Received: 7 February 2022 | Revised: 12 April 2022 | Accepted: 23 April 2022

DOI: 10.1111/jgs.17888

COMMENTARIES

Journal of the
American Geriatrics Society

Modeling success: How to work effectively with your biostatistician

Jiha Lee MD, MHS[1] | **Biren B. Kamdar MD, MBA, MHS**[2] | **Jaclyn Bergstrom MS**[3] | **Terrence E. Murphy PhD**[4] | **Thomas M. Gill MD**[5]

[1]Department of Internal Medicine, Division of Rheumatology, University of Michigan, Ann Arbor, Michigan, USA

[2]Division of Pulmonary, Critical Care & Sleep Medicine, Department of Internal Medicine, University of California, San Diego (UCSD), La Jolla, California, USA

[3]Division of Epidemiology, Department of Family Medicine and Public Health, University of California, San Diego (UCSD), La Jolla, California, USA

[4]Section of Geriatrics and Program on Aging, Department of Internal Medicine, Yale School of Medicine, New Haven, Connecticut, USA

[5]Division of Geriatric Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, Connecticut, USA

Correspondence
Jiha Lee, Department of Internal Medicine, Division of Rheumatology, University of Michigan, 300 North Ingalls Building, Room 7C27, Ann Arbor, MI 48109–5422, USA.
Email: jihalee@umich.edu

Funding information
Clin-STAR Coordinating Center, Grant/Award Number: U24AG065204; GEMSSTAR, Grant/Award Number: R03AG067975; Paul B. Beeson Emerging Leaders Career Development Award, Grant/Award Number: K76AG059936; Yale Claude D. Pepper Older Americans Independence Center, Grant/Award Number: P30AG021342; UM Claude D. Pepper Older Americans Independence Center, Grant/Award Number: P30AG024824

"Our goal is to provide early-stage clinician-investigators with information that will facilitate an effective and mutually beneficial working relationship with a biostatistician and, in turn, ensure the asuccessful completion of their research projects while also enriching their career development."

The quality of a clinical research project depends, among other things, on an appropriate statistical analysis plan, reproducible analysis of the data, and interpretable reporting of the results, preferably in partnership with a biostatistician. For early-stage clinician-investigators, understanding how to effectively collaborate and communicate with a biostatistician is vital, as a failure to do so can lead to unwanted project delays, manuscript rejections, or worst, complete failure of a research project. At the 2021 Annual Meeting of the Clinician-Scientists Transdisciplinary Aging Research (Clin-STAR) Coordinating Center, we hosted a session where two clinician-investigator and biostatistician dyads shared their experiences and insights on how to work effectively in a complementary and collaborative manner.[1] In this commentary, we discuss lessons learned, focusing on a core set of issues and approaches that should be considered at different phases of a research project. Our goal is to provide early-stage clinician-investigators with information that will facilitate an effective and mutually beneficial

working relationship with a biostatistician and, in turn, ensure successful completion of their research projects while also enriching their career development.

CROSS-CUTTING THEMES

Although early-stage clinician-investigators should be familiar with best practices in data management[2] and have a fundamental knowledge of biostatistical principles and some comfort in performing basic statistical analyses, most clinical research projects requiring quantitative data analysis will benefit from the expertise of a well-trained and experienced biostatistician. We discuss key considerations related to identifying and working effectively with a biostatistician (Table 1).

HOW TO FIND A BIOSTATISTICIAN

Research mentors are generally the first and most valuable resource when early-stage investigators are trying to determine their biostatistical needs. Mentors often have well-established working relationships with one or more biostatisticians for their own research and are also knowledgeable about the availability of other biostatisticians through the institution's Clinical and Translational Science Awards (CTSA) Program, School of Public Health, Pepper Center, or comparable programs. If such inquiries are unsuccessful, the early-stage clinician-investigators should reach out to the Division Chief and/or Department Chair for additional guidance and suggestions. At some institutions, there is a shortage of well-trained biostatisticians and forming a collaborative working relationship can be challenging and time-consuming.

Mentors should be engaged early in the process to define the level of statistical support needed, discuss how to identify candidate biostatisticians, and explore sources of funding. Biostatisticians differ considerably in terms of their prior training (masters vs. PhD), areas of expertise (longitudinal studies, clinical trials, genomics, etc.), and role within the institution (faculty vs. research staff). Some may be trained in a related quantitative field such as epidemiology, bioinformatics, demography, or data science. Biostatisticians can work in different capacities; they may serve as a consultant with a task-limited role of providing advice on methods or analytic output, or as a core team member involved in all phases of the research project, including the planning, analysis, and interpretation of data, and presentation of findings. With the support of their mentor, early-stage investigators should identify and explicitly define the best model of working with a biostatistician to ensure that their needs are met.

Key points

- To rigorously address clinically meaningful scientific questions, early-stage clinician-investigators should engage biostatisticians early in the research planning process and set clear expectations and realistic timelines.
- Early-stage clinician-investigators should identify and explicitly define the best model of working with a biostatistician to ensure that their needs are met. Biostatisticians differ considerably in terms of their training and can work in different capacities.
- Early-stage clinician-investigators and biostatisticians should identify roles and responsibilities during different phases of a research project to facilitate an effective and mutually beneficial working relationship.

Why does this paper matter?

Producing high-quality research depends on effective teamwork between the clinician-investigator and biostatistician. We share information on how early-stage clinician-investigators can effectively collaborate and communicate with biostatisticians so that they can successfully complete their research projects and advance their career development.

HOW TO BUILD A COMPLEMENTARY AND COLLABORATIVE WORKING RELATIONSHIP

Biostatisticians generally have many competing commitments as they often work with several investigators on multiple projects simultaneously. For this reason, it is important to set clear goals and expectations on issues such as project timelines and authorship. Early-stage investigators should set priorities and anticipate rate-limiting steps, which often involve the creation of an analytic dataset. The division of labor between data management and statistical analysis should be discussed. Some biostatisticians may assume data management responsibilities, while others will not. To ensure the timely completion of research, the resources required for data management and statistical analysis should be enumerated for new projects and grant proposals.

TABLE 1 Key considerations for working effectively with a biostatistician

Common questions	Approaches
When to engage a biostatistician?	• Early in a research project (i.e., during the design stage)
How to identify a biostatistician?	• Discuss with mentor • Explore institutional resources (e.g., CTSA programs, School of Public Health, Pepper Center, biostatistics office hours) • If needed, reach out to the Division Chief and/or Department Chair for additional guidance and suggestions
What are the desired attributes of your biostatistician?	• Level of training (e.g., Masters vs PhD) • Skills and experiences relevant to the project • Proficient with specific analytic programs (e.g., SAS, STATA) • Knowledge of relevant datasets (e.g., Medicare files) • Available (or can supervise a colleague) to help with your project
Roles and responsibilities?	*Clinician investigator* • Define scientific questions and hypothesis • Provide supporting citations or previous work as a framework • Apply for grants and budget for analytic support • Coordinate creation of dataset and data management • Co-develop analytic plan • Attempt simple analyses (e.g., descriptive analysis, bivariate regression models) • Co-design Tables and Figures • Review and understand completed analyses • Write and revise manuscript *Biostatistician* • Draft analytic plan • Provide direction regarding dataset design • Perform statistical analyses • Retain log of analytic code and output • Input results for Tables and Figures • Review interpretation of results and presentation/visualization of findings • Write statistical methods section and review manuscript
What are the components of an analytic plan?	• Overall objective • Specific aims and hypotheses • Analytic sample, with inclusion/exclusion criteria • Outcomes of interest (dependent variables) • Exposures of interest (independent variables) • Covariates • Outline of the statistical analysis plan
How to manage the data?	• Practice consistency in data entry and organization • Create data dictionary • Use data capture tools to ensure data integrity • Access institutional resources to support and obtain training in data management • Be prepared to import raw data, clean dataset, generate variables, address missing data • Make regular backups of data, in multiple secure locations • Review NIH Data Sharing Policy available online at: https://grants.nih.gov/grants/policy/data_sharing/

Setting goals and expectations?	• Explicitly communicate tasks, deliverables, and timelines
	  ○ When should dataset be ready for analysis?
	  ○ Identify firm deadlines (e.g., conference abstracts, manuscript revisions, grants)
	• Discuss authorship
	• Establish schedule (e.g., work timelines, regular meetings at appropriate intervals)
	• Gain knowledge from mentor about biostatistician compensation (for current and future projects)

(Continues)

TABLE 1 (Continued)

Common questions	Approaches

How to conduct a productive analytic meeting?	*Clinician investigator*
	• Prepare for each meeting
	  ○ Initial meeting: provide thorough background on the topic with relevant publications; discuss prior work, statistical approaches, and pitfalls
	  ○ Subsequent meetings: Develop an agenda, revisit tasks and timelines, identify next steps
	• Follow-up with each meeting with concise summary, action items, and date/time for next meeting
	Biostatistician
	• Share interim analytic output files/summary for review before meetings
	• Communicate analytic challenges and ask clarifying questions in-between meetings

Common pitfalls and barriers?	*Clinician investigator* • Uses statistical terms loosely or incorrectly, leading to miscommunication • Is afraid to ask for clarification • Is hesitant to admit not understanding the study design or analytical approach • Assume there is such a thing as a "quick question" • Asks biostatisticians to re-run analyses to obtain significant results *Biostatistician* • Is not familiar with relevant clinical issues that influence creation of analytic sample and subsequent data analyses • Often has multiple competing priorities and projects that limit availability • Presents analytic outputs in a format that is difficult to understand/interpret

Where to acquire statistical software and codes?	• Institutional resources
	  ○ Online bookstore (free software often available)
	  ○ Available licenses within department/division
	  ○ Startup/Mentor funds to purchase software
	• Geriatrics Research Algorithms & Statistical Programs (GRASP), available online at: https://www.peppercenter.org/public/grasp.cfm

HOW TO COMMUNICATE EFFECTIVELY WITH YOUR BIOSTATISTICIAN

Clinician-investigators and biostatisticians often speak different languages; the biostatistician may not be familiar with the relevant clinical issues, and the clinician-investigator may use some analytic terms (e.g., effect, mediation, significant) loosely or incorrectly. Early-stage investigators and biostatisticians should ask clarifying questions to reduce potential misunderstandings. Regularly scheduled analytic meetings, supplemented by email exchanges and ad hoc calls, can optimize the likelihood that projects are completed on time. Early-stage investigators should be responsible for establishing the meeting agenda, providing background information about a specific project, and identifying relevant publications that could serve as potential templates for the planned analysis. Before completion of the meeting, next steps should be identified, and the timeline should be revisited. After the meeting, the early-stage investigator should provide a concise summary, focusing specific attention on important decisions and corresponding modifications to the analytic plan.

HOW TO WORK EFFECTIVELY WITH A BIOSTATISTICIAN ON A RESEARCH PROJECT

Successfully implementing a research project requires several steps, including study design, data collection, data analysis, interpretation of results, and dissemination of findings. Although early-stage clinician-investigators should define the scientific question(s) and drive the research project, coordinated collaboration with well-trained biostatisticians will increase the likelihood that appropriate analytic approaches are adopted, thereby enhancing scientific rigor. Taking a team-science approach, we discuss the roles and responsibilities of the clinician-investigator and biostatistician during different phases of a research project (Figure 1).

PLANNING PHASE

The clinician-investigator should be responsible for formulating research question(s) and related hypotheses. After project conception, early-stage investigators should

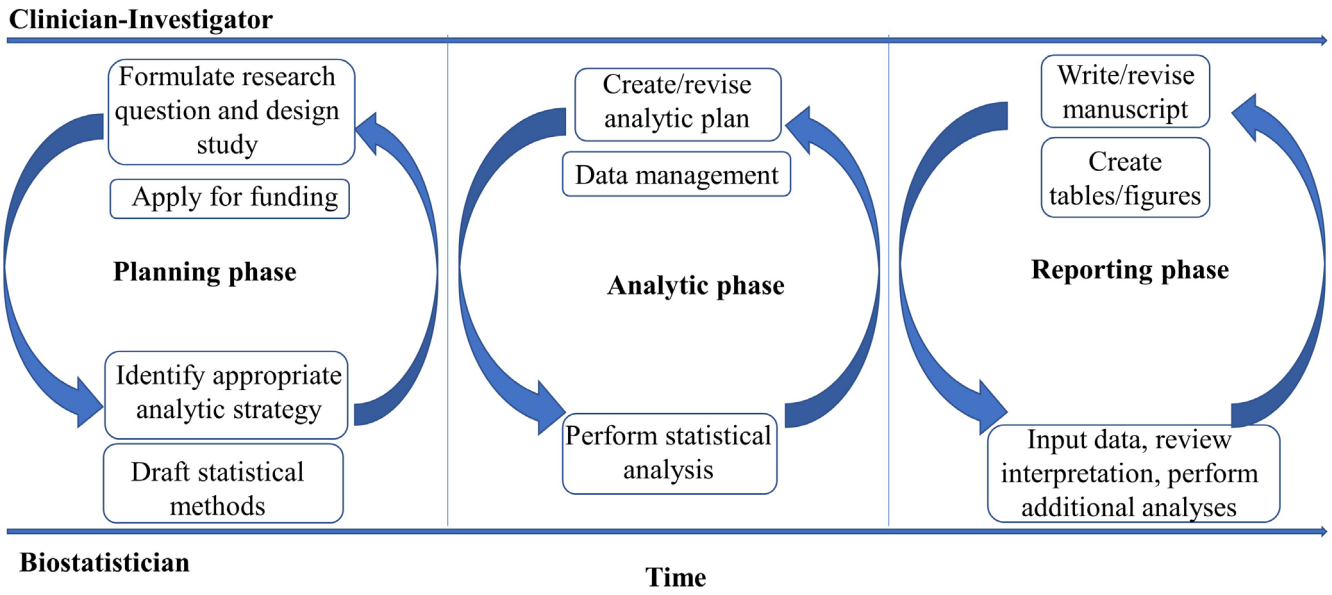

FIGURE 1 Team-science approach to working collaboratively with a biostatistician during the different phases of a research project. The arrows are intended to highlight the collaborative relationship and iterative procedures between the clinician-investigator and biostatistician

engage their biostatistician to discuss whether the planned study design and data source(s) are appropriate, confirm that the hypothesis is testable, obtain power and sample size calculations, and identify the best modeling strategies. Adequate time should be allocated to accomplish these tasks. Biostatisticians usually take the lead in drafting the statistical methods sections for grant applications, and they should provide relevant citations to support the proposed analytic strategies. Clinician-investigators should not be afraid to ask questions if they do not understand the statistical analysis plan or recommendations of the statistician. In the planning phase, the research questions and/or hypotheses may need to be revised or refined. However, the clinician-investigator must ensure that the resulting scientific questions/hypotheses remain clinically meaningful. In addition, clear expectations on the scope of work, authorship, and realistic timelines should be established.

ANALYTIC PHASE

Data management is an integral component of analysis and involves collecting, organizing, and securely storing the data.[2] Adopting sound data management practices that include data validation techniques and use of tools such as REDCap (Research Electronic Data Capture)[3] can reduce common errors during data handling, streamline data analysis, and facilitate visualization and reporting.[4] A data dictionary that includes information on the definition, structure, and content of the data elements is a key tool that is available through REDCap. For protected health information, clinician-investigators, and biostatisticians should be familiar with and adhere to institutional and governmental data sharing regulations/guidelines[5] and obtain required Institutional Review Board approval(s). In most cases, only de-identified data may be shared, and extreme caution should be exercised with management of highly sensitive identifiable data. Many academic institutions provide high-performance computing environments that ensure secure analysis of sensitive data sets regulated by federal privacy policies, proprietary access agreements, or confidentially agreements.

Under the guidance of their mentor(s), early-stage investigators should draft a short (i.e., one-page) analytic proposal that provides information on the overall objective, specific aims, and related hypotheses, analytic sample with inclusion/exclusion criteria, outcomes of interest, independent variables, covariates, and brief outline of the analysis plan. If possible, the clinician-investigator should share published reports that could serve as potential templates for the proposed analysis. The more complete analysis plan is typically formulated by the biostatistician in partnership with the clinician-investigator. This should help to ensure that the best modeling strategies are selected based on the study design, that is, cross-sectional, longitudinal, clinical trial, etc. Analytic proposals can serve as the focus of discussion between the clinician-investigator and biostatistician. Changes to the proposal should be clearly documented. To enhance efficiency, opportunities to adapt existing statistical code should be investigated.

During the analytic phase, biostatisticians are typically open to summarizing the methods or sharing their codes. Over time, these interactions provide ample opportunities for reciprocal training in biostatistical methods and clinical science.

REPORTING PHASE

Although practices may vary, the statistical analysis section is usually drafted by the biostatistician, while the rest of the manuscript is commonly written by the clinician-investigator, including the abstract, introduction, remainder of methods, results, and discussion. The clinician-investigator should also be responsible for creating table shells and templates for figures so that the biostatistician can provide or enter relevant results from their statistical analyses. Early-stage investigators should carefully review and revise the statistical analysis section to enhance clarity and consistency and to ensure that the description of the statistical methods is understandable to a clinical audience.

Manuscripts submitted to peer-reviewed journals almost always require revision, and reviewers may raise concerns about the analytic strategy or request additional analyses. The clinician-investigator should respond to each reviewer comment and enlist the assistance of their biostatistician when attempting to address analytic questions or concerns. Maintaining detailed notes from analytic team meetings and logs of analytic code and output can facilitate appropriate responses to reviewer comments. If additional analyses are requested and deemed appropriate, the biostatistician should be informed of the resubmission deadline so that sufficient time and effort can be allocated and delays in the publication process can be avoided.

CONCLUSION

Producing high-quality research depends on effective teamwork between the clinician-investigator and biostatistician. To rigorously address clinically meaningful scientific questions, early-stage investigators should engage biostatisticians early in the research planning process and set clear expectations and realistic timelines. The tenets summarized in this article, with individualized adaptation over time, should help nurture long-term productive collaborative relationships and contribute to career development.

AUTHOR CONTRIBUTIONS

All authors meet the criteria for authorship stated in the Uniform Requirements for Manuscripts Submitted to Biomedical Journals.

FUNDING INFORMATION

This work was supported, in part, by the Clin-STAR Coordinating Center (U24AG065204). Dr. Gill and Dr. Murphy are supported by the Yale Claude D. Pepper Older Americans Independence Center (P30AG021342). Dr. Lee is supported by the GEMSSTAR Program (R03AG067975). Dr. Kamdar is supported by a Paul B. Beeson Emerging Leaders Career Development Award in Aging (NIAK76AG059936).

CONFLICT OF INTEREST

Lee, J, None; Kamdar, B, None; Bergstrom, J, None; Murphy, T, None; Gill, T, None.

SPONSOR'S ROLE

The American Federation for Aging Research (AFAR) provided organizational and administrative support for the 2021 Annual Clin-STAR Meeting. The recording of the sessions is available online at: https://vimeo.com/647347222.

ORCID

Jiha Lee https://orcid.org/0000-0002-6402-6563
Biren B. Kamdar https://orcid.org/0000-0002-9245-6229
Jaclyn Bergstrom https://orcid.org/0000-0002-0212-0802
Terrence E. Murphy https://orcid.org/0000-0002-9649-517X
Thomas M. Gill https://orcid.org/0000-0002-6450-0368

TWITTER

Jiha Lee @JihaRheum

REFERENCES

1. Recording of the 2021 Clin-STAR Annual Meeting session is available online at https://vimeo.com/647347222.
2. Broman KW, Woo KH. Data organization in spreadsheets. *Am Statis*. 2018;72:2-10.
3. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42:377-381.
4. Briney KA, Coates H, Goben A. Foundational practices of research data management. *Res Ideas Outcomes*. 2020;6:e56508. doi:10.3897/rio.6.e56508
5. NIH. Central resources for grants and funding information. NIH data sharing policy. https://grants.nih.gov/grants/policy/data_sharing/.

How to cite this article: Lee J, Kamdar BB, Bergstrom J, Murphy TE, Gill TM. Modeling success: How to work effectively with your biostatistician. *J Am Geriatr Soc*. 2022;70(8):2449-2454. doi:10.1111/jgs.17888