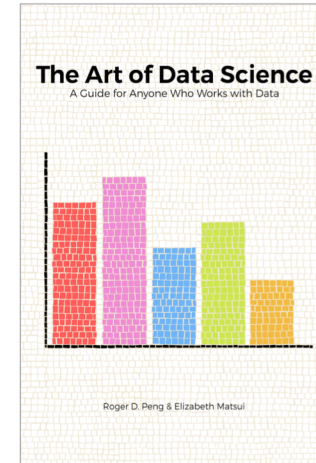
Samsa, G. A day in the life of a collaborative biostatistician deconstructed: Implications for curriculum design. *Journal of Curriculum and Teaching*. (2018) 7(1): 20 – 31. (<https://files.eric.ed.gov/fulltext/EJ1167186.pdf>)



# Example of Successful Collaboration

- **Elizabeth Matsui, MD** (Professor of Population Health and Pediatrics, Dell Medical School UT Austin)
- **Roger Peng, PhD** (Professor of Biostatistics, Johns Hopkins Bloomberg School of Public Health)

- Book: *The Art of Data Science* (<https://leanpub.com/artofdatascience>)



- Blog: *Simply Statistics* (<https://simplystatistics.org>)



- Podcast: *The Effort Report* (<http://theeffortreport.com/>)



# The Care and Feeding of the Biostatistician

- Title of a *Simply Statistics* blog post by Elizabeth Matsui
  - <https://simplystatistics.org/2013/10/08/the-care-and-feeding-of-the-biostatistician/>
- Her suggestions/comments:
  - The biostatistician is not a computational tool.
  - Teach them your language.
  - Try to learn their language.
  - Draw pictures.
  - Be willing to share your data and your ideas.
  - Be respectful of time. (Not in original list.)

# Biostatistician's Impact in a Grant Review

- Consider R01 scoring – Having a biostatistician as a co-investigator can impact:
  - **Investigators**
    - Reviewers tend to appreciate statistical expertise among investigators.
  - **Approach**
    - Sound statistical analysis section is crucial.
  - **Innovation**
    - Could possibly include statistical methods development or application of appropriate cutting edge methods.
  - **Overall Score**

# Biostatistician's Impact in a Manuscript Review

- Many journals have statistical reviewers who focus on soundness the statistical analyses used.
  - Even if no dedicated statistical reviewer, usually have at least one savvy reviewer who can tell when things aren't right.
- If reviewers find issues with statistical analysis, they often suggest finding a biostatistician.
  - It's challenging to bring in a biostatistician after the paper has been submitted.
  - Possibly longer time to resubmission.

# Logistics

# Finding a Biostatistician I

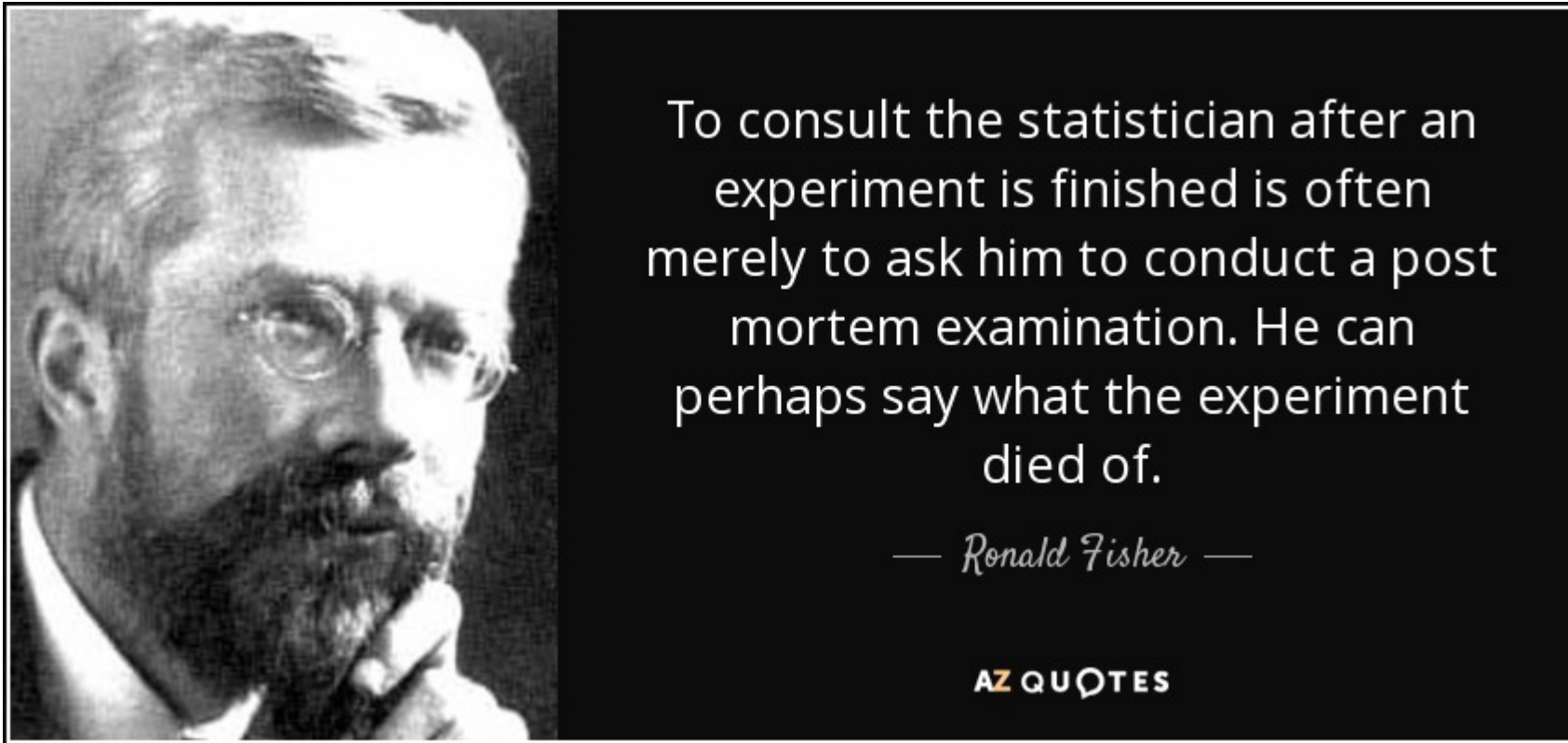
- Does your department/division/group have a dedicated biostatistician?
- Is there a Biostatistics Department at your institution?
  - Many Biostatistics Departments have consulting services available for brief consultations.
  - There may be a Biostatistics, Epidemiology, and Research Design (BERD) program as part of a CTSI.
- You may need to reach outside of your institution
  - If so, there may be a lot to think about: e.g., compensation structure, data access, etc.

# Finding a Biostatistician II

- Go in with the mindset of wanting to find a collaborator.
- If you have options for who can be your biostatistician:
  - Meet briefly with your potential collaborators.
  - Take time to review CVs, publications, grants.
  - Try to get an idea of whether the biostatistician has interests/skills that could serve to strengthen your research.

# Initiating a Project with a Biostatistician

- When you have a project/grant idea, reach out to your biostatistician ASAP.
- Waiting too long to involve a biostatistician can be ruinous.



To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of.

— *Ronald Fisher* —

AZ QUOTES



# Initiating a Project with a Biostatistician

- When you have a project/grant idea, reach out to your biostatistician ASAP.
  - Waiting too long to involve a biostatistician can be ruinous.
- Bring as detailed a draft as possible and be open to refining research questions.
  - The biostatistician is likely to ask questions about the variables that you plan to collect and associations that you plan to investigate.
- Be clear about what you are expecting.
  - Randomization
  - Data management
  - Data archiving
  - ...

# Time and Effort I

- No hard-and-fast rules for determining percent effort.
- Your institution/group may have norms that you should follow.
- In my experience, no one wants to be on 20 projects at 5% effort each.
- 5% effort about covers the services rendered in preparing a grant application.
  - Ex. 5% might cover:
    - planning meetings,
    - power and sample size analysis,
    - writing grant sections, and
    - maybe 1 – 1.5 hour weekly meetings during grant period
    - no analysis

# Time and Effort II

- Common for PhD level biostatisticians to require additional funding to cover master's level biostatistician working with them.
- For many grants, may make sense to increase percent effort over life of the grant.

- Ex. 

Year 1	Year 2	Year 3	Year 4	Year 5
PhD: 5% MS: 0%	PhD: 5% MS: 0%	PhD: 10% MS: 10%	PhD: 10% MS: 10%	PhD: 15% MS: 10%

- Resources:

- *Estimating Percent Effort*: <https://stattrak.amstat.org/2020/02/01/estimating-percent-effort/>
- Perkins, S., *et al.* Best practices for biostatistical consultation and collaboration in academic health centers. *The American Statistician*. (2016) 70(2): 187 – 194. <https://doi.org/10.1080/00031305.2015.1077727>
- Guidelines for Estimating Biostatistician Effort and Resources on Grants:  
[https://health.ucdavis.edu/ctsc/area/biostatistics/Documents/UCD\\_Biostat\\_Effort\\_Guidelines.pdf](https://health.ucdavis.edu/ctsc/area/biostatistics/Documents/UCD_Biostat_Effort_Guidelines.pdf)

# Involving a Biostatistician Late in a Project

- Try to avoid this. When it cannot be avoided, have realistic expectations.
- Ex. You run analyses, write up paper, submit, and reviews come back saying that you need a biostatistician.
  - If you have funds:
    - Give time to get familiar with the data.
    - Biostatistician may take a different approach to analysis.
    - Results may be different.
  - If you plan to run new analyses yourself:
    - Biostatistician may not be able to give best advice without seeing the data.
    - May not be proficient with the software you are using.
    - ...

# Additional Considerations

# Discussing Hypotheses

- Operationalize:
  - X (predictor)
  - Y (outcome)
  - P (population, e.g. geriatric depressed patients)
- Biostatistician will push you to identify key variables and how they are measured.
  - Think about causal relationships between variables. Potential confounders.
- Specify primary, secondary, and exploratory aims/hypotheses.
- Power and sample size geared towards primary aims.

# Power and Sample Size Considerations

- For well defined/operationalized primary hypotheses
- What does the literature say about effect sizes?
- Do you have results from pilot data that can help inform power/sample size.
- If you give a firm sample size – the biostatistician may not come back with a power that you “like”
- What goes into a power/sample size calculation?
  - Consider results under different assumptions.
  - Simulations may be necessary.

# Secondary and Exploratory Aims

- Typically don't need to power for these.
- Mediation, Moderation (sub-group analysis)
- Development of predictive models



# Help statisticians understand the data

- Consider bringing the data analyst (might be master's level statistician or PhD student) in to witness data collection if possible
  - For example, cognitive assessment battery, lab-generated data
  - Help understand
    - the meaning of the variables
    - quality of the data
    - what could go wrong – why data might be missing or miscoded
    - how to interpret anomalies

# P-Values and Multiple Testing

- P-values have become controversial
  - “Replication crisis” in science (see work of Ioannidis)
  - Some journals have banned p-values in favor of confidence intervals
  - Some researchers have suggested more stringent thresholds like  $p < 0.005$

# P-Values and Multiple Testing

- P-values have become controversial
  - “Replication crisis” in science (see work of Ioannidis)
  - Some journals have banned p-values in favor of confidence intervals
  - Some researchers have suggested more stringent thresholds like  $p < 0.005$
- “P-hacking” or “Garden of Forking Paths”

The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time\*

Andrew Gelman<sup>†</sup> and Eric Loken<sup>‡</sup>

14 Nov 2013

# P-Values and Multiple Testing

- P-values have become controversial
  - “Replication crisis” in science (see work of Ioannidis)
  - Some journals have banned p-values in favor of confidence intervals
  - Some researchers have suggested more stringent thresholds like  $p < 0.005$
- “P-hacking” or “Garden of Forking Paths”
  - Trying lots of stuff and cherry-picking the best results
    - If you selectively report, p-values are not really valid
  - Best to pre-determine the analysis plan
    - “pre-registration” / “registered reports”
  - If your study is exploratory, that is fine! Be explicit about it.

# High Dimensional Data and Machine Learning Methods

- If you plan to obtain and use high dimensional data (e.g., brain imaging data, data from wearable devices, etc.) try to find someone who specializes in analysis of this kinds of data
- Not all biostatisticians are well-versed in these methods.

# Statistical Sections of Grants

- Make sure that the analysis can assess the hypothesis.
- Know the audience of reviewers and tailor.
- Usually best to leave out jargon. Don't use it to impress if it doesn't add anything.
- Think about missing data.
- Convince reviewer that you have considered all of the possible issues.
- Who is responsible for data management and data sharing?
  - Often not the biostatistician.

# Other general advice

- Clarify the research question / hypotheses!
  - Well-motivated by past work
  - Try to formulate so that the result is publishable regardless of outcome of your statistical test(s)
  - Exploratory is fine, but have some general outline of how it will proceed
- Think about need for validation (for parameter tuning in predictive models) and test samples (to assess model accuracy on new data)
  - Especially important for exploratory analyses and predictive models
- Collect (lots of) high quality data! (Garbage in, garbage out)

# From a (junior) biostatistician's perspective

- Want to have multiple options for collaborative projects
  - Ability to say no to collaborations that are not a good fit
  - Pick projects and collaborators carefully
- If collaboration is going badly
  - Communicate better – this may be challenging for junior person, so
  - Have a senior mentor act as mediator if possible
- Ability to have senior mentor intervene is one reason why statisticians generally don't want to be isolated, working as the sole statistician in a given setting
  - Workload can be overwhelming, collaborators may not appreciate how difficult the statistical work is



# Wrap-Up: Final Thoughts

- Sound application of statistical methods is crucial for successful projects, grants, and manuscripts.
- Try to develop a collaborative relationship with your biostatistician.
- Be prepared to teach each other and to learn from each other.
- Have realistic expectations about the biostatistical support you need and engage as early as possible.

Thank You!