NHS
National Institute for Health Research

The NIHR Research Design Service for the East Midlands

The NIHR Research Design Service for Yorkshire & the Humber

Experimental Design

Authors

Heather Wharrad

With revisions by

Paul Silcocks

This Resource Pack is one of a series produced by The NIHR RDS for the East Midlands / The NIHR RDS for Yorkshire and the Humber. This series has been funded by The NIHR RDS EM / YH.

Reference as:
Wharrad H. and Silcocks P. An Introduction to Experimental Design. The NIHR RDS EM / YH, 2009.

Heather Wharrad
School of Nursing
University of Nottingham
Medical School
Queen's Medical Centre
Nottingham
NG7 2UH

Paul Silcocks
Clinical Senior Lecturer
CTSU / The NIHR RDS EM / YH
University of Nottingham
Medical School
Queen's Medical Centre
Nottingham
NG7 2UH

Last updated: 2009

The NIHR RDS for the East Midlands
Division of Primary Care,
14th Floor, Tower building
University of Nottingham
University Park
Nottingham
NG7 2RD
Tel: 0115 823 0500

www.rds-eastmidlands.nihr.ac.uk
Leicester: enquiries-LNR@rds-eastmidlands.org.uk
Nottingham: enquiries-NDL@rds-eastmidlands.org.uk

The NIHR RDS for Yorkshire & the Humber
ScHARR
The University of Sheffield
Regent Court
30 Regent Street
Sheffield
S1 4DA
Tel: 0114 222 0828

www.rds-yh.nihr.ac.uk
Sheffield: rds-yh@sheffield.ac.uk
Leeds: rds-yh@leeds.ac.uk
York: rds-yh@york.ac.uk

Table of Contents

1. Introduction

LEARNING OBJECTIVES

The aim of this pack is to provide an overview of the principles underlying experimental research design to allow the reader to be able to read research literature and evaluate research proposals critically.

This includes the following objectives:

1. To introduce the reader to experimental method as it relates to primary health care.
2. To describe different experimental designs and give examples of when they may be used.
3. To determine when the experimental approach is an appropriate research method for a particular research question.
4. To identify strengths and weaknesses of the experimental method.

1.1 Identifying the problem

The word 'experimenting' is used in our everyday language to describe our actions when we are testing or trying out something new, for example, 'I am experimenting with this new recipe' or 'they are experimenting with a different team'. In such situations the experimentation is performed albeit unwittingly, in a fairly ad hoc or 'uncontrolled' manner. In experimental research however the design and progress of an experiment is carefully controlled by the researcher.

Any research problem is inevitably multifactorial in nature. Let us consider an example. A health worker might wish to investigate the progression of pressure sores in patients. Many factors, in this context known as **variables**, might influence the development of pressure sores, some of these are shown in Figure 1. Some of the variables relate to the characteristics of the patient, others relate to the medical or nursing environment to which the patient is exposed, and others reflect the disease itself. Descriptions of the pressure sores themselves are also variables, so the number of sores, the size of sores and the grade of sore are all variables.

Figure 1.

Patient	age sex mobility nutrition psychological condition
Environment	mattress nursing practice medical/surgical interventions
Disease	number of sores grade of sores size of sores

1.2 Dependent and independent variables

In experimental research, the researcher manipulates one or more of the variables, these are termed the **independent variables** (also referred to as **"predictor"** or **"explanatory"** variables), and monitors how subjects react by measuring a response in one or more outcome measures or **dependent variables** (also known as **"response"** or **"outcome"** variables). In the simplest experimental design, a researcher might manipulate just one independent variable and measure a change in one dependent variable, so in our example in Figure 1, a health worker might investigate the relationship between mobility (the independent variable) and the grade of pressure sore which develops in a patient (dependent variable). If a researcher wanted to investigate the effect of posture on pulse rate, he/she might measure pulse rate in a group of subjects who were sitting, then repeat the measurement with the subjects standing. Here the posture adopted is the independent variable as this is the variable being changed by the experimenter. The dependent variable is pulse rate - this is the variable being measured during the experiment. The change in the independent variable is planned by the researcher before the experiment is carried out, whereas, how the dependent variable will change is not known until the experiment has been performed - it is 'dependent' on the change in the independent variable.

It is important to distinguish the outcome *variable* being studied (change in pulse rate in this example), which is measured on individuals, from the parameter that measures the effect of the intervention which is an aggregate property of the group being studied (in this example the **effect parameter** is the *average* change in pulse rate.

EXERCISE 1

Extract 1

Pressure sores and pressure-decreasing mattresses: controlled clinical trial.

A.Hofman, RH.Geelkerken, J Wille, JJ Hamming, PJ Breslau
Lancet 1994; 343: 568-571.

Summary

Pressure sores are a problem, especially in elderly patients. Our study was designed to determine the effectiveness in pressure sore prevention of a new interface-pressure decreasing mattress.

In a prospective randomised controlled trial we tested the Comfortex DeCube mattress (Comfortex, Winona USA) against our standard mattress in 44 patients with femoral-neck fracture and concomitant high pressure sore risk score. In addition both groups were treated according to the Dutch consensus protocol for the prevention of pressure sores. On admission, and 1 and 2 weeks after admission, pressure sores were graded. The two groups were similar in patient characteristics and pressure sore risk factors. At 1 week, 25% of the patients nursed on the DeCube mattress and 64% of the patients nursed on the standard mattress had clinically relevant pressure sores (grade 2 or more). At 2 weeks the figures were 24% and 68% respectively. The maximum score over the several body regions of the pressure sore grading, measured on a 5-point scale was significantly different in favour of the DeCube mattress at 1 week (p=0.0043) and 2 weeks (p=0.0067) post-operatively.

We show that the occurrence of pressure sores and their severity can be significantly reduced when patients at risk are nursed on an interface-pressure decreasing mattress.

1. What is the dependent variable in this experiment?

2. What is the independent variable?

3. What would you choose as the effect parameter?

Answers to exercises are given at the back of this resource pack.

As we shall see later in this pack, more complex experimental designs allow the investigation of the relationships between a number of independent and dependent variables within one experiment. One word of warning though, a common mistake made by inexperienced (and sometimes the more experienced!) researcher is to try to address too many research

questions in one experiment, this not only leads to difficulties when analysing and interpreting the data but also can be time-consuming for researchers and subjects alike.

The results of a well designed experiment, along with appropriate statistical analyses (See The NIHR RDS EM / YH Resource Pack: *Using Statistics in Research*), allows the experimenter to make inferences about possible associations between the independent and dependent variables. The next section will introduce the basic principles behind experimental designs.

2. Basic principles of experimental design

2.1 Developing research questions and formulating hypotheses

Research questions develop from the identification of a 'problem' or 'issue' derived from our own professional experiences within the health care setting and from studying relevant literature on the subject. From these questions we formulate **hypotheses** to test. A **hypothesis** is a testable proposition about the outcome of an experiment, if you like it is an 'educated guess' based on your knowledge and experience about how the independent variable will affect the dependent variable. Every experiment tests at least one hypothesis, more often several hypotheses, in each case predicting the relationship between variables.

Not all questions are *formulated* as hypotheses, sometimes it is easier to formulate a research question in terms of estimating a quantity with desired precision. However, a question such as this can often be expressed as "wanting to rule out" a particular value, in which case the question can be seen as a hypothesis test in disguise.

EXERCISE 2

The hypothesis is often not overtly stated in published research papers. Can you identify the statement in Extract 1 which infers the research question?

2.2 Null hypothesis

Experimental research relies on statistical analysis to determine whether the change in the independent variable is related to the change in the dependent variable. Statistical analysis is the subject of another pack in this series, however it is worth introducing at this point the term **Null hypothesis**. The **Null hypothesis** is a statement of no difference or no relationship between the variables and it is a fundamental concept on which all statistical tests are based.

Examples of null hypotheses

(i) There is no difference in pulse rate in the sitting and standing positions.

(ii) Mobility has no effect on grade of pressure sore.

(iii) Drug X has no effect on diastolic blood pressure.

EXERCISE 3

Try altering the statement below into a Null hypothesis - it is the one posed in Extract 1.

> 'Our study was designed to determine the effectiveness in pressure sore prevention of a new interface-pressure decreasing mattress.'

To re-iterate the important point made in the introduction, the more complex the pattern of interaction of variables the less likely it is an experiment can be designed. It is worth remembering that when asking broad questions about subject's opinions a **qualitative**[1] approach might be considered as a more appropriate research method.

2.3 Selection and manipulation of variables

One essential component of the experimental design is the evaluation of manipulated change, the assumption is made that the dependent variable will not vary unless the independent variable changes. The independent variable is often termed the **intervention** or **treatment**. The choice of independent and dependent variables should be carefully considered. This section summarises the factors to take into account when selecting the dependent and independent variables for your experiment.

Reliability

A measurement tool is said to be reliable if it can be consistently used by other researchers to obtain **reproducible** results when applied in an identical (as far as is possible) setting. If the dependent variable was something fairly simple like height or weight, the choice of measurement instrument is not a problem. However, even with these simple instruments, it is important to further **standardise** the measurement by defining standard conditions to be followed when making the measurement such as, removing shoes and checking that the scales are zeroed before each measurement. The omission of even simple instructions such as these can lead to unreliable results particularly if different researchers are making the measurements. Consider another fairly simple measurement, estimation of arterial blood pressure using a sphygmomanometer. There are four main sources of variation and error in the technique leading to unreliability in this measurement: the observer, the patient, the technique and the equipment. These have been published in more detail in a paper by O'Brian and Davison (1994). Factors that enhance the accuracy of arterial blood pressure measurement and recording are listed in Table 1.

[1] Qualitative research - a system of inquiry describing and summarising actions. Data is in the form of words and language rather than numbers. Useful ways of gathering these forms of data are participant observation, interviews and the studying of documents (see also study pack on qualitative research)

Patient	Sitting or lying for at least 3 minutes No clothing on arm Arm supported level with heart Psychologically calm No recent smoking or alcohol consumption
Equipment	Matching cuff dimensions to patient's arm circumference Sphygmomanometer calibrated 6-monthly Sphygmomanometer and stethoscope regularly maintained
Observer	Agreement about phase IV and V Korotkoff sounds for diastolic pressure Reading to the nearest 2mm Hg i.e. no rounding to the nearest 5 or 10 mm Hg Patient's arm adequately supported Cuff fitted firmly Centre of bladder cuff over brachial artery

Table 1. Factors to consider when standardising the technique for blood pressure measurement that enhance the accuracy of the recording (adapted from O'Brian and Davison, 1994).

These considerations are important for health workers in practice as well as for researchers. If there are variations in the way blood pressure is recorded by different researchers then a true change in blood pressure due to an intervention might be missed. An erroneous reading which is consistently made by an observer, such as always rounding up the reading to the nearest 5mm Hg instead of reading to the nearest 2mm Hg is called a **systematic error,** which is one way in which study results may become biased.

EXERCISE 4

Which other errors in blood pressure measurement might be considered to be systematic errors?

It is also important that the intervention (the independent variable) is applied in exactly the same way to all the subjects throughout the experiment. In the same way that the measurement of the dependent variable is **standardized,** so should the application of the independent variable. This is achieved by verifying techniques and by using a standard protocol. A pilot study (See Page 12) will verify whether the protocol is workable.

Validity

Measurement tools also need to be **valid**. A valid test or tool is one which measures what it purports to measure. This type of validity is known as **face validity**. There is a good example which is commonly used to explain the concept of validity (See also Clegg, 1994). You can use a tape measure to determine head size and record it in centimetres and thus have a perfectly reliable measurement which would be reproducible if other people repeated the same measurement. It is also a valid measurement of head size. However if you were to use it as a measure of intelligence then the measurement is still reliable but it is certainly not valid!

Intelligence is an example of many important dimensions of human subjects that are so complex and multivariate that it is impossible to gain universal agreement about how it can be measured validly and reliably.

EXERCISE 5

Before reading on, spend a minute or two thinking about other aspects of human behaviour or experiences that might come under this category.

Some of the aspects of human behaviour you might have come up with are:

- pain
- stress or anxiety
- health care
- attitudes
- personality

Some workers have used physiological or biochemical indices to measure changes in stress or pain. For example, Boore et al (1978) measured levels of corticosteroids excreted as her dependent variable for determining stress levels. Some researchers argue that this reductionist [2] approach in order to obtain objectivity, is inappropriate for human subjects. Another difficulty is that these are **surrogate variables** – ie substitutes for the "thing" you are really interested in. The advantage of a surrogate variable is that it may give quicker and cheaper results, but there is always the risk that the surrogate may not respond to an intervention in the same way as the real variable of interest – so you end up getting a precise answer to the wrong question!

Many 'tools' in the form of questionnaires are now available for assessing these complex emotions in a more holistic way (See examples in Table 2 below). These types of tools require consideration of what is termed **content validity**. This means does the instrument (questionnaire) include all the relevant questions related to the hypothesis? In other words will this tool give valid results for the research question being investigated?

Dependent variable	Measurement tool	Experimental study using tool
Pain	McGill pain questionnaire (MPQ)	Pain on a surgical ward Melzack et al (1987)
Anxiety	Spielberger STAI-X1	Who's afraid of informed consent? Kerrigan et al (1993)

Table 2.

Where a recognised reliable or valid measurement tool for assessing the dependent variable does not exist, researchers must develop one themselves ensuring to check it for reliability and validity.

[2] Reductionist - human processes or behaviours are broken down into simpler components such as specific physiological events or molecular or biochemical processes. This is opposite to the holistic approach.

Potency

It is important to ensure that the intervention is of sufficient **potency** to produce a measurable change in the dependent variable otherwise we assume wrongly that an intervention has no effect (in statistics this is termed a type II error). The simplest example of this is giving a drug in too low a dose. In drug trials, early experiments take the form of dose-response curve to find out which dose is effective. A related issue is that the dependent variable must be a quantity which is **sensitive** to the intervention. To some extent sensitivity can be improved by reducing noise (ie measurement error), for example by taking replicate measurements and averaging them. Sensitivity is also greater when a continuous variable (like blood pressure) is measured directly rather than being categorised into "high" or "low".

Example

In a trial of middle-aged men of the effect of advice to stop smoking, the intervention was a series of information-giving interviews. The experimenters felt that to change an entrenched behaviour such as smoking habits by providing information via a pamphlet or booklet rather than using interviews was unlikely to be a potent enough intervention (See Rose et al, 1978).

The pilot study

The procedures and protocol of an experiment are normally tested on a small number of people prior to the main study. A **pilot study** is a trial run. It allows researchers to check whether equipment is functioning properly or whether respondents understand the phrasing of questions in a questionnaire. It is rare, however well planned, for a few unforeseen problems not to arise in the course of an experiment. The subjects who take part in the pilot study should be from the same general population as the subjects selected for the main study (See also Section 2.5).

Example

Study of the effects of education on patient's knowledge and acceptance of indwelling urethral catheters. Roe, BH (1990) *J.Adv.Nursing 15 223-231.*

> 'The information booklet, written on the basis of a literature review, was tested at a pre-pilot stage by lay people (non-catheter users) who were asked to check its readability and sense. Amendments were made and it was then formally tested in the pilot study.'

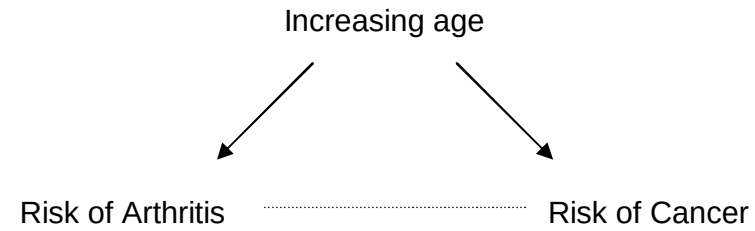
2.4 Control groups

Consideration of the independent variables leads neatly on to the discussion of the term **control**. If the aim of the experimental method is to determine the relationship between the independent and dependent variables it is obviously important to ensure that all the other known possible influences are held constant. Thinking back to the classic experiments we did at school in plant biology, it was fairly easy to keep constant factors other than those we were investigating so when looking at the effect of light on leaf colour we could ensure that temperature, amounts of water etc were identical for all plants in the experiment, that is we **control** for as many of these unwanted factors as possible. The only variable manipulated would be the independent variable - light. These unwanted factors are called **extraneous**

variables. Extraneous variables matter when they are associated with both the independent and the dependent variable (when they are known as **confounding variables).**

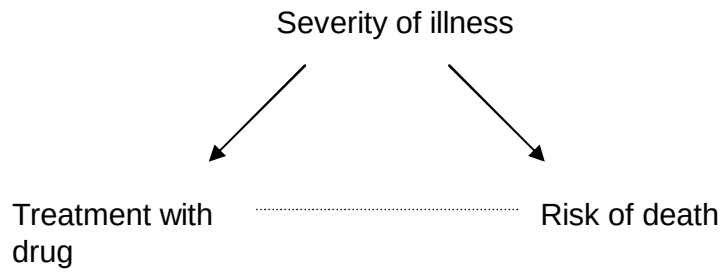
For example, advancing age increases the risk of developing cancer. Also advancing age increases the risk of developing arthritis. Unless age is allowed for, age will confound the association between Arthritis and Cancer, i.e. there will be a spurious association between arthritis and cancer (denoted by the dotted line), as in the Figure below:

Figure 2a.



Another possible example might be:

Figure 2b.



The example in Figure 2b shows how confounding could result in the mistaken conclusion that treatment with a particular drug increased the risk of death. This possibility is referred to in pharmacoepidemiology circles as "confounding by indication" – the indication for the drug being the disease severity, which acts as a confounding variable.

In human research, controlling other variables is more problematic than in botany or the physical sciences! The complexities of human behaviour mean that not only is it difficult to eliminate confounding variables but it is often impossible to identify all of them in the first place!

EXERCISE 6

Extract 1 summarises a study describing the effect of mattress type on the development of pressure sores. List the confounding variables that might have influenced the dependent variables in this study - you may find Figure 1 helpful!

One way of accounting for any influence of confounding variables on the dependent variables is by means of a **control group**. Subjects in the control group are in theory identical with the subjects in an experimental group in all aspects, except for the intervention or treatment and they therefore act as a reference to indicate "what would have happened otherwise", sometimes termed the ***counterfactual***. Without a control group it is difficult for researchers to be sure that the observed change in the dependent variable was solely due to the independent variable or due to other influences (confounding variables). In practice of course this ideal will not be met – even if the average age and proportion of females, say, are similar in the control and intervention groups, they may differ with respect to age-sex combinations, and also with respect to other variables. Additional methods have to be used to deal with confounding in practice.

In recent years the nature of the control group has changed. Rather than receiving no treatment at all, control groups may receive a placebo (dummy) treatment – this enables the researcher to deduct the "**placebo effect**", that is, the response observed simply as a consequence of the attention given or the expectation of some effect based perhaps on the colour of a pill (or some other aspect not central to the hypothesis). In addition, because of raised awareness of ethical responsibilities, it is recognised that it is better to have a control arm which is provided with a placebo *in addition to* either palliative treatment (if no specific remedy is known) or to standard care. It is not ethical to withdraw or withhold a beneficial treatment in a group of control subjects simply to satisfy the conditions of experimental research (See also Section 4.1 – Humanity & Ethics).

Extract 2

Who's afraid of informed consent?

DD Kerrigan et al , BMJ 1993; 306: 298-300.

Objective - To test the assumption that patients will become unduly anxious if they are given detailed information about the risks of surgery in an attempt to gain fully informed consent.

Design - Pre-operative anxiety assessed before and after patients were randomly allocated an information sheet containing either simple or detailed descriptions of possible post-operative complications.

Setting - Four surgical wards at two Sheffield hospitals.

Subjects - 96 men undergoing elective inguinal hernia repair under general anaesthetic.

Main outcome measure - Change in anxiety level observed after receiving information about potential complications.

Results - Detailed information did not increase patient anxiety (mean Spielberger score at baseline 33.7 (95% confidence interval 31.3 to 36.2), after information 34.8 (32.1 to 37.5); p=0.20, paired t test). A simple explanation of the facts provided a

statistically significant degree of reassurance (mean score at baseline 34.6 (31.5 to 37.6), after information 32.3 (29.8 to 34.9); p=0.012), although this small effect is likely to be clinically important only in those whose baseline anxiety was high (r=0.27, p=0.05).

Conclusions - In men undergoing elective inguinal hernia repair, a very detailed account of what might go wrong does not increase patient anxiety significantly and has the advantage of allowing patients a fully informed choice before they consent to surgery, thus reducing the potential for subsequent litigation.

Extract 3

West Berkshire perineal management trial

J Sleep et al, BMJ 1984; 289: 587- 590.

One thousand women were allocated at random to one of two perineal management policies, both intended to minimise trauma during spontaneous vaginal delivery. In one the aim was to restrict episiotomy to fetal indications, in the other the operation was to be used more liberally to prevent perineal tears. The resultant episiotomy rates were 10% and 51% respectively. An intact perineum was more common among those allocated to the restrictive policy. This group experienced more perineal and labial tears, however, and included four of the five cases of severe trauma. There were no significant differences between the two groups either in neonatal state or in maternal pain and urinary symptoms, ten days and three months post partum. Women allocated to the restrictive policy were more likely to have resumed sexual intercourse within a month of delivery.

These findings provide little support either for liberal use of episiotomy or for claims that reduced use of the operation decreased postpartum morbidity.

EXERCISE 7

From Extract 1, Extract 2 and Extract 3, complete the Table 3, indicating what the 'control' treatment is in each study.

Study	Authors	Description of Control Group
Extract 1: Pressure sores and pressure-decreasing mattresses controlled clinical trial.	*A.Hofman et al* *Lancet 1994; 343: 568-571*	
Extract 2: Who's afraid of informed consent?	*DD Kerrigan et al* *BMJ 1993; 306: 298-300*	
Extract 3: West Berkshire perineal management trial.	*J Sleep et al* *BMJ 1984; 289: 587- 590*	

Table 3.

The placebo

The introduction of placebo-controlled drug trials in the 1980s arose because of the need to account for any improved outcome in a subject arising simply due to the increased attention being afforded them by virtue of being involved in research ('Hawthorne' effect - see also Section 4.1. Page 33). Potentially it could be unclear whether any physiological or behavioural response in a subject was due to the drug treatment or just due to being involved in an experiment. A placebo is an inert substance with no known physiological effects but which looks the same as the active drug. A placebo is given to control subjects taking part in drug trials so that any changes due to the Hawthorne effect can be identified and compared with any 'real' changes due to the drug itself.

EXERCISE 8

Before moving on, read the following description of an experimental study to check your understanding of the terms that have been introduced so far. This intervention study was carried out in 1985 by a research unit based at the University of London. The group wanted to address the question 'does social support given to women during pregnancy improve women's satisfaction after delivery and infant birth weight.

Oakley, A. (1985) Social support in pregnancy: the 'soft' way to increase birthweight?
Social Science & Medicine, 21, 11:1259-1268.

509 women agreed to take part in the study over a 15 month period. They were randomly allocated into two groups. One group received social support by one of four midwives who visited the women at home throughout the women's pregnancy and offered them various forms of practical and emotional help when required. The other group of subjects received no special support, they received only the existing maternity care provisions. Women in both groups had previously given birth to one

low-birth weight baby in the past. Women's satisfaction and infant birth weight were evaluated after delivery using obstetric case note information from the hospitals and by sending the women a questionnaire to complete and return.

1. *What is the independent variable in this study?*
2. *What are the dependent variables in this study?*
3. *Which variables have been controlled?*
4. *Can you think of other variables apart from the independent variable that might influence the dependent variable?*
5. *What name is given to these variables?*
6. *What is the name given to the group who received the standard care?*

2.5 Selection of subjects and random allocation to groups

An experiment is carried out on a **sample** of subjects. Sampling is the selection of a representative portion of the **population**. It is important that the sample is a true representation of the population. The population should be clearly defined using strict criteria decided upon during the planning of an experiment. A researcher should avoid working with an atypical or **biased** sample and ensure that the sample represents the heterogenetic range within the population. Clearly, the larger the sample the more likely it is accurately representing the population. However, in experimental research time, costs and availability of subjects mean that sample sizes have to be relatively small. **Sampling** is a complex topic which is covered in detail in another research pack in this series (See The NIHR RDS EM / YH Resource Pack: *Sampling*). It is often considered essential that any researcher undertaking experimental research needs to ensure that their sample is selected in an unbiased way. **Random sampling** is the most commonly used method of ensuring that an unbiased sample is chosen. The term random in this context means 'by chance' rather than 'haphazard'. In fact, this is a bit of a red herring. All experimental research will define eligibility criteria that determine whether a person can be included in a study (an obvious example is that pregnant women should be excluded from pharmaceutical trials – unless the aim is to treat a condition of pregnancy). After applying these criteria, and excluding those potential research subjects who do not wish to take part, the researcher will clearly have a non-representative sample. However this only affects the **external validity** or **generalisability** of the study results. The internal validity will not be affected, provided there is **random allocation** of subjects to the control or **intervention** (experimental) groups.

The process of randomly allocating subjects to the control or experimental group can be done in many different ways, for example by tossing a coin or using a table of random numbers. The favoured modern method is to allocate using a computer-generated sequence of "pseudo-random" numbers. This is easy in Excel for example by using the rand() function and allocating to treatment A if the random number is <0.5 or treatment B otherwise. In randomised controlled trials more sophisticated versions of this are used in practice. Randomisation ensures that subjects in a control and a treatment group will be similar on average except for the intervention (chance imbalances of characteristics can easily occur in small studies but become less likely the larger the study). Another way of

looking at randomisation is that in Figure 2b it breaks the association between the drug treatment and the confounding variable. The value of randomisation is lost however if the sequence of allocations can be discovered (this is possible if a randomisation list is generated in advance) because then it is possible for investigators to "steer" subjects towards one or the other intervention and thus bias the results. Modern randomisation schemes therefore place as much or more emphasis on concealment of the sequence as on generation of the sequence of random allocations.

Oakley in her study of social support in pregnancy outlines the process and problems she encountered in their study:

> 'Randomisation was done by the midwives telephoning us at TCRU with the names of women who had agreed to take part. The study 'secretary' had sheets of allocations derived from a table of random numbers and she entered each woman in order, then informing the midwife of the result of the allocation.'

The midwives in this study however were often unhappy about this process because the women could not choose whether they received extra support and also because the midwives were concerned about those women who they felt needed more support, being allocated to the control group. Hence there was conflict between the midwives professional opinions and the need to maintain scientific rigour.

Nowadays trials are organised to prevent such problems – in the first place ideally there should be "**equipoise**", for example if opinion among the midwives is evenly divided as to the value of the intervention. If not, or if the midwives could not be persuaded that they should bury their differences in order to obtain a clear answer one way or the other, then the study should not have been done at that Unit. Secondly, letting the study secretary see the list of treatment allocations has the risk that she could disclose an allocation in advance and hence give the midwife the chance to "steer" a patient towards the treatment she favoured. As mentioned above, the modern approach would be to keep the list of allocations secret, eg kept on a computer, so that only the current allocation can be accessed.

A more detailed explanation of randomisation is given in *D.Altman and M.Bland*, "Statistical notes: How to randomise", which may be accessed online with other BMJ articles in the same series at http://www-users.york.ac.uk/~mb55/pubs/pbstnote.htm.

2.6 Informed consent

This topic is covered in more detail in another Trent Focus Resource Pack "*Ethical Considerations in Research*", however it is important for the researcher to be aware of how informed consent specifically relates to experimental research. Prospective subjects must have adequate information about the potential health costs and benefits to themselves before taking part in the experiment, so that they can make an informed choice about whether or not they wish to take part in the study. This principle of **informed consent** is perhaps more pertinent in experimental research (though necessary in other types of research as well) because subjects are exposed to an intervention such as a new drug or procedure. The intervention may be beneficial to the individual, but it may be harmful, if only in terms of costs in time, money or inconvenience.

Informed consent can be sought at different times in the experimental protocol, the issues relating to these different possibilities are discussed in Oakley's 'Who's afraid of the Randomized Controlled Trial". Figure 3 shows one of these options, perhaps the most popular choice. Informed consent is gained immediately following the selection of the subjects. The process of random allocation to control and experimental groups is carried out on those who agree to continue with the experiment.

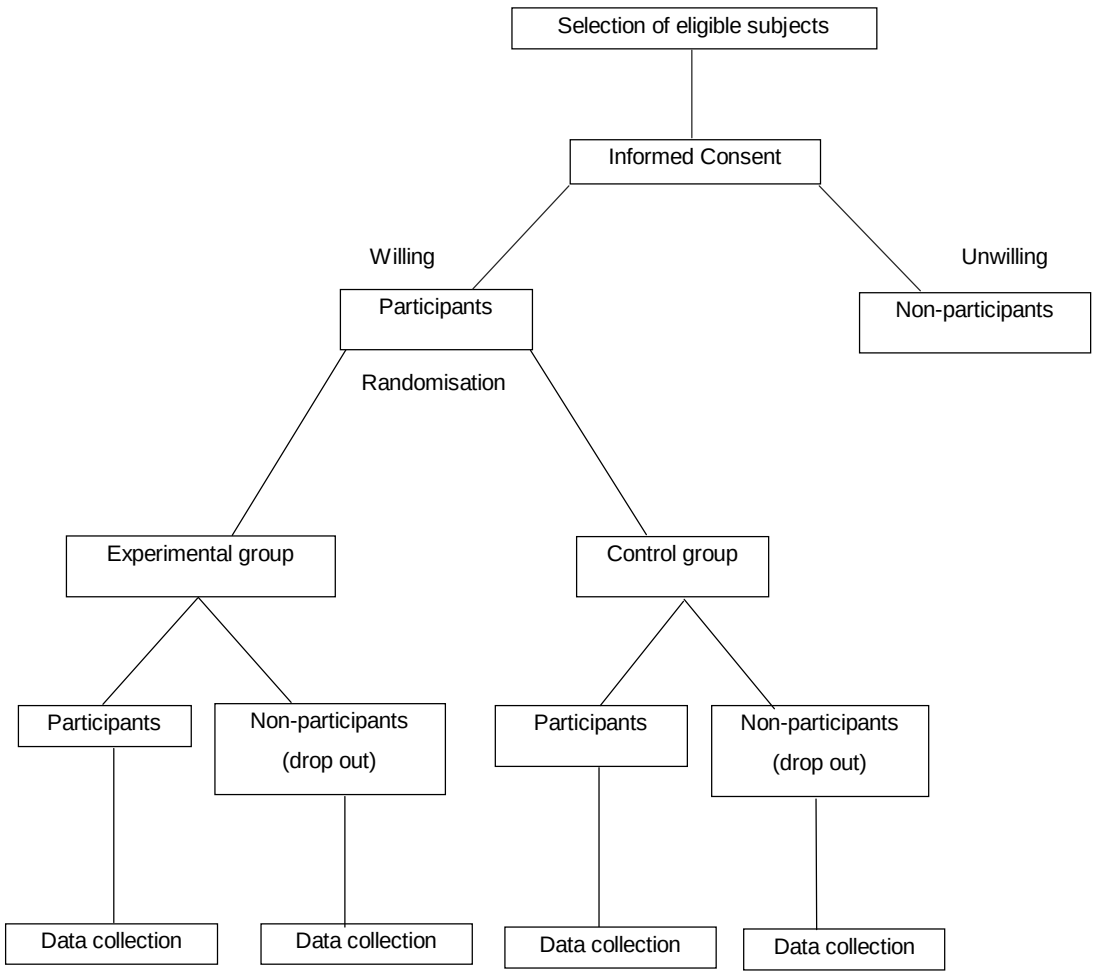


Figure 3.

The amount of information given to the subject varies depending on the research question and the subject's level of understanding. In some experiments, giving too much information can invalidate the study. Note that consent to research differs from consent to treatment.

Example

> Kerrigan, DD et al (1993) Who's afraid of informed consent? *BMJ, 306: 298-300*
>
> 'Patients were simply told that we were aware that emotions could influence how they reacted to illness and were asked to participate in a survey investigating their response to hospital admissions for minor surgery. They were not aware that they were taking part in a randomised comparison of two different information sheets as this would have invalidated the study, which received ethical approval from the Royal Hallamshire Hospital ethical committee.'

The basic principles of informed consent are that:

- participants must "know what they are signing up to" - which includes understanding the possible risks, benefits (if any) and what will happen if something goes wrong
- there mustn't be any coercion or inducement to take part
- participants are free to withdraw consent at any time

EXERCISE 9

Test your understanding of this section by answering the following questions:

1. How important is it to randomly select subjects for an experimental study?

2. Why should subjects be randomly allocated to the experimental and control groups in an experiment?

Summary

When reading literature that reports on a study using an experimental design or when planning your own experiment, you should assess the following in the methods:

- validity and reliability of the dependent variable
- standardisation of the techniques and protocol (independent variable)
- control of confounding variables

Where these haven't been considered in the methods, they should be taken into account when interpreting and discussing the results as shown in the example below.

Example

Specialist nurse support for patients with stroke in the community: a randomised controlled trial*. Forster & Young (1996).*

This was a pragmatic randomised trial with a heterogeneous population including patients with aphasia, cognitive impairment, and diverse ethnic backgrounds. The study needed widespread local support, and therefore knowledge of the trial's objectives may have raised other community staff's awareness of psychosocial problems related to stroke. Moreover, intermittent professional contact with the specialist nurses will have further emphasised this effect. Inevitably there will have been contact between the two groups of patients with stroke at day centres, stroke clubs, or local community gatherings. All these potential confounding factors would have been in the direction of reducing the measured effectiveness of the trial. Lastly, during the 12 months, other significant life events - such as new physical illness or bereavement - took place and are likely to have influenced the psychological state and social functioning of the patients; we found no evidence, however, that such adverse events were more frequent in one or other group.

3. Research designs

This section of the pack will introduce you to a number of common research designs. The previous section described features that needed to be considered when designing experiments. In an ideal world all experiments would be designed so that they were fully randomised and controlled with no confounding variables! In reality of course this is not often possible to achieve and research designs with less rigour are chosen because opportunities arise to obtain data and time, financial constraints or transitory 'ideal' circumstances mean that randomisation or pre-interventioning are not feasible. As long as the limitations of the experiment are recognised and the effect on the validity of the results acknowledged then some reduced rigour is accepted.

3.1 Within subject designs

Single-subject

The simplest possible design is to apply an intervention to a single subject and measure the response. Say you suspected that a drug was causing patient's temperature to increase, you could investigate this further by monitoring the temperature to establish a baseline, lets call this control condition A, then a dose of the drug is given and the temperature continues to be monitored (experimental condition B). If a rise in temperature coincides with the time that the drug is known to take to enter the blood stream then you may surmise that the drug is raising body temperature. This is represented schematically in Figure 4.

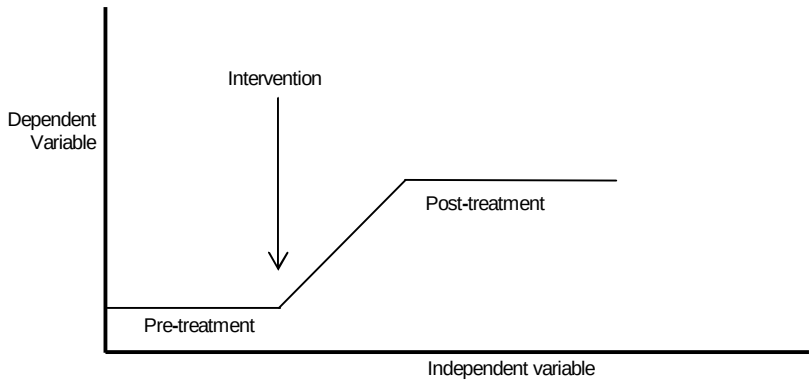


Figure 4. Schematic diagram of a single-subject experimental design.

Following on from this you could then continue to monitor temperature as the drug is eliminated from the body then re-administer the drug when the temperature had returned to baseline levels this then becomes an ABAB design. Studies of single subjects are often referred to as 'N of one' studies.

EXERCISE 10

Continue the graph in Figure 4 to show what you might expect to see in an ABAB design.

A stable baseline (condition A) is essential in single subject designs since, if this is fluctuating considerably, then it is impossible to assess the influence of condition B. The baseline measurement should be made for a long enough time for the researchers to be confident that a baseline level has been determined and reliable. The validity will be improved if the intervention is applied a number of times (i.e. ABABAB). Even better will be if the sequence of treatments is random.

These single subject designs are important in health sciences as in clinical practice settings health care workers are always dealing with individual patients who are unique in the way their condition progresses and the treatment they require. Hence when assessing the results of a particular therapy in a single case these designs are acceptable, the problem arises when a researcher wishes to generalise the results to a wider population. Single subject experiments provide useful preliminary data, which can then be extended to larger group studies in which the treatment can be evaluated.

Within-subject group (Cross-over) designs

A group design involves a group of subjects in an experiment. Each subject is exposed to more than one experimental condition. Table 4 shows a design in which subjects receive a standard treatment (control condition) and then a new treatment.

Subject	Week 1	Week 2
1	standard treatment	new treatment
2	standard treatment	new treatment
3	standard treatment	new treatment
4	standard treatment	new treatment

Table 4.

EXERCISE 11

There is a major flaw in this design. Can you think what it is?

Consider if something happened in week 1, which affected the dependent variable in one or more of the subjects which did not occur in week 2? For example say the dependent variable was body temperature and the independent variable was a new drug treatment or the standard drug treatment (control treatment). If all the subjects were, unknown to them or the researchers, exposed to an infection during week 1, this is likely to raise body temperature. Our baseline measurement would be inaccurate and the comparison with

temperature measured during the new treatment period in week 2 invalid. The term **maturation** refers to incidents such as the infection, which may arise during the course of an experiment, other examples include policy changes, practice effects and developmental changes so that the longer an experiment continues the more likely it is that external events might influence the dependent variable (the more statistical term for this is a **period effect**). This can be overcome by **balancing** the design as shown in Table 5. This is also called a **cross-over design**.

Subject	Week 1	Week 2
1	standard treatment	new treatment
2	new treatment	standard treatment
3	standard treatment	new treatment
4	new treatment	standard treatment

Table 5.

Now any unknown confounding variable affecting the subjects in week 1 will be spread across the treatments and will have less influence on the results. An extension of this type of design is known as a **repeated measures** design; that is, subjects repeat their performance but under slightly different conditions. More than two conditions are allowed in repeated measures designs but it is worth bearing in mind that subjects understandably get bored and tired when participating in long experiments. These **fatigue** effects are more important within subject's designs, so it is more likely that subjects will drop out of the study or exhibit reduced performance. When subjects are lost from an experiment they are known as **drop outs**.

So what are the advantages of using within subjects designs? There is a large variability between subjects both in their physiological responses and behaviour; this source of variation is reduced by using the same subjects in an experiment. Unfortunately it is often not possible to use the same subjects because in many experiments exposure to one intervention means that they cannot participate again. An example would be giving information to a subject about symptoms they may experience on returning home after surgery to see whether it relieved anxiety, a subject could not undertake this twice! In these circumstances a between subjects design is used.

Between subjects

In these designs different subjects are used for the control and intervention groups. In the previous section we talked about random allocation of subjects to ensure that extraneous variables are spread equally between the groups. Variation in subject characteristics should be similar for control and treatment groups, but randomisation cannot be expected to make the groups identical in all respects. The characteristics of all the groups may be compared informally for variables such as age, sex, disease severity and previous exposure to any experience, which might influence the outcome variable and the pre-intervention (**baseline**) values of the dependent variable should also be comparable between the groups. Formal statistical testing should *not* be performed to detect "imbalance" however, because if the

randomisation has worked, any differences will be due to chance (and some – 1 in 20 – will be significant at the 5% level). Secondly, if the randomisation hasn't worked, there are more sensitive methods of detecting this. Thirdly, what actually matters is not the size of any imbalance as such, but the extent to which this will distort the outcome; moreover the outcome may be distorted because of the combination of variables, none of which alone is statistically different between the groups. The solution is to stratify the analysis on these variables or perform statistical modelling (for example multiple regression) – procedures which can deal with any imbalances remaining after randomisation.

Random allocation can eradicate what is termed **systematic bias** since every person has exactly the same chance of being selected for the intervention or control groups, although chance differences may still arise. In the table below which is part of a table in a published paper reporting the effect on mood of lowering cholesterol concentration, you can see how randomly allocating the subjects into the control and treatment group has led to remarkably comparable baseline measures in the groups.

An additional set of procedures to handle baseline imbalance between groups of subjects is called stratification. This can be done in two ways. The first is **stratified allocation** – patients are grouped into categories (**strata**) based on combinations of a variety of attributes such as age, gender and occupation. Within each stratum a separate randomisation is carried out, and to ensure that the proportions on each treatment are balanced within each stratum, the randomisation is usually blocked, ie carried out in small groups (a multiple of the number of treatments) within which each treatment is assigned equally frequently, but in a different order or permutation between each block (hence the term **random permuted blocks**). The second approach is to perform simple randomisation, but group the subjects into the strata ***after*** the data are collected (this is known as **post-stratification**). Either way the statistical analysis should take account of the stratification – by performing a **stratified analysis**, but a post-stratified analysis loses surprisingly little compared with stratified allocation, provided the study is reasonably large. A more modern approach, would be to perform statistical modelling, using the stratification variables directly as covariates (which avoids the need to create strata altogether).

Stratification also allows separate sub-group analyses to be carried out.

Example

	Simvastatin 20mg or 40mg (n=334)	Placebo control (n=157)
Mean age (years)	63.3	63.8
Mean total cholesterol (mmol/l)	7.0	6.9
Mean low density lipoprotein cholesterol (mmol/l)	4.8	4.7
Mean low density lipoprotein cholesterol (mmol/l)	1.17	1.14
Mean triglycerides (mmol/l)	2.52	2.59

Table 6. Wardle et al (1996) Randomised placebo controlled trial of effect on mood of lowering cholesterol concentration. BMJ, 313, 75-78

3.3 Matched designs

The **matched pair** design also uses different subjects in the groups. Random allocation does not guarantee that the groups are equivalent in a particular study but only that, on average, they will be similar. In matched pair designs the researcher first identifies characteristics that might influence the outcome and then selects subjects who are matched for these characteristics, matching for age and sex is common. Random allocation to the control or treatment groups occurs after the matching procedure. This is effectively a very "tight" stratification procedure.

Example

> Specialist nurse support for patients with stroke in the community: a randomised controlled trial *Foster & Young (1996).*
>
> Once recruited, the patients were stratified by whether they had been admitted to hospital or stayed at home, by level of social activities before their stroke (Frenchay activities index, categories 0-10, 11-30, 31-45), and by functional ability (the Barthel index, categories 0-9, 10-14, 15-19, 20).

The allocation of subjects to treatment or control groups in health research raises a number of ethical issues. In particular a researcher may be concerned when a subject is denied a potentially beneficial treatment if allocated to the control group (remember the midwives' concerns in Section 2.5).

EXERCISE 12

From the Extracts from research papers below, decide what type of experimental design has been used in these studies.

1. Pads and pants for urinary incontinence

> 51 female patients, incontinent of urine, were asked to compare 2 different combinations of pants and pads used in ambulatory management of their incontinence

2. Who's afraid of informed consent?

> Preoperative anxiety assessed before and after patients were randomly allocated an information sheet containing either simple or detailed descriptions of possible post-operative complications

3. Specialist nurse support for patients with stroke in the community.

> To evaluate whether specialist nurse visits enhance the social integration and perceived health of patients with stroke or alleviate stress in carers in long term stroke

3.4 Variations in experimental design

Pre-intervention/post-intervention designs

In the classical experiment the dependent variable is measured both before and after the intervention or treatment in a control or experimental group. The pre-intervention measurement gives what is often called a baseline measure of the dependent variable. This allows two comparisons to be made in the analysis. Firstly, the change from baseline in the control and experimental groups and secondly, a comparison of post-intervention measures of the dependent variable in the control and experimental groups. Don't waste time doing significance tests on the change from baseline in each group! Only the comparison of the response between the groups gives information about the possible cause-effect relationship of the independent and dependent variables. If it's possible that the baseline value may affect the outcome then measure the final response, with the baseline itself as either a stratification variable or a covariate in a statistical model.

Example

In a trial studying the effects of weight-bearing exercise on bone in post-menopausal women, McMurdo et al (1997) compared a group taking calcium supplements only with a group taking calcium supplements and exercise. The bone mineral content was measured at 2 sites in the forearm and in the lumbar spine both at entry into the study (baseline or pre-intervention measure) and after 2 years. Table 7 shows some of the results they obtained.

Site	Mean % change	
	Calcium group (n=48)	Calcium and exercise (n=44)
Ultradistal forearm	-2.6	1.14
Distal forearm	-1.38	-2.18
Lumbar spine	-2.65	-0.91

Table 7. Change in bone mineral content and density over 2 years of study.

Pre-intervention bone density has been subtracted from the post-intervention values to allow a percentage change in bone density to be calculated. This percentage change can then be compared in the control (Calcium only) and experimental (Calcium and exercise) groups. Statistical analysis would then show whether the percentage changes and the comparisons between groups were statistically significant.

Stage:	1	2	3
Experimental Group 1	pre-intervention measurement	intervention	post-intervention measurement
Experimental group 2	pre-intervention measurement	intervention	post-intervention measurement

Table 8.

Post-intervention only designs

Sometimes it is not appropriate to do a pre-intervention measurement and just post-intervention measurements are carried out in control and treatment groups, the post-intervention results from the two groups are then compared. This is potentially a weaker design and difference between groups may have been present before the intervention. However if the study is "large" and randomised, this effect is likely to be small.

EXERCISE 13

Read Extract 3 in Section 2.4 describing **West Berkshire perineal management trial,** it is easy to see why a pre-intervention measurement was not possible in this trial! Can you identify all the post-intervention measurements (outcome measures) made by the researchers mentioned in this abstract.

Number of dependent variables

It is also permissible to measure more than one dependent variable, for example, in a pressure sore experiment, the researcher might want to measure number of sores, grade of sores and size of sores. This increases and complicates the statistical analysis of the data however and subjects may get fatigued if many measurements are made on them.

Number of independent variables

It was stated in the introduction that the researcher should not be tempted to try to answer too many research questions within one experiment. This implies that one should not manipulate more than one independent variable at a time. Sometimes a researcher does want to study the individual and combined effects of two or more independent variables. Such designs are called factorial designs. If the two independent variables have two levels (See the example below) the design is called a 2 x 2 factorial design and four groups of subjects would be needed. Factorial designs permit the testing of multiple hypotheses in a single experiment, but in a way that is totally distinct from the commoner practices of having multiple study endpoints and/or subgroup analyses.

Example

Melnyck et al (1994) studied the effect of receiving or not receiving two types of information on mothers' and children's' ability to cope with an unplanned childhood hospitalisation. The design was as shown below (Table 9) with the number of subjects receiving neither, one, or both sets of information shown in the cells:

		Child behavioural information	
		Received	Did not receive
Parental role information	Received	27	22
	Did not receive	26	23

Table 9.

The two independent variables are child behavioural information and parental role information and there are two levels of measurement for the two independent variables: receiving and not receiving the information.

Factorial designs aimed at getting "two studies for the price of one" require two unrelated questions and two essentially different diseases, which can still be studied in the same individuals. For instance The Physicians' Health Study was a 2x2 factorial trial of aspirin and beta-carotene among US physicians. The aspirin was investigated for its effect on cardiovascular disease, and the beta-carotene for its effect on cancer risk (a useful overview can be found at http://phs.bwh.harvard.edu/phs1.htm#ppt). The assumption (which may not be true) is that the intervention of one study will not affect the outcome of the other (that is, there will be no **interaction** between the treatments). Alternatively two treatments for the same condition may be studied, alone and in combination as in Table 9. An interaction may exist, but such studies are generally underpowered to detect it, partly because the precision of an interaction is much less than that of the original comparison and partly because interactions are generally relatively small. Studies actually aimed at finding an interaction must therefore be specifically designed for this, and tend to be larger.

A factorial design may detect interactions between treatments, but interactions between treatments and patient characteristics don't require a factorial design. Age, ethnicity, sex and treating hospital are all factors which might modify the therapeutic effect of a drug, ie interact with it. Sometimes in clinical trials, investigators look for interactions by analysing subgroups, but it is better to do a formal statistical test for interaction. Interactions may be found in observational studies too.

Example of sex by treatment interaction. The Canadian Cooperative Study Group found that aspirin *reduced* the risk of stroke in men by 48% but *increased* it in women by 42%. The response to treatment is clearly different. However there is another lesson here: some findings are spurious even though statistically highly significant and in fact additional trials and meta-analyses eventually showed that women too benefit from aspirin.

Community Trials

A community trial is an experimental study conducted on a whole community rather than on a small sample. The best example to use is the study done by the British government on the fluoridation of water. In 1955, the British government fluoridated the water supply in Watford, Kilmarnock and parts of Anglesey and measured the proportion of children with caries free teeth, so for example in Holyhead (fluoridated) and Bodafon (non-fluoridated) the proportions, before fluoridation, were 13% and 12% respectively. 10 years later after fluoridation in Holyhead, they did a post-intervention measurement of the proportion of children with caries free teeth, the proportions were 40% for Holyhead and 14% for Bodafon.

The above is an example of a **cluster randomised** trial – the clusters may be whole communities, or they may be groups of individuals defined in other ways (eg patients in a hospital ward, patients with a certain condition attending a GP). The allocation to treatments is carried out at the group level – thus wards within a hospital may be randomly allocated to a one or another hand-washing protocol, or GP practices may be allocated randomly to receive a pack on management of diabetes.

If a cluster randomised trial consists, say, of 20 hospital wards each with 30 patients the analysis must take account of the fact that patients within each ward are likely to be less variable than the patients taken as a whole. A simple example is if the wards are single sex – between wards the proportion of women might be 50% but within a ward it will either be 100% or 0%. A similar but less extreme similarity is likely with respect to other characteristics

such as age, disease severity and so forth. The extent of the similarity can be assessed by a quantity called the ***intraclass correlation***. The implication is that you should not analyse the 600 patients as if they were individually randomised to treatments, because the effective sample size is smaller than it appears and this increases the variance of quantities estimated from the data. There are several ways of doing an analysis that allows for the intraclass correlation but these will not be discussed here.

For cluster randomised studies, the sample size for the entities randomised may not be very large, but the sample size in terms of the total number of individuals involved in the study, may be considerably larger than needed if the individuals were directly randomised (this is the converse of the problem of analysing such a study). The correction factor is obtained by multiplying the initial sample size estimate by a quantity called the ***design effect***, given by $[1+(n-1)\rho]$, where ρ is the intraclass correlation and n is the average number of subjects per cluster. Because in a cluster design the number of subjects per cluster can be large, the effect on the sample size or the analysis can be large even though intraclass correlations (which are not the same as an ordinary correlation) tend to be surprisingly small (often around 0.05).

Cluster randomised studies do have certain advantages however (especially if the intervention is a policy or involves behavioural/lifestyle change which could be copied by participants in the control arm).

Summary

When planning an experiment or reading about experimental research you need to consider whether the design chosen is appropriate for the research question posed. A researcher may choose:

- a between subjects, within subjects or matched pair design
- a pre-intervention/post-intervention design or a post-intervention only design
- a factorial design when the effect of individual or combined effects of more than one independent variable are to be investigated

4. Benefits and limitations of experimental research

4.1 Introduction

The advantages and limitations of experimental research have been implied throughout the previous sections of this pack. This section will provide a summary of the main benefits and limitations of experimental research and when this method should be used. The special considerations relating specifically to human experimentation can be grouped as factors relating to causality, validity and humanity and ethics.

Causal relationships between variables

The previous sections have identified three features that characterise a true experiment. In experimental research, the researcher:

- manipulates the experimental situation by systematically varying the independent variable and measuring the response in the dependent variable
- introduces some control over the experimental situation by eliminating the influence of variables other than the independent and dependent variables
- randomly selects and allocates the subjects to a control group (no treatment or standard treatment) and experimental group (treatment)

Experimental research is the most powerful method for inferring causal relationships between variables. This is because, in theory, the researcher has eliminated all possible factors that could account for a change in one or more outcome (or dependent) variable(s) other than the influence of one or more explanatory (or independent) variable(s).

Sometimes it is not possible to fulfil all these requirements so for example it may be unethical to have a control group or it may not be possible to carry out randomisation. Such experiments are then called **quasi-experiments**. In Woolf's hierarchy of evidence (Figure 5), well-designed randomized control trials (true experiments) are considered the most useful evidence for establishing causal relationships with experiments with less control (quasi-experiments) lower down the hierarchy. **Observational studies** such as case-control and cohort studies are quasi-experiments which though not randomised, do include a control group. The cohort study can be thought of as an experiment in which the allocation is made by choice or circumstance rather than by randomisation. These studies are not within the remit of this pack but an excellent source is the book by dos Santos Silva (See Further Reading). Don't be put off by its title.

I	Well designed randomised controlled trials
II-1	Other types of trial; well designed controlled trial without randomisation; quasi experiments
II-2	Well designed cohort (prospective) study, preferably from more than one centre
II-3	Well designed case control (retrospective) study, preferably from several centres
III	Large differences from comparisons between time and/or places with or without intervention
IV	Opinions of respected authorities based on clinical experience; descriptive studies and reports of expert committees

Figure 5. Woolf's hierarchy of evidence.

Some caution should be exercised when inferring cause-effect relationships between the independent and dependent variables. Few non experimental studies are so well designed that the researchers are 100% confident that the change in the independent variable **caused** the change in the dependent variable. More often researchers acknowledge an **association** or **relationship** between the variables; causality must be inferred (never proven) and many studies may need to be carried out for causality to be generally accepted if the effect is small.

Internal and external validity

Of all the research methods, experimental research is said to have high **internal validity** because it is a more controlled approach, any factor that interferes with the design or implementation of an experiment potentially threatens the internal validity of that experiment. To check whether an experiment has internal validity you should ask the question 'are the changes in the dependent variable only due to the intervention (the independent variable) and not due to other factors?'

EXERCISE 14

Think back over what we covered in Section 1. Before reading further, write down some of the features of experimental design that ensure high internal validity.

Some of the factors you might have written down are:

1. Control of confounding variables

2. Reliable instruments

3. Appropriate choice of independent and dependent variables

4. Random selection to groups

5. Standardisation of protocol

6. Minimising factors such as maturation, history and mortality

EXERCISE 15

Would a quasi-experiment have a higher or lower internal validity than a true experiment? Explain why and add 'true experiment' and 'quasi-experiment' to Figure 6.

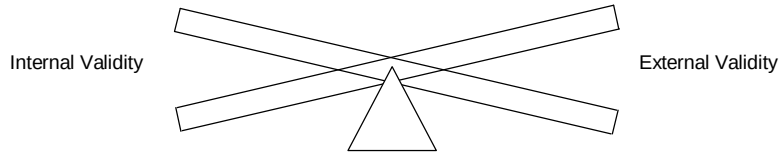


Figure 6.

Some researchers believe that the true experiment produces an unreal situation because of the strict control over conditions, a situation that would not normally occur in everyday life, hence results from experiments are not generalisable to the wider population. **External validity** answers the question 'can the results be generalised to the wider population?'

External validity can be improved by **replicating** experiments that is repeating the experiment under similar conditions. Researchers might wish to use a slightly modified research tool, a different setting or subjects with slightly different characteristics; all of these changes would increase the generalisability of the research and improve the external validity.

Humanity and ethics

The researcher using human subjects in experimental research may frequently come across situations where ethics and considerations for fellow human beings conflicts with the strict control required in the methodology. Experiments involving human subjects who do not behave as predictably as inanimate objects, and are less compliant, require some special considerations listed in Table 10.

1	The need to protect human rights
2	The importance of preventing unnecessary suffering and risk to subjects
3	Subjects should give informed consent even though this may affect the results and conclusions (See Section 2.6)
4	The need for subjects to comply with protocol while recognising the importance of 'free will'
5	Being objective when studying fellow human beings whom the researcher might like or dislike
6	Difficulty of measuring dependent variables when they concern subjective & complex emotions or behaviours eg stress, pain, intelligence (See Section 2.3 - Validity)
7	Ensuring all variables are held constant except the independent variable(s) (See Section 2.4)
8	Withholding a possible beneficial treatment from clients in the control group (See Section 2.4 and 2.5)
9	Subjects' behaviour may change when they know they are being studied ('Hawthorne effect')
10	Artificiality of human behaviour in experiments compared to 'real life' (See above)

Table 10. Considerations when designing experiments using human subjects

Some of these issues have been discussed in other parts of the pack. Ethical issues need to be considered when planning and implementing any research, this topic is covered in another pack in this series (See The NIHR RDS EM / YH Resource Pack: *Ethical Considerations in Research*).

EXERCISE 16

Check your understanding of this section by listing 3 advantages and 3 disadvantages of experimental research.

4.2 When to use experimental designs

We have considered some of the advantages and disadvantages of experimental research throughout this pack and by now you should have some idea of when to use an experimental approach to research. If your research question is fairly specific and you want to compare the effectiveness of one or more treatments (or interventions), then the experimental approach is the best to use. Other methods such as observational, surveys and qualitative approaches should be used when the research aims are broader and an overview of opinions or behaviour is required. The validity of the research will be threatened if an experiment is badly designed or inappropriate for the research question for example if it is not possible to impose the necessary control and randomisation which are fundamental to experimental design perhaps for ethical or practical reasons. Reading through studies that have used an experimental approach will help you to understand the type of research questions requiring an experimental approach. Examples of a range of experimental studies can be found in the reference list at the end of this pack.

Answers to exercises

Exercise 1	Dependent variable - pressure sore grade
	Independent variable - mattress type
	Effect parameter – the abstract doesn't give one, but possibilities are a) the difference between the average of the maximum score/person in the two groups b) the proportion of patients with a grade 2 or above pressure sore
Exercise 2	'Our study was designed to determine the effectiveness in pressure sore prevention of a new interface-pressure decreasing mattress.'
Exercise 3	There is no difference in the effectiveness in pressure sore prevention of the new interface pressure decreasing mattress and the standard mattress
Exercise 4	Potentially there are many sources of systematic error; however some you may have identified are:-
	Uncalibrated sphygmomanometer which always reads higher or lower than the true pressure
	Not taking account of the mercury meniscus when reading the pressure
	Sphygmomanometer not zeroed
Exercise 5	Feedback in text
Exercise 6	psychological condition
	age
	sex
	mobility
	nursing/medical care
	nutrition
	medical/surgical intervention
	All these factors are potentially confounding variables. The authors have said in the abstract however that both groups were treated using a standard protocol for the prevention of pressure sores and that the two groups were similar in patient characteristics and pressure sore risk factors

Exercise 7 Either of these treatments could be considered to be the 'standard' policy. This will depend on the individual midwives and obstetricians clinical judgement.

Study	Authors	Description of Control Group
Extract 1: Pressure sores and pressure-decreasing mattresses controlled clinical trial.	*A.Hofman et al* *Lancet 1994; 343: 568-571*	nursed on a standard mattress
Extract 2: Who's afraid of informed consent?	*DD Kerrigan et al* *BMJ 1993; 306: 298-300*	allocated simple information sheet
Extract 3: West Berkshire perineal management trial	*J Sleep et al* *BMJ 1984; 289: 587- 590*	liberal or restrictive policy *

Table 3.

Exercise 8

1. Social support
2. Womens satisfaction & infant birth weight

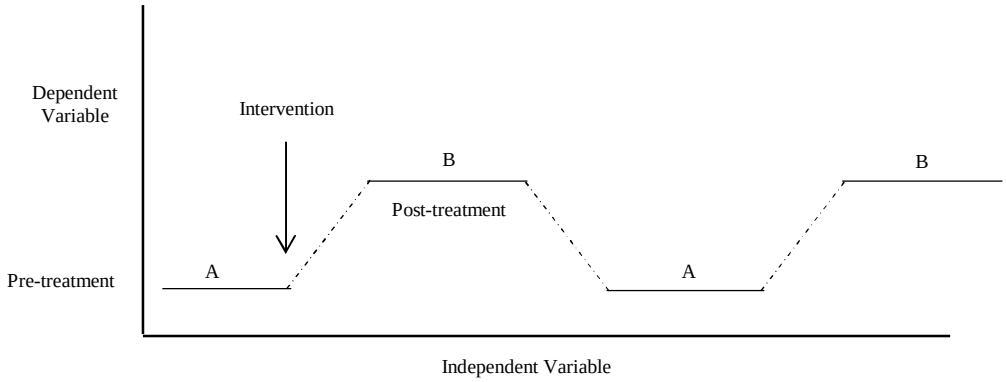
All women had given birth to one low birth weight baby. There are many confounding variables in a study of this kind which is why a qualitative approach is often preferred for this type of research question. Some of the confounding variables you might have identified are:

- some mothers may have received social support from other sources eg relatives and friends
- women in the control group may have felt more supported just because they were taking part in a research project (the 'Hawthorne effect' see Section 4.1).

Exercise 9

- Random selection may help avoid selecting a biased or atypical sample from a population which will reduce generalisability, but it will not affect the internal validity of an experiment
- To increase the likelihood that the control and experimental groups will be comparable in terms of subject characteristics and baseline measures.

Exercise 10



Exercise 11 Feedback in text.

Exercise 12
1. Within subject design
2. Between subject design
3. Between subject design

Exercise 13 Dependent variables (outcome measures)
Perineal and labial tears
Neonatal state, maternal pain and urinary symptoms at 10 days and at 3 months
Likelihood of resumption of sexual intercourse within 1 month

Exercise 14 Feedback in text.

Exercise 15 Lower internal validity; because there are less strict controls in the design of quasi-experiments

Exercise 16
- Advantages: High internal validity, researcher has more control over the research; specific relationships between variables can be studied.
- Disadvantages: Low external validity; only consider a relatively small number of variables; ethical and practical difficulties of imposing control and randomisation

References

Bainton, D., Blannin, JB., Shepherd, AM. (1982) Pads and pants for urinary incontinence. BMJ, 285; 419-420.

Boore, JRP. (1978) Prescription for recovery: The effect of pre-operative preparation of surgical patients on post-operative stress, recovery and infection. Royal College of Nursing.

Clegg, F. (1994) Simple Statistics. 12th edition Cambridge University Press.

Forster & Young (1996) Specialist nurse support for patients with stroke in the community: a randomised controlled trial. BMJ, 312: 1642-1646.

Hofman, A., Geelkerken, RH., Wille, J ., Hamming, JJ., Breslau, PJ. (1994) Pressure sores and pressure-decreasing mattresses: controlled clinical trial. Lancet, 343: 568-571.

Kerrigan, DD., Thevasagayam, RS., Woods, TO., McWelch, I., Thomas, WEG., Shorthouse, AJ. & Dennison, AR. (1993) Who's afraid of informed consent? BMJ, 306: 298-300.

McMurdo, MET., Mole, PA & Paterson, CR. (1997) Controlled trial of weight bearing exercise in older women in relation to bone density and falls. BMJ, 314: 569.

Melnyck, B.M. (1994) Coping with unplanned childhood hospitalization. Effects of informational interventions on mothers and children. Nursing Res., 43: 50-55.

Melzack, R (1975) The McGill pain questionnaire: major properties and scoring methods. Pain, 1: 277-299.

Melzack, R., Abbott, FV., Zackon, W., Mulder, DS. & Davis, MWL. (1987) Pain on a surgical ward: a survey of the duration and intensity of pain and the effectiveness of medication. Pain, 29, 67-72.

Oakley, A (1990) Who's afraid of the Randomised Controlled Trial? In: Women's Health Counts. Ed. Helen Roberts, publ. Routledge.

Oakley, A. (1985) Social support in pregnancy: the 'soft' way to increase birth weight? Social science & Medicine. 21, 11:1259-1268.

O'Brian, D. & Davison, M. (1994) Blood pressure measurement: rational and ritual actions. Br.J.Nursing, 3: 393-396.

Roe, BH. (1990) Study of the effects of education on patients knowledge and acceptance of indwelling urethral catheters. J Adv.Nursing, 15: 223-231.

Rose, G. & Hamilton, PJS. (1978) A randomised controlled trial of the effect on middle-aged men of advice to stop smoking. J Epid. & Comm. Health, 32: 275-281.

Sleep, J., Grant, A., Garcia, J., Elbourne, D., Spencer, J. & Chalmers, I. (1984) West Berkshire perineal management trial. BMJ, 289: 587- 590.

The Canadian Cooperative Study Group, (1978) A randomised trial of aspirin and sulfinpyrazone in threatened stroke. N Engl J Med 299: 53–59.

Wardle, J., Armitage, J., Collins, R., Wallendszus, K., Keech, A. & Lawson, A. (1996) Randomised placebo controlled trial of effect on mood of lowering cholesterol concentration. BMJ, 313: 75-78.

Wilson-Barnet, J. (1991) The Experiment: Is it worthwhile? Int.J.Nurs.Stud., 28:77-87.

Woolf SH et al (1990): Assessing the clinical effectiveness of preventive manoeuvres: analytical principles and systematic methods in reviewing evidence and developing clinical practice recommendations. J Clin. Epidemiol., 43, 891-905.

.

Further reading and resources

dos Santos Silva I (1999) Cancer Epidemiology: Principles and Methods. Lyon. IARC (ISBN 92 832 0405 0).

Polgar, S & Thomas, SA (1995) Introduction to research in the health sciences (3rd Edition) Churchill Livingstone.

Polit, DF & Hungler, BP (1995) Nursing Research - Principles and Methods (5th edition) JB Lippincott & Co.

Rose, G. & Hamilton, PJS. (1978) A randomised controlled trial of the effect on middle-aged men of advice to stop smoking. J.Epidem. & Comm. Health, 32: 275-281.

Sapsford, R & Abbott, P (1992) Research methods for nurses and the caring professions. Oxford University Press.

Wilson-Barnet, J. (1991) The Experiment: Is it worthwhile? Int.J.Nurs.Stud., 28:77-87.

Helpful discussions of many of the statistical points raised can be found in the Statistics notes series published in the British Medical Journal. These can be readily accessed at http://www-users.york.ac.uk/~mb55/pubs/pbstnote.htm

Glossary

Create your own glossary of terms as you work through the pack. All the words shown below appear in bold in the text.

baseline

between subjects design

bias

carry-over effects

cluster randomised

community trials

content validity

control

confounding variables

control group

counterfactual

cross-over

dependent variables

design effect

double-blind

effect parameter

equipoise

external validity

extraneous variables

face validity

fatigue

generalisability

Hawthorne effect

independent variables

informed consent

interaction

internal validity

intervention

intraclass correlation

hypothesis

matched pair design

mortality

multi-centred RCT.

null hypothesis

observational studies

period effect

pilot study

placebo effect

population

post-stratification

potency

pre-intervention/post-intervention designs

post-intervention only designs

qualitative

quasi-experiments

random allocation

random permuted blocks

random sampling

randomised clinical (controlled) trials (RCTs)

reliability

repeated measures

replicating

reproducible

single-subject designs

sample

sampling

sensitive

standardise

strata

stratified analysis

surrogate variable

systematic bias

systematic error

treatment

validity

variables

within-subject