

## Quantitative Research Design

### Will G Hopkins PhD

Department of Physiology and School of Physical Education,  
University of Otago, Dunedin, New Zealand 9001

Email: [will.hopkins=AT=otago.ac.nz](mailto:will.hopkins=AT=otago.ac.nz)

Sportscience 4(1), [sportsci.org/jour/0001/wghdesign.html](http://sportsci.org/jour/0001/wghdesign.html), 2000 (4318 words)

Reviewed by Greg Atkinson PhD, Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, Liverpool, UK

In quantitative research your aim is to determine the relationship between one thing (an independent variable) and another (a dependent or outcome variable) in a population. Quantitative research designs are either descriptive (subjects usually measured once) or experimental (subjects measured before and after a treatment). A descriptive study establishes only associations between variables. An experiment establishes causality.

For an accurate estimate of the relationship between variables, a descriptive study usually needs a sample of hundreds or even thousands of subjects; an experiment, especially a crossover, may need only tens of subjects. The estimate of the relationship is less likely to be biased if you have a high participation rate in a sample selected randomly from a population. In experiments, bias is also less likely if subjects are randomly assigned to treatments, and if subjects and researchers are blind to the identity of the treatments.

In all studies, subject characteristics can affect the relationship you are investigating. Limit their effect either by using a less heterogeneous sample of subjects or preferably by measuring the characteristics and including them in the analysis. In an experiment, try to measure variables that might explain the mechanism of the treatment. In an unblinded experiment, such variables can help define the magnitude of any placebo effect.

[Reprint](#) · [Help](#)

**KEYWORDS:** controlled trial, crossover, descriptive, experimental, mechanism, placebo effect, sample size

[Types of Study](#)

[Samples](#)

[Sample Size](#)

[What to Measure](#)

Quantitative research is all about quantifying relationships between variables. Variables are things like weight, performance, time, and treatment. You measure variables on a sample of subjects, which can be

tissues, cells, animals, or humans. You express the relationship between variable using effect statistics, such as correlations, relative frequencies, or differences between means. I deal with these statistics and other aspects of analysis [elsewhere at this site](#). In this article I focus on the design of quantitative research. First I describe the types of study you can use. Next I discuss how the nature of the sample affects your ability to make statements about the relationship in the population. I then deal with various ways to work out the size of the sample. Finally I give advice about the kinds of variable you need to measure.

## TYPES OF STUDY

Studies aimed at quantifying relationships are of two types: **descriptive** and **experimental** (Table 1). In a descriptive study, no attempt is made to change behavior or conditions--you measure things as they are. In an experimental study you take measurements, try some sort of intervention, then take measurements again to see what happened.

Table 1: Types of research design	
<b>Descriptive or observational</b>	
<ul style="list-style-type: none"> <li>● case</li> <li>● case series</li> <li>● cross-sectional</li> <li>● cohort or prospective or longitudinal</li> <li>● case-control or retrospective</li> </ul>	
<b>Experimental or longitudinal or repeated-measures</b>	
<ul style="list-style-type: none"> <li>● without a control group               <ul style="list-style-type: none"> <li>time series</li> <li>crossover</li> </ul> </li> <li>● with a control group</li> </ul>	

## Descriptive Studies

Descriptive studies are also called **observational**, because you observe the subjects without otherwise intervening. The simplest descriptive study is a **case**, which reports data on only one subject; examples are a study of an outstanding athlete or of a dysfunctional institution. Descriptive studies of a few cases are called **case series**. In **cross-sectional** studies variables of interest in a sample of subjects are assayed once and the relationships between them are determined. In **prospective** or **cohort** studies, some variables are assayed at the start of a study (e.g., dietary habits), then after a period of time the outcomes are determined (e.g., incidence of heart disease). Another label for this kind of study is **longitudinal**, although this term also applies to experiments. **Case-control** studies compare **cases** (subjects with a particular attribute, such as an injury or ability) with **controls** (subjects without the attribute);

comparison is made of the **exposure** to something suspected of causing the cases, for example volume of high intensity training, or number of alcoholic drinks consumed per day. Case-control studies are also called **retrospective**, because they focus on conditions in the past that might have caused subjects to become cases rather than controls.

A common case-control design in the exercise science literature is a comparison of the behavioral, psychological or anthropometric characteristics of elite and sub-elite athletes: you are interested in what the elite athletes have been exposed to that makes them better than the sub-elites. Another type of study compares athletes with sedentary people on some outcome such as an injury, disease, or disease risk factor. Here you know the difference in exposure (training vs no training), so these studies are really cohort or prospective, even though the exposure data are gathered retrospectively at only one time point. The technical name for these studies is **historical cohort**.

## Experimental Studies

Experimental studies are also known as **longitudinal** or **repeated-measures** studies, for obvious reasons. They are also referred to as **interventions**, because you do more than just observe the subjects.

In the simplest experiment, a **time series**, one or more measurements are taken on all subjects before and after a treatment. A special case of the time series is the so-called **single-subject design**, in which measurements are taken repeatedly (e.g., 10 times) before and after an intervention on one or a few subjects.

Time series suffer from a major problem: any change you see could be due to something other than the treatment. For example, subjects might do better on the second test because of their experience of the first test, or they might change their diet between tests because of a change in weather, and diet could affect their performance of the test. The **crossover** design is one solution to this problem. Normally the subjects are given two treatments, one being the real treatment, the other a control or reference treatment. Half the subjects receive the real treatment first, the other half the control first. After a period of time sufficient to allow any treatment effect to wash out, the treatments are crossed over. Any effect of retesting or of anything that happened between the tests can then be subtracted out by an appropriate analysis. **Multiple crossover** designs involving several treatments are also possible.

If the treatment effect is unlikely to wash out between measurements, a **control group** has to be used. In these designs, all subjects are measured, but only some of them--the **experimental group**--then receive the treatment. All subjects are then measured again, and the change in the experimental group is compared with the change in the control group.

If the subjects are assigned randomly to experimental and control groups or treatments, the design is known as a **randomized controlled trial**. Random assignment minimizes the chance that either group is not typical of the population. If the subjects are **blind** (or **masked**) to the identity of the treatment, the design is a **single-blind** controlled trial. The control or reference treatment in such a study is called a

**placebo**: the name physicians use for inactive pills or treatments that are given to patients in the guise of effective treatments. Blinding of subjects eliminates the **placebo effect**, whereby people react differently to a treatment if they think it is in some way special. In a **double-blind** study, the experimenter also does not know which treatment the subjects receive until all measurements are taken. Blinding of the experimenter is important to stop him or her treating subjects in one group differently from those in another. In the best studies even the data are analyzed blind, to prevent conscious or unconscious fudging or prejudiced interpretation.

Ethical considerations or lack of cooperation (compliance) by the subjects sometimes prevent experiments from being performed. For example, a randomized controlled trial of the effects of physical activity on heart disease may not have been performed yet, because it is unethical and unrealistic to randomize people to 10 years of exercise or sloth. But there have been many short-term studies of the effects of physical activity on disease risk factors (e.g., blood pressure).

## Quality of Designs

The various designs differ in the quality of evidence they provide for a cause-and-effect relationship between variables. Cases and case series are the weakest. A well-designed cross-sectional or case-control study can provide good evidence for the **absence** of a relationship. But if such a study does reveal a relationship, it generally represents only suggestive evidence of a causal connection. A cross-sectional or case-control study is therefore a good starting point to decide whether it is worth proceeding to better designs. Prospective studies are more difficult and time-consuming to perform, but they produce more convincing conclusions about cause and effect. Experimental studies provide the best evidence about how something affects something else, and double-blind randomized controlled trials are the best experiments.

**Confounding** is a potential problem in descriptive studies that try to establish cause and effect. Confounding occurs when part or all of a significant association between two variables arises through both being causally associated with a third variable. For example, in a population study you could easily show a negative association between habitual activity and most forms of degenerative disease. But older people are less active, and older people are more diseased, so you're bound to find an association between activity and disease without one necessarily causing the other. To get over this problem you have to **control for potential confounding factors**. For example, you make sure all your subjects are the same age, or you include age in the analysis to try to remove its effect on the relationship between the other two variables.

## SAMPLES

You almost always have to work with a **sample** of subjects rather than the full **population**. But people are interested in the population, not your sample. To **generalize** from the sample to the population, the sample has to be **representative** of the population. The safest way to ensure that it is representative is to use a **random** selection procedure. You can also use a **stratified** random sampling procedure, to make

sure that you have proportional representation of population subgroups (e.g., sexes, races, regions).

When the sample is not representative of the population, **selection bias** is a possibility. A statistic is biased if the value of the statistic tends to be wrong (or more precisely, if the expected value--the average value from many samples drawn using the same sampling method--is not the same as the population value.) A typical source of bias in population studies is age or socioeconomic status: people with extreme values for these variables tend not to take part in the studies. Thus a high **compliance** (the proportion of people contacted who end up as subjects) is important in avoiding bias. Journal editors are usually happy with compliance rates of at least 70%.

Failure to randomize subjects to control and treatment groups in experiments can also produce bias. If you let people select themselves into the groups, or if you select the groups in any way that makes one group different from another, then any result you get might reflect the group difference rather than an effect of the treatment. For this reason, it's important to randomly assign subjects in a way that ensures the groups are **balanced** in terms of important variables that could modify the effect of the treatment (e. g., age, gender, physical performance). Human subjects may not be happy about being randomized, so you need to state clearly that it is a condition of taking part.

Often the most important variable to balance is the pre-test value of the dependent variable itself. You can get close to perfectly balanced randomization for this or another numeric variable as follows: rank-order the subjects on the value of the variable; split the list up into pairs (or triplets for three treatments, etc.); assign the lowest ranked subject to a treatment by flipping a coin; assign the next two subjects (the other member of the pair, and the first member of the next pair) to the other treatment; assign the next two subjects to the first treatment, and so on. If you have male and female subjects, or any other grouping that you think might affect the treatment, perform this randomization process for each group ranked separately. Data from such pair-matched studies can be analyzed in ways that may increase the precision of the estimate of the treatment effect. Watch this space for an update shortly.

When selecting subjects and designing protocols for experiments, researchers often strive to eliminate all variation in subject characteristics and behaviors. Their aim is to get greater precision in the estimate of the effect of the treatment. The problem with this approach is that the effect generalizes only to subjects with the same narrow range of characteristics and behaviors as in the sample. Depending on the nature of the study, you may therefore have to strike a balance between precision and applicability. If you lean towards applicability, your subjects will vary substantially on some characteristic or behavior that you should measure and include in your analysis. See [below](#).

## **SAMPLE SIZE**

How many subjects should you study? You can approach this crucial issue via statistical significance, confidence intervals, or "on the fly".

### **Via Statistical Significance**

[Statistical significance](#) is the standard but somewhat complicated approach. Your sample size has to be big enough for you to be sure you will detect the smallest worthwhile effect or relationship between your variables. *To be sure* means detecting the effect 80% of the time. *Detect* means getting a statistically significant effect, which means that more than 95% of the time you'd expect to see a value for the effect numerically smaller than what you observed, if there was no effect at all in the population (in other words, the p value for the effect has to be less than 0.05). *Smallest worthwhile effect* means the smallest effect that would make a difference to the lives of your subjects or to your interpretation of whatever you are studying. If you have too few subjects in your study and you get a statistically significant effect, most people regard your finding as publishable. But if the effect is not significant with a small sample size, most people regard it (erroneously) as unpublishable.

## Via Confidence Intervals

Using [confidence intervals](#) or confidence limits is a more accessible approach to sample-size estimation and interpretation of outcomes. You simply want enough subjects to give acceptable precision for the effect you are studying. *Precision* refers usually to a 95% confidence interval for the true value of the effect: the range within which the true (population) value for the effect is 95% likely to fall. *Acceptable* means it won't matter to your subjects (or to your interpretation of whatever you are studying) if the true value of the effect is as large as the upper limit or as small as the lower limit. A bonus of using confidence intervals to justify your choice of sample size is that the sample size is about half what you need if you use statistical significance.

## "On the Fly"

An acceptable width for the confidence interval depends on the magnitude of the observed effect. If the observed effect is close to zero, the confidence interval has to be narrow, to exclude the possibility that the true (population) value could be substantially positive or substantially negative. If the observed effect is large, the confidence interval can be wider, because the true value of the effect is still large at either end of the confidence interval. I therefore recommend getting your [sample size on the fly](#): start a study with a small sample size, then increase the number of subjects until you get a confidence interval that is appropriate for the magnitude of the effect that you end up with. I have run simulations to show the resulting magnitudes of effects are not substantially biased.

## Effect of Research Design

The type of design you choose for your study has a major impact on the sample size. Descriptive studies need hundreds of subjects to give acceptable confidence intervals (or to ensure statistical significance) for small effects. Experiments generally need a lot less--often one-tenth as many--because it's easier to see changes within subjects than differences between groups of subjects. Crossovers need even less--one-quarter of the number for an equivalent trial with a control group--because every subject gets the experimental treatment. I give [details](#) on the stats pages at this site.

## Effect of Validity and Reliability

The precision with which you measure things also has a major impact on sample size: the worse your measurements, the more subjects you need to lift the signal (the effect) out of the noise (the errors in measurement). Precision is expressed as **validity** and **reliability**. Validity represents how well a variable measures what it is supposed to. Validity is important in descriptive studies: if the validity of the main variables is poor, you may need thousands rather than hundreds of subjects. Reliability tells you how reproducible your measures are on a retest, so it impacts experimental studies: the more reliable a measure, the less subjects you need to see a small change in the measure. For example, a controlled trial with 20 subjects in each group or a crossover with 10 subjects may be sufficient to characterize even a small effect, if the measure is highly reliable. See the [details](#) on the stats pages.

## Pilot Studies

As a student researcher, you might not have enough time or resources to get a sample of optimum size. Your study can nevertheless be a **pilot** for a larger study. Perform a pilot study to develop, adapt, or check the feasibility of techniques, to determine the reliability of measures, and/or to calculate how big the final sample needs to be. In the latter case, the pilot should have the same sampling procedure and techniques as in the larger study.

For experimental designs, a pilot study can consist of the first 10 or so observations of a larger study. If you get respectable confidence limits, there may be no point in continuing to a larger sample. Publish and move on to the next project or lab!

If you can't test enough subjects to get an acceptably narrow confidence interval, you should still be able to publish your finding, because your study will set useful bounds on how big and how small the effect can be. A statistician can also combine your finding with the findings of similar studies in something called a **meta-analysis**, which derives a confidence interval for the effect from several studies. If your study is not published, it can't contribute to the meta-analysis! Many reviewers and editors do not appreciate this important point, because they are locked into thinking that only statistically significant results are publishable.

## WHAT TO MEASURE

In any study, you measure the **characteristics of the subjects**, and the **independent and dependent variables** defining the research question. For experiments, you can also measure **mechanism variables**, which help you explain how the treatment works.

## Characteristics of Subjects

You must report sufficient information about your subjects to identify the population group from which

they were drawn. For human subjects, variables such as sex, age, height, weight, socioeconomic status, and ethnic origin are common, depending on the focus of the study.

Show the ability of athletic subjects as current or personal-best performance, preferably expressed as a percent of world-record. For endurance athletes a direct or indirect estimate of maximum oxygen consumption helps characterize ability in a manner that is largely independent of the sport.

## Dependent and Independent Variables

Usually you have a good idea of the question you want to answer. That question defines the main variables to measure. For example, if you are interested in enhancing sprint performance, your **dependent variable** (or **outcome** variable) is automatically some measure of sprint performance. Cast around for the way to measure this dependent variable with as much precision as possible.

Next, identify all the things that could affect the dependent variable. These things are the **independent variables**: training, sex, the treatment in an experimental study, and so on.

For a descriptive study with a wide focus (a "fishing expedition"), your main interest is estimating the effect of everything that is likely to affect the dependent variable, so you include as many independent variables as resources allow. For the large sample sizes that you should use in a descriptive study, including these variables does not lead to substantial loss of precision in the effect statistics, but beware: the more effects you look for, the more likely the true value of at least one of them lies outside its confidence interval (a problem I call [cumulative Type 0 error](#)). For a descriptive study with a narrower focus (e.g., the relationship between training and performance), you still measure variables likely to be associated with the outcome variable (e.g., age-group, sex, competitive status), because either you restrict the sample to a particular subgroup defined by these variables (e.g., veteran male elite athletes) or you include the variables in the analysis.

For an experimental study, the main independent variable is the one indicating when the dependent variable is measured (e.g., before, during, and after the treatment). If there is a control group (as in controlled trials) or control treatment (as in crossovers), the identity of the group or treatment is another essential independent variable (e.g., Drug A, Drug B, placebo in a controlled trial; drug-first and placebo-first in a crossover). These variables obviously have an affect on the dependent variable, so you automatically include them in any analysis.

Variables such as sex, age, diet, training status, and variables from blood or exercise tests can also affect the outcome in an experiment. For example, the response of males to the treatment might be different from that of females. Such variables account for **individual differences** in the response to the treatment, so it's important to take them into account. As for descriptive studies, either you restrict the study to one sex, one age, and so on, or you sample both sexes, various ages, and so on, then analyze the data with these variables included as **covariates**. I favor the latter approach, because it widens the applicability of your findings, but once again there is the problem of cumulative Type 0 error for the effect of these

covariates. An additional problem with small sample sizes is loss of precision of the estimate of the effect, if you include more than two or three of these variables in the analysis.

## Mechanism Variables

With experiments, the main challenge is to determine the magnitude and confidence intervals of the treatment effect. But sometimes you want to know the **mechanism** of the treatment--how the treatment works or doesn't work. To address this issue, try to find one or more variables that might connect the treatment to the outcome variable, and measure these at the same times as the dependent variable. For example, you might want to determine whether a particular training method enhanced strength by increasing muscle mass, so you might measure limb girths at the same time as the strength tests. When you analyze the data, look for associations between change in limb girth and change in strength. Keep in mind that errors of measurement will tend to obscure the true association.

This kind of approach to mechanisms is effectively a descriptive study on the difference scores of the variables, so it can provide only suggestive evidence for or against a particular mechanism. To understand this point, think about the example of the limb girths and strength: an increase in muscle size does not necessarily cause an increase in strength--other changes that you haven't measured might have done that. To really nail a mechanism, you have to devise another experiment aimed at changing the putative mechanism variable while you control everything else. But that's another research project. Meanwhile, it is sensible to use your current experiment to find suggestive evidence of a mechanism, provided it doesn't entail too much extra work or expense. And if it's research for a PhD, you are expected to measure one or more mechanism variables and discuss intelligently what the data mean.

Finally, a useful application for mechanism variables: they can define the magnitude of **placebo effects** in unblinded experiments. In such experiments, there is always a doubt that any treatment effect can be partly or wholly a placebo effect. But if you find a correlation between the change in the dependent variable and change in an *objective* mechanism variable--one that cannot be affected by the psychological state of the subject--then you can say for sure that the treatment effect is not all placebo. And the stronger the correlation, the smaller the placebo effect. The method works only if there are individual differences in the response to the treatment, because you can't get a correlation if every subject has the same change in the dependent variable. (Keep in mind that some apparent variability in the response between subjects is likely to be random error in the dependent variable, rather than true individual differences in the response to the treatment.)

Surprisingly, the objective variable can be almost anything, provided the subject is unaware of any change in it. In our example of strength training, limb girth is not a good variable to exclude a placebo effect: subjects may have noticed their muscles get bigger, so they may have expected to do better in a strength test. In fact, any noticeable changes could inspire a placebo effect, so any objective variables that correlate with the noticeable change won't be useful to exclude a placebo effect. Think about it. But if the subjects noticed nothing other than a change in strength, and you found an association between change in blood lipids, say, and change in strength, then the change in strength cannot all be a placebo

effect. Unless, of course, changes in blood lipids are related to susceptibility to suggestion...unlikely, don't you think?

---

[editor=AT=sportsci.org](#) · [Homepage](#) · ©2000

Webmastered by Will Hopkins

Uploaded 4 May 00

Updated extensively from an [earlier version](#) (1998)