

When do you need to involve a statistician?

Once you have developed your research question and carefully thought about the methodology it is time to consult a biostatistician. You might think this is too early, as you have no data but the biostatistician will need to help you with the study design, a data analysis plan and calculating sample size for your study.

It is very distressing if after all your efforts of collecting the data you find that your results are not able to fully address your research question due to a shortcoming in your design or are inconclusive because your sample size was too small. These are important reasons why you consult before finalising your proposal and embarking on your study.

Even if your study is a descriptive study, an audit or a case series it is still worthwhile discussing the study with a biostatistician. For these studies we normally use descriptive analysis but doing comparisons within the studies enriches these studies. This is where the biostatistician will help.

Statistical programs are widely available for doing a sample size calculation or analysis. You can certainly use these but getting advice from a statistician will help you to avoid studies that are too small or too large, incorrect statistical analysis and embarrassment when your article is reviewed.

TIP: To ensure that you can have a good discussion with the biostatistician, read through studies that have done something similar to what you want to do and give good attention to the study design, samples size and analysis that those researchers have used.

TIP: You and the biostatistician are both novices! You are a novice since you are not a statistician and the statistician is a novice since he is not a clinician or basic researcher. Do not confront the statistician with the most technical stuff – keep it basic from him/her to get an understanding of what it is you want to do. Who, what where, when and how are you going to measure and obtain your data? The consultation between you and the statistician is a process to understand what each one of you is talking about.

TIP: To understand some of the concepts the statistician will be talking about the basic vocabulary (Appendix 4) is a useful resource. Print out and read the glossary of statistical terms prior to meeting the biostatistician.
Link to statistical glossary
<http://www.csm-oxford.org.uk/statistical-resources/statistical-glossary/>

When you go and see the biostatistician you should have a clear idea of the following:

1. The study design you intend to use. The biostatistician will be able to help you with the study design if you are uncertain.
2. The level of significance you require. This is normally set at $p < 0.05$
3. The power required. This is termed a type II error and is normally set by convention at 90%.
4. The magnitude of the difference you expect between the 2 groups you are investigating. This can be very frustrating as this difference is often not known and this lack of knowledge is the reason for you doing the research. You will then have to use the published data or your clinical experience to make the educated guess. Often you mentor will be able to help you with this aspect. If you are investigating the difference in the complication rate that arises post operatively in patients that are HIV infected compared to those uninfected, the literature might indicate that the difference is 3 fold greater while your clinical experience indicates a 7 fold difference. With the aid of the biostatistician you might agree to set the magnitude of difference at 5 fold. (the larger the difference the smaller the sample size required)
5. Some idea of your analysis plan. What are the primary and secondary outcomes and what variables do you intend to collect?
6. What database do you intend to use. Some databases are easily compatible. You will save a lot of time if your database is compatible with the database used by the biostatistician.

Once you have discussed the primary outcome you might then discuss if you will be able to show a statistical difference for your secondary outcomes. The statistician might indicate that you have to enlarge your sample size to ensure statistical significance for secondary outcomes. You might then decide whether this is feasible or not.

In some cases it might be necessary to do a pilot study to collect some data to inform the calculation of the sample size. A pilot study can also be valuable in suggesting additional variables that should be collected.

TIP: See the biostatistician early. You cannot fix a flawed study after you have completed it.

You will need to visit the biostatistician again during the data analysis step.

What do you need for before you revisit the biostatistician?

1. Clean your data to make sure your data is correct and that you do not have missing data or incorrect data. (See STEP 12)
2. Your final analysis plan.
3. What help you will require from the biostatistician when you write up/report you study. You may want him/her to help you with a specific analysis that might not be expected from you as a non-statistician.
4. Your expected timelines. You cannot expect the statistician to perform a miracle a day before you have to submit some report or conference abstract.



TIP:

Always have a look at a summary of your raw data before you visit the biostatistician. You might see some interesting findings.

Once you have completed your first few drafts of your article remember to involve the biostatistician again to ensure that the description of the statistical analysis is correct. Remember the biostatistician needs to be included as an author on the paper or acknowledged in the paper for his/her help.

Gie, R., & Beyers, N. (2014). Getting started in clinical research: Guidance for junior researchers. Cape Town: Department of Paediatrics and Child Health, Faculty of Medicine and Health Sciences, Stellenbosch University.