INVITED FEATURE: STATISTICS PRIMER

How to collaborate with a Biostatistician

Robin M Turner, Claire Cameron, Ari Samaranayaka

Introduction

Health researchers in various disciplines often seek advice from, or collaboration with, biostatisticians. This advice and collaboration can vary from guidance on a specific point through to extensive collaboration over many years across all aspects of the research process. Anecdotally, the level of support that health researchers request is often driven by the researchers' own beliefs about what biostatisticians do and what they can contribute. This does not always align with what biostatisticians actually do. In this article we summarise, from our own experience, the benefits of collaborating with academic biostatisticians by explaining what our jobs involve and how we collaborate best.

What is biostatistics?

Biostatistics is its own discipline. It involves applying statistical methods to health-related research and requires mixing statistical theory with unique knowledge of the application in health sciences. Statistics is the foundation of biostatistics, but additional knowledge is required to be a biostatistician. This often includes health-specific study design expertise, epidemiology, and an understanding of the source of the data. Other elements may be required depending on the health application. Being a good statistician is just one element of being a good biostatistician.

What do academic biostatisticians do?

As academic biostatisticians, our work is not restricted to statistical aspects alone. Rather, we have a particular interest in all quantitative elements of research that we are involved in. This includes translating research ideas to answerable research questions, designing studies to meet research aims given resource constraints, appropriately collecting data, producing analysis plans, analysing data, interpreting results within the context and limitations of the study, and generating new research questions. A diagram (Figure 1), published in a recent primer[1] and reproduced here with permission, shows the cycle that research projects often take and how the biostatistical elements are woven throughout the entire cycle. Alongside our collaborators, at the design stage, we often re-frame research questions to be answerable, refine study designs to incorporate important biostatistical elements, and tailor analysis plans to accommodate features specific to the study design. These aspects are often critical for successful applications for grant funding. After starting the study, we then assess how subsequent decisions can influence the analysis and conclusions, and we alert the research team to design changes that may undermine the integrity of the study. In other words, as experts in the application of statistical methods to health research, we provide methodological leadership. We have a large role to play in carrying out the analyses, ensuring accurate reporting of results, and responding to questions or criticism on the statistical analyses from reviewers. We also (usually) write the statistical analysis section and the interpretation of results
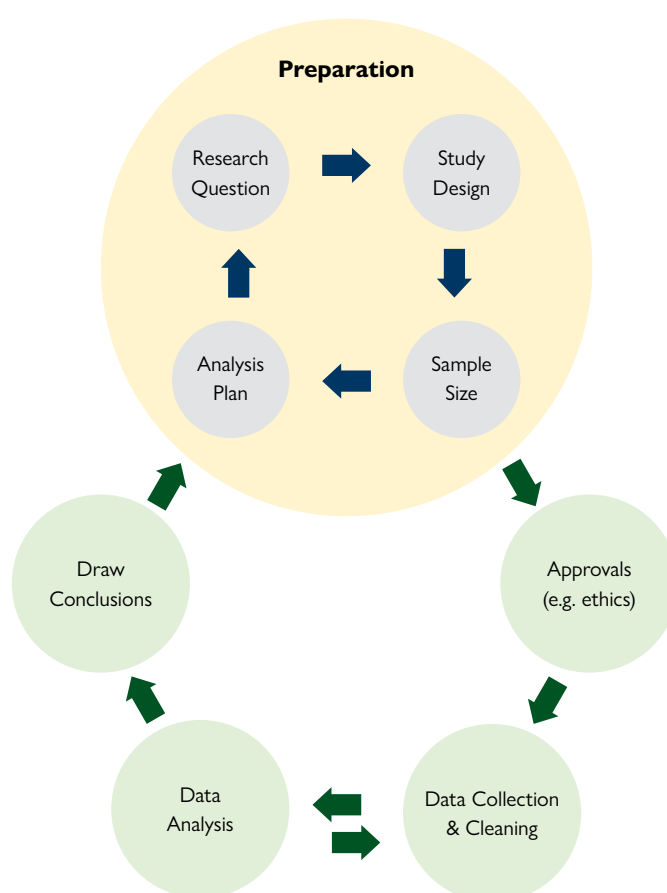

Figure 1: The research cycle with biostatistical elements woven throughout. Reproduced with permission of the authors.[1]

for journal articles and reports. Biostatisticians not only contribute to academia as active members of research teams; due to their unique ability to understand the methods needed for the whole research process from conceptualisation to dissemination, they also play important roles within grant funding and ethics committees.

We spend a significant (pun intended) amount of time doing analyses and all analyses take longer than anticipated. Every analysis requires careful checking, all methods must be justified for the specific context and applied correctly, and all assumptions need to be investigated. For these reasons, even "simple" analyses can take a lot of time. It goes with the territory that biostatisticians often, especially on smaller projects, end up managing the study data which also requires

considerable time. Data management can include transforming data into usable formats and types, looking for errors in the data and correcting them with agreement from the content experts, and deciding on appropriate ways of handling incomplete data. Despite often being tasked with data management, we would never claim to be data managers — that is a highly skilled occupation of its own.

Biostatisticians can also take on roles overseeing the analyses, instead of conducting the analyses themselves. This can be an uncomfortable arrangement unless there is an excellent working relationship and good communication within the team. When unexpected problems arise, which are common with real world health data, difficult conversations may need to occur. Occasionally, our role is purely advisory, where we provide guidance without being an active team member. This can work well when we have an established relationship with an experienced quantitative researcher. However, we do not have the same responsibility for the project if we are not collaborators, and it can be difficult to give sound advice without knowing the level of detail about the study that a collaborator would know. In our experience, using a biostatistician in an advisory capacity does not work well with postgraduate students. Most students need much more than just advice on the biostatistical elements; they usually need formal biostatistics supervision. If you are a postgraduate student and feel you need biostatistical guidance, our recommendation is to meet with a biostatistician (take your supervisor along as well) and discuss the options available, including whether a biostatistician is needed in your supervisory team.

At what stage of the research project should you involve a biostatistician?

The simplest answer to that question is: at the very beginning.

A common misconception is that biostatisticians only conduct data analysis and do not have a role prior to this. Therefore, some researchers only reach out once their data is collected. As we have discussed in a previous article,[1] the biostatistical threads are woven into the study design and research question from the very beginning. If the data are collected without input from a biostatistician, it may turn out that critical elements have been missed and that it is not possible to answer the intended research question. It is never too early to talk to a biostatistician.

Having said that, we do not turn people away if they reach out to us later in the research process. Biostatisticians are good at joining projects late with a sense of optimism that the research question can be answered. However, including a biostatistician will not be a quick fix as the biostatistician will need time to think through the statistical issues and the study design amongst other things. There are occasions where a biostatistician who joined the team after the data is collected determines that the study design is unsuitable for the research question and/or incompatible with the analysis plan that the researchers had in mind. This is a situation that we always try to avoid.

The data collection for a study is absolutely crucial. You need to be sure that you are collecting what you need in the form that you need it. That sounds obvious, but there are many examples of people collecting their data and finding once they have finished that they forgot to collect a critical piece of information, or that the data is collected in a form that requires a lot of extra work before it can be used. Ideally, a biostatistician would be involved before data collection starts. They can confirm what form the data should be in. The analysis plan should be considered at the study design stage, as this helps avoid a lot of data issues later on. Depending on the complexity of the study, you may also need to involve a data manager.

What can you expect when you first meet with a biostatistician?

In the first meeting, it is common to talk about both the research idea and how the collaboration will work. Some points you might want to discuss are listed below, although not all will be relevant if you have made contact at the beginning of your project.

SPECIFICS OF THE RESEARCH:
1. What is your research question?
2. What stage is the work at?
   a. Are you planning the study?
   b. Have you started collecting data?
3. Do you have a study design in mind?
4. Who will your participants be?
   a. How will you choose them?
   b. Will you be collecting information directly from them?
5. How will the data be collected?
   a. Does the data already exist (for example, administrative data) or do you need to collect it?
6. What information do you want to collect and at what time points?
7. What are the possible practical limitations for your project?

HOW THE COLLABORATION MAY WORK:
1. Do you have a background in statistics (this sets the scene for the conversation)?
2. What role do you see for a biostatistician in your project?
   a. Will they be a collaborator, a supervisor, or an advisor?
   b. How much contact will they have with the research group?
3. What is the intended authorship of any planned journal publication (this relates to the intended role in 2a)?

How does good communication fit in?

A biostatistician once pointed out (informally) that researchers (non-biostatisticians) think in ideas, whereas biostatisticians think in variables. When we hear researchers' ideas and research questions, our (biostatistical) minds go instantly to: What will be measured? How do we summarise those measurements? How do we assess those associations? What is happening with the variability? How does the study design impact these measurements? And so on. Because of this, when biostatisticians become involved in a project, their perspective can be seen as being very detail-driven. For example, it may feel like we are fussing over the wording of the research question, but from our point of view, the wording of the research question drives the entire study design and analysis, and therefore needs to be very precise and is well worth fussing over. Good communication is critical to weave together the different perspectives of researchers on a project. Research teams must be able to communicate from their own areas of expertise to support good decision making that respects the integrity of the research from all sides. It can feel at times as if biostatisticians and non-biostatisticians are speaking different languages. The reality is that everyone taking part in health research is highly skilled, highly qualified, and experienced in their own areas. Everyone in a research team needs to be heard and their point of view considered. A keenly observant biostatistical eye can save a project from disaster.

There is something you can do to improve your statistical communication (if you have not done so already): consider taking an introductory biostatistics course. The more biostatistical knowledge researchers have, the better their understanding of quantitative research methods, and the easier it is to have complex discussions over the biostatistical elements of a project. This ultimately can lead to better research with better integration across content and methodological aspects of the study.

But I just have a quick question?

We understand why people think they have "quick questions". For example, "I just want to compare these two groups, should I use a t-test?" sounds like a very simple question. In reality, the answer is never quick, and to answer that properly, we need to know:

> What has been measured (is a mean appropriate)?
> How was the data collected (in detail)?
> Are the groups independent, and observations within groups independent?
> What are the assumptions for the method in question, and will they be met given the study design and measures?

> What important information might be missing? For example, if measurements are from a scale, how was the scale created and what sort of answers are possible?
> Are there people with missed measurements or those with missed components in the measurement?

A seemingly quick question is never quick to answer, because of the amount of detail required to give the correct answer. Instead of approaching biostatisticians with "quick questions", researchers should invest time in building up a good collaborative relationship with a biostatistician at the very beginning of their research.

Conclusion

Hopefully, this article clarifies some of the mysteries surrounding biostatisticians' involvement in research. We see a need for more involvement of biostatisticians in health-related research. We can only guess why researchers can be reluctant to include biostatisticians in their research teams, but we suspect that the following points may have a role to play:

1. Many people think biostatisticians are people who only analyse data, so there is no need to include them until after data collection
2. Many people have negative feelings about mathematics and statistics, and the idea of talking to a biostatistician may be daunting
3. Many people have learnt some statistics as part of their degree and believe that is enough to see them through the research project
4. Biostatisticians always seem so busy, there are not very many around, and it may appear difficult to involve them

We have shown that (1) is not the case. To (2) we say, please come and see us, we understand that some people love numbers and some people struggle with numbers. Our work requires good communication, so we can talk about your project and its biostatistical components at a level tailored to you. To (3) we say, having done some statistics is very helpful. However, real-world research is very different to situations covered in pure academic courses, and not comparable to the wealth of experience biostatisticians bring to a project. If you do have it covered, great, but if you are not sure, please come and talk to us. Finally, to point (4), we say, you make a good point! We agree there probably aren't enough of us, but have a look at your institution's website, as there may be a biostatistics group who you can contact.

We hope that sharing this insight into what we do and how we like to work can help improve collaborations and the quality of health-related research.

References

1. Samaranayaka A, Cameron C, Turner R. Sample size in health research. New Zealand Medical Student Journal. 2021;32:52-4. doi: https://doi.org/10.57129/XPSC9819

About the authors

> Professor Robin Turner, BSc(Hons), MBiostat, PhD, is the Director of the Biostatistics Centre, Division of Health Sciences, University of Otago.
> Associate Professor Claire Cameron, BSc(Hons), DipGrad, MSc, PhD, is a Biostatistician in the Biostatistics Centre, Division of Health Sciences, University of Otago.
> Dr Ari Samaranayaka, BSc, MPhil, PhD, is a Senior Research Fellow and Biostatistician in the Biostatistics Centre, Division of Health Sciences, University of Otago.