than you were at the beginning. Whatever the reason, this means that if you do all the measurements on one treatment group first and then all those of another treatment group, you risk introducing systematic differences between the groups because of changes in the accuracy of the methods. It is far better to organize your sampling procedure so that individuals are measured in a random order (see Chapter 5 for more about this sort of problem).

Avoid introducing time of measurement as a confounding factor.

3.4.6 Random samples and representative samples

Imagine you ask a computer program to randomly pick five unique random numbers between 1 and 100. It is possible that it will randomly generate something that looks non-random, say the set of numbers {1, 2, 3, 4, and 5}. This is very unlikely, but will happen occasionally. What should you do? If you think that this will affect the quality of your results, then you should discard this set of random numbers and ask the computer to try again. For example, imagine that over the years you have written 100 reports, and you want to measure the occurrence of the word 'however' in these. Taking a random sample of the 100 reports seems a good idea, but if that sample is numbers 1 to 5, then you are only sampling reports that you wrote very early in your career. This seems less good; hence you should ask the computer to try again. This is not cheating. One way to think about this is that your aim is not to get a random sample, but a representative sample. Almost all random samples are representative, but some are not. But be careful that you don't cheat. If you carry out statistical analyses and get a result that you find inconvenient, and you then conclude that this was because your random sample was unrepresentative, then we would be very uncomfortable about you discarding this experiment and trying again with another random sample. Ask yourself: would you also have decided that the sample was unrepresentative if the results of the statistical analyses were different? If the answer is 'no', then you are rejecting inconvenient results on spurious grounds by deciding that the sample was unrepresentative. Decide whether or not to reject a random sample immediately, generally before you have looked at the data from that sample. This stops any suggestion of cheating. Let us return to the case of the 100 reports. If the sample had been numbers 20, 40, 60, 80, and 100, this also looks odd, but is it unrepresentative? Probably not, as long as there is no regularity in the sort of reports you write. However, if every fifth report is a summary of the preceding four, then you do have a problem, as you are only sampling summary reports and not the original

sources, which make up 80% of the reports you write. In this case 20, 40, 60, 80, and 100 would be an unrepresentative sample.

Ideally, your decision to reject random samples as unrepresentative should be based on a well-defined rule that was decided in advance. However, unrepresentative random samples occur so infrequently that very few scientists take the time to think up such rules that will very rarely come into operation. You can be certain that the Devil's advocate will take a very strong interest in any time you decide that a random sample is unrepresentative, so when you do discard a sample, make sure that you have a very strong case for justifying this decision.

You should discard random samples as unrepresentative only very infrequently and after very careful thought.

3.5 Selecting the appropriate number of replicates

Replication is the basis of all experimental design, and a natural question that arises in any study is: how many replicates do we need? As we saw earlier, the more replicates we have, the more confident we can be that differences between groups are real and not simply due to chance effects. So, all things being equal, we want as many replicates as possible. However, as with all things in life, all things are not equal. Increasing replication incurs costs. These costs might be financial: if an experiment involves expensive chemical reagents then doubling the number of replicates will result in a large increase in cost. More likely (and more importantly for many people), experiments will involve time costs.

Probably most important of all, if experiments involve using humans or animals (or materials obtained from humans or animals), then there may well be welfare or conservation costs of increasing sample sizes.

Q 3.9 You are planning a study on wild birds which will involve catching them, which is a potentially stressful experience for the birds. A colleague tells you that you should limit your study to 10 birds to minimize animal suffering. Do you agree?

These costs mean that the general answer to the question 'How big should my experiment be?' is that it should be big enough to give you confidence that you will be able to detect any biologically meaningful effects that exist, but not so big that some sampling was unnecessary. How can we decide how big this actually is? There are two general approaches that can be used to answer this question: we can either make educated guesses based on similar studies or carry out formal power analysis.

'The more samples, the better' is an over-simplification.