

Experimental (quantitative) studies: An overview and general issues by Stephanie Tierney (stephanie.tierney@manchester.ac.uk)



Experimental designs are said to be *the* approach for obtaining information about causal relationships (Robson, 1993), allowing researchers to assess the correlation (relationship) between one variable and another. A principle factor of such designs is that one element is manipulated by the researcher to see whether it has any impact upon another. The element being manipulated by researchers (e.g. introducing a teenage pregnancy preventative intervention) is known as the independent variable, whereas the change (or outcome) resulting from the implementation of the independent variable (e.g. teenage pregnancy rates) is the dependent variable.

General terms associated with experimental studies

Hypotheses

Experimental designs are developed to answer hypotheses, or testable statements, formulated by researchers to address specific questions; for example, is family therapy an effective means of helping children with mental health problems? The investigators set up an experimental study (e.g. a randomised controlled trial, see below) and collect and analyse data, which will support or disprove the hypothesis.

Hypotheses can be based on theory, on earlier research findings or on a 'hunch' that the researcher may wish to examine further. A two-tailed hypothesis refers to those testable statements that can go in one of two directions. For example, is there a difference in the outcome of people with mental health problems who are given a drug compared to psychotherapy? Alternatively, investigators may decide to employ a one-tailed hypothesis, for which the direction of the outcome would be pre-stated, such as a prediction that mental health patients receiving psychotherapy would fare better than those treated with medication.

When testing a hypothesis, it is convention to state that no difference will occur. This is defined as the *null hypothesis*. Hence, at the start of a study, a researcher might produce a null hypothesis stating that family therapy is no better than standard care for children with mental health problems. This is then tested via an experimental study and if the intervention demonstrates superiority over the control condition, the null hypothesis can be rejected.

Internal and external validity

An experimental approach is said to be an effective means of strengthening:

- *Internal Validity* – This relates to how far a study has established whether a variable or condition under scrutiny has had an effect. Controlling against extraneous variables (see below) strengthens internal validity.
- *External Validity* – This relates to whether findings from a specific sample in a study can be generalised to a larger, target population.

Extraneous and confounding variables

Extraneous variables constitute all factors, other than the independent variable, that have an effect on the dependent variable. Since an underlying aim of the experimental approach is to see if alterations in the independent variable (the variable manipulated by a researcher - such as the introduction of a new service) leads to changes in the dependent variable, it is essential that extraneous variables are considered when planning a study. Foster and Parker (1995) divide extraneous variables into three forms:

- Participant variables – e.g. gender, being left or right handed, IQ.
- Investigator variables – e.g. different investigators collecting data from different respondents.
- Situational variables – “any aspect of the task, physical environment, or temporary features of the respondents (such as hunger, fatigue), which might influence performance” (Foster and Parker, 1995: 25).

Potential extraneous variables should not prejudice the relationship between the independent and dependent variables. Extraneous variables that have such an unwanted effect are known as confounding variables. Confounding variables lead to ambiguous results. In an experiment, every effort is made to control for confounding variables in order to be more confident of the cause-effect relationship. In the case of the early pregnancy preventative package referred to above, the researchers will be keen to know whether any decrease (or increase) in pregnancy rates are due to the new intervention and not to alternative factors (e.g. family background, educational levels, ethnicity, personality). It is important to realise that extraneous variables may not necessarily be confounding and it may prove impossible to account for all extraneous variables. Nevertheless, researchers should take steps to ensure that variables do not become confounding, thus strengthening the legitimacy of claims made about the relationship between dependent and independent variables. A principle means of overcoming extraneous variables is through the use of a study design known as the randomised controlled trial (RCT), which is described in detail below.

The Placebo Effect

When people know they are taking medication or receiving a treatment, they can gain a psychological and neurobiological boost which improves their condition. This is known as the placebo effect. It can be induced not by the specific properties of the pill taken or treatment received, but by the very fact of having some sort of intervention, whether or not it is 'active' or 'real'. It has been noted that it is the expectation that an intervention will be beneficial that results in such amelioration. Examples of the placebo effect include caffeine free coffee inducing the same effects as regular coffee when those drinking it believe they are consuming 'normal' coffee; warts disappearing when brushed with an inert, coloured dye, after patients are assured by their doctor that it will help; patients with wisdom tooth pain obtaining as much relief from a fake application of ultrasound as from a real one (see <http://skeptdic.com/placebo.html> for more examples, checked 06/07/08). This is why it is important to include a control group within a study, which may improve in a trial sometimes to a similar extent to the intervention group. In an experimental study, researchers will look for any improvements in the intervention group over and above those reported in the controls. The use of a control group receiving nothing or an alternative intervention (e.g. standard care) means that any differences between groups can be more firmly related to the independent variable (e.g. the intervention under consideration).



Laboratory versus field experiments

A distinction should be made between laboratory and field experiments. The former take place in an artificial setting, with the advantage of being able to control precisely for confounding variables. Laboratory experiments allow for the setting up of artificial situations that would not occur in everyday life and for ease of replication if other researchers wished to conduct a similar investigation. However, the types of variables and situations that can be studied under laboratory conditions are restricted, and the degree of artificiality makes the generalisation of findings to real life more problematic. Field experiments have good ecological validity, i.e. they can be transferred more easily to other, real life settings because they take place in familiar environments, such as a school, clinic, day care centre. However, confounding variables are less easily controlled in field settings compared to those performed in a laboratory.

Types of experimental studies

Randomised controlled trials (RCTs)

RCTs are considered to be the strongest form of primary evidence when examining questions of effectiveness (Greenhough, 2001) because they are structured to minimise bias, as far as possible, through randomisation and use of a control arm:

- *Randomisation* – The use of this technique should produce a representative sample, typical of the population under investigation (e.g. consumers of a specific service). Care must be taken in the mode of randomisation employed, with ‘pseudo’ methods of randomisation avoided, such as clients attending a service on alternative days, because bias can arise. Biased sampling occurs when the sample chosen fails to represent the true makeup of the overall population of interest. Random sampling should overcome such bias, by ensuring that every individual within the target population has the same chance of being chosen to participate. It is assumed that through randomisation confounding variables will be evenly distributed across different groups so that they are comparable, with only the independent variable distinguishing them.
- *Use of a control group* – Individuals in control groups will either not receive any intervention, or will receive an alternative treatment, often the standard applied to the specific problem or situation. The control group used should be as closely matched as possible to the intervention group to avoid any confounding factors from interfering with results produced, such as age, gender, social class. People within the control group will not have received the intervention under investigation (the ‘independent variable’). This allows for a comparison between control and intervention groups. Since the former did not receive the independent variable, any differences between the two should relate to the effect of the intervention, as long as the groups are of a similar makeup at the outset.



Factors that can affect the results of RCTs

Although RCTs are considered to be the gold standard when assessing the effectiveness of an intervention, they can still be affected by various processes that can have an impact on results:

- *Selection bias* – According to Jadad (1998: 30-31), selection bias occurs “when the outcomes of a trial are affected by systematic differences in the way in which individuals are accepted or rejected for a trial, or in the way in which the interventions are assigned to individuals once they have been accepted into a trial.” The use of RCTs, compared to other research designs, helps to reduce the risk of selection bias because random allocation should mean that all individuals have an equal chance of being allocated to the intervention or control group. However, selection bias can occur if participants are allocated to one arm of a study or another in such a way that it is possible to ‘fiddle’ this process. For example, if a practitioner responsible for the care of patients taking part in a trial had the job of randomising people to one group or another she might be tempted to change the sequence of allocation to ensure that those she felt would benefit most from the intervention ended up in this group. The best way of

ensuring that this type of 'tinkering' does not occur is to have randomisation carried out by someone unconnected with participants, preferably independent of those conducting the research.

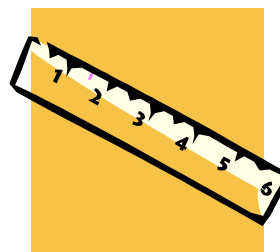


- *Performance bias* – Once participants have been allocated to intervention or control groups, the next danger is that something will have an effect on how interventions are delivered. Performance bias refers to the fact that the behaviour of participants or investigators may alter outcomes simply because they are participating in a study. For example, if a participant knows that they are in the intervention rather than the control group,

they may be more motivated to improve. Likewise, if a practitioner in a study thinks intervention A is less effective than intervention B, he might be more attentive to those receiving intervention A to compensate for this perceived deficiency (Crombie, 1996). This is why *blinding* (or *masking*) is used in RCTs. Blinding is a term that comes from medical trials, and refers to a 'not knowing'. For example, in a drug trial a doctor may be given two boxes of pills, one labeled A, the other B. Tablets in bottles A and B look identical, but one set is the drug being investigated, the other is a sugar pill or placebo. The doctor has no idea which is which, and neither does the patient receiving them. This is known as *double blinding*, i.e. neither participant nor those providing the intervention know whether or not the drug or placebo is being received. A *single blind study* means that those receiving an intervention do not know whether they are taking the drug or the placebo, but the doctor giving them out does (or vice versa). It is easy to see how blinding might work in a drug trial, but when investigating psychosocial interventions, things are a little trickier. For example, it is impossible for a therapist not to know that they are providing individual cognitive-behavioural therapy as opposed to systemic family therapy. However, assessor blinding is possible (see below).

In relation to performance bias, it is important to consider whether participants in a study received any additional treatment, aside from their allocated intervention, which might have had an effect on their progress.

- *Measurement bias* – People assessing outcomes in a study should not know the groups to which individual participants have been allocated, otherwise knowledge of group allocation could affect assessments they make (e.g. giving people in an intervention group better scores). If those assessing outcomes know which intervention participants have received there is a danger that they may only focus on data supporting the intervention.



If participants involved are aware of being in the intervention or control group, it may impact on their response to questions about their functioning or quality of life. This can be overcome by a) blinding participants to allocation, so they are unaware of whether or not they have received the intervention or control (see above); b) using a standardised, non-emotive means of assessing outcome; c) obtaining outcome data from independent sources, such as casenotes (see Elwood, 1998).

Another issue relating to measurement bias is that the way in which assessment of outcome is carried out should be the same for each group (e.g. carried out using the same timeframe and using the same assessment format).

- *Attrition bias* – Researchers conducting RCTs must consider any withdrawals from their sample because there may be a systematic difference between those completing and those not completing a study. For example, people may have dropped out because they felt they were making little progress. If too many individuals fail to take part in follow-up assessments, results may be biased.

Quasi-experimental studies

Quasi-experiment refers to studies in which participants are not randomised to conditions. In this type of design, researchers do not have complete control of independent variables because the intervention is already in place, or because it is impossible or ethical to manipulate the variable (e.g. when measuring the effects of smoking on people's health, it would be unethical to randomise people to a smoking or non-smoking group). Researchers rely on existing populations (e.g. people already smoking versus those who do not smoke). Hence, a control group is included, but individuals are not randomly allocated to condition; usually groups are naturally occurring (e.g. people in a part of the country where a new service has been established compared to individuals in another part of the country where the service has not been set up).

The problem with such designs is that any differences between two groups are harder to control for, giving less certainty of the cause and effect relationship. But for many studies this may be the only design option available. Therefore, it is the researcher's task to "tease out the threats to valid inferences about causation present...and to evaluate how far these threats can be discontinued in a particular study, taking into account the specific features of the study and the pattern of results obtained" (Robson, 1993: 46-7).

Single subject designs

Group designs (as described above) help give a generalised picture of response to a particular condition/intervention/treatment. However, it prevents the investigator from making any conclusions about how specific individuals react.

This is where single case (or small-N) designs come in useful, a tool that practitioners can employ to monitor progress of their own work.

Such designs may be a simple A-B makeup. A researcher planning to carry out such an experiment with an individual would firstly need to get a baseline measure of the response or behaviour under investigation. Once this has been established, the researcher then proceeds to intervene, changing conditions and measuring how this affects the participant's response.



The A-B-A is an improvement on the above, adding a reversal stage to the design, removing the intervention and then taking another measurement, looking for a return to baseline scores as a consequence. Problems with such a design include: a) there may be other factors accounting for changes observed aside from the intervention, and b) practitioners may have reservations about removing treatment when they are supposed to be helping someone. In an attempt to try and overcome ethical problems related to treatment withdrawal, a researcher could try using an A-B-A-B design, in which an individual is given an intervention, it is withdrawn and then it is given once again. This process can be repeatedly followed creating an A-B-A-B-A-B design, increasing the investigator's confidence in the relationship between the intervention and outcome.

The single case may not necessarily be an individual; it could be an individual unit (a school class, a specific service). The main focus of such research is that repeated measures of the same individual (or unit) are taken at specific times, usually pre and post intervention.

Operationalising definitions

In experimental research it is important that the researcher is clear about what they are doing from the outset. Therefore, before embarking on an experiment, researchers must have 'operationalised' the variables (independent and dependent) that are to be measured. For example, when measuring the effectiveness of an intervention to improve the quality of life for older people in the community, how is quality of life to be measured? What aspects of this rather abstract concept are to be considered in assessing whether any form of progress or change has been established? One way is to use a standardised measure of quality of life, complemented by semi-structured interviews, in which participants are asked to define for themselves whether or not they feel an intervention has helped and, if so, what specific areas of their life have improved. A clear definition of concepts involved in the experiment not only ensures its validity, but also increases the chances that it could be replicated by other researchers wishing to carry out a follow-up or similar investigation.

References

Crombie, I.K. (1996) *The pocket guide to critical appraisal*. BMJ Publishing: London.

Elwood, M. (1998) (2nd Edition) *Critical appraisal of epidemiological studies and clinical trials*. Open University Press: Buckingham.

Foster, J.J and Parker, I. (1995) *Carrying out investigations in psychology: Methods and statistics*. The British Psychological Society: Leicester.

Greenhalgh, T. (2001) (2nd Edition) *How to read a paper: The basics of evidence based medicine*. BMJ Publishing: London.

Jadad, A. (1998) *Randomised controlled trials*. BMJ Books: London.

Robson, C. (1993) *Real world research: A resource for social scientists and practitioner-researchers*. Blackwell: Oxford; Cambridge.