**WHICH QUANTITATIVE RESEARCH DESIGN SHOULD I USE?**

**by Simon Moss**

|  |
| --- |
| **Introduction** |

Suppose you plan to conduct a quantitative study—a study in which you analyse numbers. For example, you might want to explore whether exercise enhances the motivation of PhD candidates, as measured on a 10 point scale. How would you conduct this study? Perhaps you could administer a questionnaire that assesses the exercise routine and motivation of PhD candidates. Perhaps you could instruct half the candidates to exercise more rigorously than usual and then test their motivation, and so forth. Indeed, you could apply a variety of research designs.

This document delineates the main research designs. Furthermore, this document provides some insight into the benefits and drawbacks of each design. Finally, for each design, this document offers some recommendations on which statistical tests you should conduct.

|  |
| --- |
| **Are you comparing distinct groups?** |

**Study that compares groups**

First, you need to decide whether you want to compare groups in your study. To demonstrates, if you wanted to examine whether exercise enhances motivation in PhD candidates

* 50 candidates could be instructed to exercise vigorously for a month, 3 times a week
* 50 candidates could be instructed to refrain from exercise for a month
* after this month, candidates could be asked to indicate how motivated they feel at university, on a rating scale that ranges from 1 to 10.

As this figure shows, some participants exercised for a month. Other participants did not exercise for a month. The researcher thus wants to compare these two groups—sometimes referred to as two conditions—on their motivation.

|  |
| --- |
| Exercise for one month, then indicate level of motivation |
| Image result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette person |
| Refrain from exercise for one month, then indicate level of motivation |

**Studies that do not compare groups**

Alternatively, if you wanted to examine whether exercise enhances motivation in PhD candidates

* all candidates could complete a survey
* during the survey, the candidates would indicate how many hours a week they exercise and how motivated they feel
* the researcher could then explore whether these two numerical variables—exercise and motivation—are related to each other.
* that is, do the candidates who exercise over a longer duration tend to be more motivated?

As this figure shows, all participants complete the same survey. This study does not compare groups. This design is sometimes called a correlational design.

|  |
| --- |
| Complete a survey that assesses exercise duration and motivation |
| Image result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette person |

**Studies that compare pre-existing groups**

Admittedly, in some instances, the distinction between studies that compare groups and studies that do not compare groups is subtle and arbitrary. For example, suppose you administered the previous survey—a survey that assesses duration of exercise and motivation in PhD candidates. After participants completed the survey

* you could divide these individuals into two groups: people who exercise and people who do not exercise at all.
* you could then compare these competing groups.

As this figure shows, all participants complete the same survey. In this sense, the study was not designed to compare groups. But, the researcher eventually compared two groups on motivation. In this sense, the study does compare two groups. For the purposes of this document, we will assume this study is a special example of the category of studies that compares groups—but groups that differ on some pre-existing characteristic, such as whether or not the individuals exercise.

|  |
| --- |
| Complete a survey that assesses exercise duration and motivation |
| Image result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette person |

|  |
| --- |
| **Between-subject, within-subject, and mixed designs.** |

**Between-subject design**

If you are comparing groups, you can choose one of three alternatives, sometimes called between-subject, within-subject, and mixed designs. To illustrate, in the following example, the two groups or conditions—participants who exercised and participants who did not exercise—were not the same people. This design is called *between-subject* or sometimes independent samples. Likewise, researchers might compare two groups of animals, specimens, organizations, and many other units.

|  |
| --- |
| Exercise for one month, then indicate level of motivation |
| Image result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette person |
| Refrain from exercise for one month, then indicate level of motivation |

**Within-subject design**

In contrast, researchers could

* recruit a set of PhD candidates who seldom exercise and measure their motivation
* ask these PhD candidates to exercise for a month
* then measure their motivation again

In essence as the following figure shows, the researcher wants to compare two groups: people who have not exercised and people who have exercised. In this example, the two groups comprise the same people. This design is called *within-subject—*because the researcher wants to compare the motivation within the same people or subjects. Likewise, researchers could examine animals, specimens, organizations, nations, and other units across time. Sometimes, this design is also called dependent samples, matches samples, or repeated measures. In addition, assessing people over time is referred to as longitudinal; assessing all people at one time is referred to as cross-sectional.

|  |
| --- |
| Complete a second survey that assesses exercise duration and motivation |
| Instruct participants to exercise for one month |
| Instruct participants to exercise for one month |
|  |
| Complete a survey that assesses exercise duration and motivation |
| Image result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette person |

Researchers have developed many variations of this design. However, the most common variant—in which researchers want to assess participants before and after some intervention—is often called a *pre-post design*.

**Within-subject designs that are not pre-post**

But within-subject designs are not always pre-post. For instance

* all participants might be instructed to complete cardio training one month
* then all participants might be instructed to complete weight training a second month
* after each month, they might then complete a survey to assess their motivation

As this figure shows, after participants completed cardio training, their motivation was 4.3 on average. After participants completed weight training, their motivation was appreciably high at 6.4. These findings imply that weight training motivates candidates more than does cardio training.

|  |
| --- |
| Complete weight training and then measure motivation: Average motivation = 6.4 |
|  |
| Complete cardio training and then measure motivation: Average motivation = 4.3 |
| Image result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette person |

**Problems that arise unless the order is counterbalanced.**

In the previous example, cardio training always preceded weight training. If all participants are exposed to these instructions in the same order, the researcher cannot conclude definitively that weight training enhances motivation more effectively than cardio training—for two reasons. The following table outlines these reasons.

|  |  |  |
| --- | --- | --- |
| Reason | Description | Illustration |
| Time effects | Perhaps the measure naturally changes over time | * PhD candidates might naturally become more motivated over time * This effect of time, rather than exercise, could explain why weight training seemed to enhance motivation |
| Transfer effects also called carry-over effects | Perhaps one condition affects the impact of another condition | * After candidates complete cardio training, their fitness improves * They might thus complete the weight training more effectively—enhancing their motivation. * Hence, the cardio training enhances the benefits of the weight training |

**Cross-over designs and Latin squares to counterbalance conditions**

To override these problems, researchers need to vary the order in which participants receive the instructions. In this example,

* half the participants complete cardio exercise first and then weight training last
* half the participants compete weight training first and then cardio exercise last
* all participants indicate the extent to which they feel motivated after cardio exercise and after weight training

Remember that each set of instructions is called a condition. For example, in this design, each individual participates in two conditions. As the following figure shows

* both clusters of participants indicated they were more motivated after weight training than after cardio training
* when combined, these findings indicate that weight training is more likely to promote motivation than will cardio training

|  |
| --- |
| Weight training: Average motivation = 6.4 Weight training: Average motivation = 5.4 |
|  |
| Cardio training: Average motivation = 4.3 Cardio training: Average motivation = 2.4 |
| Image result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette person |

The following table also represents this design. Each row represents a distinct cluster of participants. For instance, each cluster might comprise 25 individuals. From this table, you can probably appreciate why this configuration is often called a cross-over design.

|  |  |  |
| --- | --- | --- |
| Cluster of participants | First set of instructions these participants receive | Second set of instructions these participants receive |
| Image result for cluster of people silhouette | Condition A  such as cardio | Condition B  such as weights |
| Image result for cluster of people silhouette | Condition B  such as weights | Condition A  such as cardio |

If you wanted to compare three distinct exercises or conditions, you would need to organize six distinct arrangements. The following table illustrates these six arrangements.

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster of participants | First set of  instructions | Second set of  instructions | Third set of instructions |
| Image result for cluster of people silhouette | Condition A  such as cardio | Condition B  such as weights | Condition C  such as yoga |
| Image result for cluster of people silhouette | Condition A  such as cardio | Condition C  such as yoga | Condition B  such as weights |
| Image result for cluster of people silhouette | Condition B  such as weights | Condition A  such as cardio | Condition C  such as yoga |
| Image result for cluster of people silhouette | Condition B  such as weights | Condition C  such as yoga | Condition A  such as cardio |
| Image result for cluster of people silhouette | Condition C  such as yoga | Condition A  such as cardio | Condition B  such as weights |
| Image result for cluster of people silhouette | Condition C  such as yoga | Condition B  such as weights | Condition A  such as cardio |

Nevertheless, six distinct clusters of participants can be hard to arrange. Instead, researchers have developed an approach, called Latin squares, to diminish the number of clusters that researchers need to arrange. For example, when researchers want to compare three conditions, such as three exercises, they can utilize the arrangement that is illustrated in the following table.

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster of participants | First set of  instructions | Second set of  instructions | Third set of instructions |
| Image result for cluster of people silhouette | Condition A  such as cardio | Condition B  such as weights | Condition C  such as yoga |
| Image result for cluster of people silhouette | Condition B  such as weights | Condition C  such as yoga | Condition A  such as cardio |
| Image result for cluster of people silhouette | Condition C  such as yoga | Condition A  such as cardio | Condition B  such as weights |

As this table shows

* this arrangement does not include every possible sequence of conditions
* however, each condition—A, B, and C—is presented first, second, and third the same number of times;
* we thus refer to these conditions as counterbalanced
* consequently, differences between these conditions or exercises cannot be ascribed to time

Latin squares have also been developed for designs that comprise four or more conditions. Or, rather than utilize Latin squares, each participant could instead complete all the conditions in a random order.

**Mixed designs**

Cross-over designs and Latin squares comprise distinct clusters of participants, each of whom complete every condition. Because these designs comprise distinct clusters of participants, the design is partly between-subject. Because participants complete every condition, the design is partly within-subject. Consequently, researchers sometimes referred to these designs as mixed.

Not all mixed designs are cross-over designs or Latin squares. To illustrate, suppose that

* 50 candidates completed cardio for a month
* 50 candidates completed weight training for a month
* all participants indicated their motivation before and after this month

In this example, people completed cardio or weight training but not both, reminiscent of a between-subject design. But, people indicated their motivation before and after the exercise, reminiscent of a within-subject design. Consequently, this design is also mixed.

|  |
| --- |
| **Categories of between-subject designs: Randomized control trials or experiments** |

**Random allocation of individuals**

Researchers differentiate a variety of between-subject designs. The design that researchers tend to recommend is called the randomized control trial—or experiment. To illustrate, if the researcher wanted to explore whether exercise improves motivation in PhD candidates

* half the candidates would be randomly assigned to the condition in which they are instructed to exercise for a month
* the remaining candidates would be randomly assigned to the condition in which they are instructed to refrain from exercise for a month
* all individuals would then indicate the extent to which they feel motivated
* because of this random assignment, the pre-existing characteristics of individuals—such as their age, intelligence, and personality—should be roughly the same in each condition
* to illustrate, in this example, the number of blue participants is the same in each condition
* consequently, if the individuals who exercise are more motivated, we can quite certain this difference can be ascribed to the exercise rather than other pre-existing differences between the conditions

|  |
| --- |
| Instructed to exercise: Average motivation = 6.7 |
| Image result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette person |
| Instructed not to exercise: Average motivation = 3.6 |

In short, in a randomized control trial, or experiment, participants are randomly assigned to conditions. A variety of methods can be used to allocate participants randomly. The researcher might toss a coin or use a computer to generate random numbers, for example.

**Randomized allocation of clusters**

In many instances, researchers cannot randomly allocate individuals to conditions. To illustrate, suppose that, to expedite your research

* you wanted to present the same instructions to all PhD candidates who attend the same university
* you randomly allocate the PhD candidates of 20 universities to the condition in which they are instructed to exercise for a month
* you randomly allocate the PhD candidates of 20 other universities to the condition in which they are instructed to refrain from exercise
* all individuals would then indicate the extent to which they feel motivated

Hence, as the following figure shows, each university, rather than each individual, is randomly assigned to each condition. This design is thus called a cluster randomized control trial, because clusters of individuals are assigned to each condition. Clusters could include organizations, regions, or other constellations of people.

|  |
| --- |
| Instructed to exercise: Average motivation = 6.7 |
|  |
| Instructed not to exercise: Average motivation = 3.6 |

**Step-wedged design**

A variant of the cluster randomized control trial is called the step-wedged design. In this design

* participants are assessed several times. For example, their motivation might be tested once a month.
* after the first assessment, one cluster is randomly chosen to receive the intervention, such as instructions to exercise.
* after the second assessment, another cluster is randomly chosen to receive the intervention
* this procedure continues until all clusters receive the intervention.

|  |
| --- |
| **Categories of between-subject designs: Quasi-experimental and regression discontinuity** |

**Quasi-experimental designs**

Sometimes, for pragmatic reasons, researchers cannot randomly allocate participants to conditions. To illustrate, researchers might be able to organize the intervention—the exercise protocol, for example—only during January. Therefore

* PhD candidates who agree to participate in January may be instructed to exercise
* PhD candidates who agree to participate during another month might be instructed to refrain from exercise
* the candidates are thus not randomly allocated to conditions

In this instance, because participants are not randomly allocated to conditions, the researcher cannot reach definitive conclusions. To illustrate, suppose the individuals who were instructed to exercise are more motivated than were the individuals instructed to refrain from exercise, as shown in the following figure.



Arguably

* the PhD candidates who agreed to participate in January may simply be more motivated in general than PhD candidates who agreed to participate in subsequent months
* consequently, despite these findings, the exercise might not have affected motivation
* instead, pre-existing differences between the participants in each condition could explain the findings

This design is sometimes called quasi-experimental because the study, at first, resembles an experiment or randomized control design. That is, participants were allocated to conditions. But actually, participants were not randomly allocated to conditions.

**Regression discontinuity**

A slight variant of this design can generate more definitive conclusions. Sometimes, researchers utilize a systematic procedure to allocate participants to conditions. For example, to manage risks, researchers might

* allocate everyone with less than 50% body fat to the condition in which participants are instructed to exercise
* allocate everyone with more than 50% body fat to the condition in which participant refrain from exercise
* some hypothetical results appear in the following figure



At first glance, this design might seem unsound. Obviously, the difference between conditions could be ascribed to the effect of body fat rather than exercise. However, researchers could later conduct statistical techniques, such as ANCOVA or multiple regression, to nullify the effects of body fat. That is, if they apply these statistical techniques, they can estimate what the effect of exercise on motivation would have been had body fat been the same in everyone. Consequently, when researchers utilize this technique, this design, called a regression discontinuity design, is almost as compelling as a randomized control trial.

|  |
| --- |
| **Categories of between-subject designs: Case control and cohort studies** |

In quasi-experimental designs and regression discontinuity designs, participants are assigned to conditions, but this allocation is not random. In some designs, however, participants are not assigned to conditions at all. Designs in which participants are not assigned to conditions are sometimes called observational studies. Two of the most renowned examples of observational studies, common in epidemiology, are case control studies and cohorts studies.

**Case control studies**

In case control studies, the researcher compares two, or sometimes more, groups of individuals. Specifically, the researcher typically compares individuals with some feature—such as a disease—to individuals who do not exhibit this feature or disease. Usually, the researcher compares these groups of individuals on characteristics that may have cause this feature or disease. The aim of these studies is to assess which characteristics may have caused the feature or disease. To illustrate

* the researcher may first assemble 50 individuals with a disease of interest—such as a fear of statistics, represented by the blue figures in the following display
* the researcher then assembles 50 individuals who do not exhibit this disease or feature—the control group, represented by the black figures in the following display
* the control group may be generated by inviting a sibling of each person in the other group, or a comparable method, to guarantee that each group is derived from same population
* the researcher then compares the two groups on characteristics that could have caused the disease or feature of interest, such as average class size during school
* the characteristics that differ between the two groups may be causes of this disease or feature.

|  |
| --- |
|  |
| Image result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette person |

Average class size during school; IQ; height; weight

**Cohort studies**

In typical cohort studies, the researcher examines a sample of individuals over time. For example

* the researcher might first assemble a sample of 100 participants—perhaps university candidates who have not yet developed a fear of statistics
* the researcher might then assess these participants on measures that could affect the outcome of interest
* for example, the researcher might distribute a survey that gauges the average class size during school, IQ, height, and weight of these individuals
* then, perhaps some time later, such as a year, the researcher gauges which of these individuals developed a fear of statistics
* again, the characteristics that differ between individuals who developed the fear and individuals who did not develop the fear may be causes of this disease or feature

|  |
| --- |
| Average class size during school; IQ; height; weight |
| Image result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette personImage result for silhouette person |

**Differences between case control studies and cohort studies**

|  |  |  |
| --- | --- | --- |
| Case control studies | Cohort studies | Rationale |
| Measures participants  at one time | Measures participants  over time |  |
| Relatively inexpensive | Slightly more expensive | * Cohort studies need to monitor participants over time and, therefore, may be more expensive than case control studies |
| Retrospective | Prospective | * Case studies explore how measures in the past affect the outcome now; that is, the measures or causes of this outcome are examined retrospectively * Cohort studies explore how measures or causes now affect outcomes in the future—and thus are called prospective |
| The level of evidence is modest | The level of evidence is better than case control studies | * Because both kinds of studies are observational—devoid of a manipulation or random allocation—we cannot be certain the measures really cause the outcome * However, in cohort studies, researchers can be more certain about direction of causality |

|  |
| --- |
| **Categories of within-subject designs and mixed designs** |

**Split body design**

Like between-subject designs, researchers can also utilize several variations of within-subject designs and mixed designs. Certainly, the pre-post design and cross-over design are the most common. But, researchers can apply other variants as well. For example, medical researchers sometimes apply the split-body design. For this design

* each person receives two or more interventions at the same time
* one intervention might be applied to one region of the body; another intervention might be applied to another region of the body

One variant of this split-body design is the split-mouth design, often used in dentistry. For example, a dental researcher might

* apply some dental treatment to one side of the mouth, such as the left teeth
* apply some control condition to other side of the mouth
* compare the two sides of the mouth to ascertain whether the treatment was effective.

**Causal pathways**

Some within-subject designs explore the direction of causality. To illustrate

* suppose researchers showed that PhD candidates who exercise are more motivated
* perhaps motivation encourages people to exercise; or perhaps exercise enhances motivation

|  |  |  |
| --- | --- | --- |
| Exercise |  | Motivation |

To differentiate these possibilities, researchers need to assess these measures—exercise and motivation—and two times. In particular,

* suppose exercise fosters motivation; therefore, exercise at one time should coincide with high motivation at a later time—even after controlling motivation at the previous time
* suppose motivation fosters exercise; thus motivation at one time should coincide with frequent exercise at a later time—even after controlling exercise at the previous time
* multiple regression can be utilized to assess these hypotheses

|  |
| --- |
| **Which design should I utilize** |

If you conduct a between-subject study, you should utilize a randomized control trial whenever feasible. If you conduct a within-subject or mixed design study, you should typically conduct a cross over trial or Latin square whenever feasible. But, how should you decide whether to utilize a between-subject design or within-subject or mixed design? The following table outlines the benefits and drawbacks of each design.

|  |  |  |
| --- | --- | --- |
| Characteristic | Which design  is better? | Rationale |
| Statistical power | Within-subject design | * To generate significant results, you need about 25% as many participants or units for within-subject or mixed designs than for between-subject designs * Thus, if participants are willing to persist on this study over time, consider within-subject designs |
| Ethics | Within-subject designs | * For between-subject designs, only a subset, often half, the participants receive the beneficial treatment—such as instructions to exercise. * Consequently, some participants are treated better than other participants |
| Withdrawal | Between-subject designs | * For within-subject designs, participants often withdraw from the study prematurely * Consequently, the final analysis often includes only a subset of participants—the participants willing to persevere * This final subset of participants might not be representative of the population; the results could be misleading |
| Asymmetric transfer |  | * Counterbalancing is intended to overcome carry-over effects, in which one condition affects the impact of a subsequent condition * But, counterbalancing does not achieve this goal in all circumstances * In particular, some studies are contaminated by a problem called asymmetric transfer: the impact of Condition A on Condition B differs from the impact of Condition B on Condition A * In this instance, within-subject designs might generate misleading results. |

|  |
| --- |
| **Summary and statistics** |

Whenever researchers utilize some design, they can conduct a range of statistical techniques to analyse the data, depending on many considerations, such as properties of the measures. Nevertheless, for each design, some techniques are especially common The following table summarizes these common techniques.

|  |  |  |
| --- | --- | --- |
|  | Design | Common statistical technique |
| No groups | Correlational design | * Correlation * Multiple regression * Logistic regression |
| Between-subject designs | Randomized control trial | * Independent t-test if two conditions * ANOVA if more than two conditions |
| Cluster randomized control trial | * Multi-level modelling or linear mixed models \* |
| Quasi experiment | * ANCOVA * Multiple regression with dummy variables * Propensity score matching \* |
| Regression discontinuity design | * ANCOVA * Multiple regression with dummy variables |
| Case control studies | * Logistic regression analysis |
| Cohort studies | * Survival analyses, such as hazard models |
| Step-wedge design | * Multi-level modelling or linear mixed models * Generalized linear mixed models \* * Generalized estimating equations \* |
| Within-subject or mixed designs | Pre-post design | * Dependent or paired-samples t-test |
| Within-subject design over more than 2 times | * Repeated-measures ANOVA |
| Cross-over or Latin square design | * Mixed model two-way ANOVA in which the cluster is a random variable * Multi-level modelling or linear mixed models \* |

\* This technique is challenging but more accurate

**Level of evidence**

To ascertain which design to utilise, you need to consider

* which designs unearth the most compelling evidence; that is, which designs most convincingly identify the causes of some outcome
* which designs are most feasible, inexpensive, and so forth

To determine which designs are most compelling, you might want to consult various guidelines that scholars have developed, sometimes called hierarchies of evidence or levels of evidence. To illustrate, one set of guidelines called the Oxford Levels of Evidence, appear in the following table. For example,

* according to this hierarchy, systematic reviews of randomised controlled trials are more compelling than systematic reviews of cohort studies
* these guidelines are often used in health and medicine
* read the document about critical thinking to learn more about why some designs might be more convincing than other designs

|  |  |
| --- | --- |
| Level of evidence | Design |
| 1 or best | * Systematic reviews of randomised control trials—if homogenous * Randomised control trials |
| 2 | * Systematic reviews of cohort trials—if homogenous * Cohort studies or flawed randomised control trials |
| 3 | * Systematic reviews of case control studies—if homogenous * Case control studies or correlational studies |
| 4 | * Case study, flawed cohort study, or flawed case control study |
| 5 or worst | * Expert opinion—based on first principles, physiology, and so forth |