RESEARCH SERIES

Overview of research designs

M J Clancy

....

The fourth paper in the research series looks at the difficulties in establishing truth for the researcher and provides an overview of research design.

....

The aims of this paper are twofold. Firstly, to describe the obstacles that lie in the way of the researcher establishing the truth and how these may be dealt with in terms of research design. Secondly, to provide an overview of the different designs and the trade offs involved.

The aim of research should be to establish the truth and research design aims to minimise or exclude the threats to the internal validity of the study (that is, that the conclusions are warranted by the observations). These threats are bias, confounding, and chance.

BIAS
Bias is a systematic deviation from the truth that distorts the results of research.[1] Bias can occur anywhere within the research process and here are some common sources.

Selection bias
This occurs when the study subjects differ systematically from the population with the same condition. For example, subjects who present to hospital may not be representative of all patients with the condition and this affects the ability to generalise the result outside the study sample (that is, external validity). Similarly, those who volunteer for studies are different from those who refuse usually being healthier and better educated than the population as a whole.

Intervention bias
For example, the greater use of diagnostic or treatment procedures on the favoured arm in a trial may overestimate the benefit of the intervention. Conversely those patients who are poorly compliant with their intervention reduce the chances of it being effective.

Follow up bias
Those patients who remain in a study may differ from those lost both in terms of personal characteristics and outcome status. Those lost may have died or not wished to be followed up because of the treatment.

Measurement and information bias
This entails misclassifying according to disease or exposure or both. Thus if an investigator knows of the exposures or treatments received this could influence his assessment of the outcome. Similarly, knowledge of the outcome may influence his assessment of exposure.

....

Correspondence to:
M J Clancy, Emergency Department, Southampton General Hospital, Tremona Road, Southampton SO16 6YD, UK
Mr M Clancy; janice.mills@suht.swest.nhs.uk
....

Some 35 types of bias have been described and the interested reader should consult the paper by Sackett.[2] The key to decreasing bias is to identify the possible areas that could be affected and to change the design accordingly. Increasing the sample size will not reduce bias. Bias is an issue of study design. Thus randomisation in interventional studies should avoid selection bias but does not protect against the other types of bias. Blinding the subjects, researchers, and statisticians can reduce bias when knowledge of a subject's treatment/exposure or case-control status may influence the results obtained. Measurement bias can be reduced by the use of repeated measures, training of the researchers, using as objective as possible measures, and using more than one source of information. Follow up bias can be reduced by using multiple methods of contact (mail, telephone, etc) and performing sensitivity analyses. Here the missing group are assumed to all have a good or a bad outcome and the impact of these assumptions on the outcome is evaluated.

CONFOUNDING
Confounding may be defined as a situation in which the effects of two processes are not separated so that the apparent effect is not the true effect and our interpretation of the results is likely to be faulty. From an epidemiological perspective it is the failure of a crude (or partially adjusted) association to reflect properly the magnitude and sometimes direction of an exposure effect because of a different distribution of extraneous risk factors among exposed and non-exposed subjects. The relation between exposure and disease may actually be attributable to another variable that is associated with both exposure and disease. Alternatively failure to control for an extraneous variable may be responsible for the apparent lack of association (see fig 1).

Lecky and Driscoll[3] provide the useful example of confounding by way of work showing that death from trauma is more likely if you are treated by a consultant than a junior house officer. If this crude unadjusted information is taken at face value erroneous conclusions may be made. However, this association is confounded by the severity of illness factor that is associated both with increased mortality and a senior doctor being involved in the patient's care.
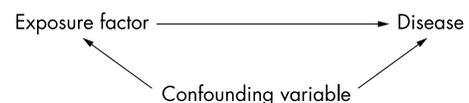

Figure 1

Figure 2

Figure 2 Study designs flow chart. Study designs branch into: Descriptive, Evaluative studies, and Qualitative research. Descriptive: Case report, Case series, Time series, Ecological studies; and Cross sectional studies, Design surveys, Surveillance studies. Evaluative studies branch into Observational (Cohort, Case-control) and Experimental (Randomised/non randomised, explanatory/pragmatic). Qualitative research: Opportunistic.
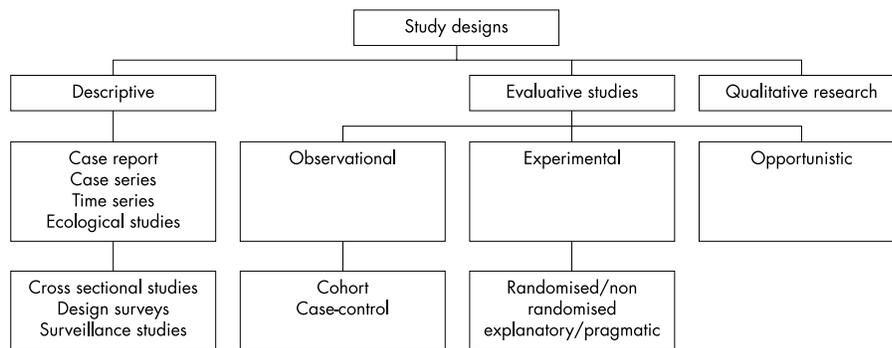
With randomised controlled trials (RCTs) the balance of known and unknown extraneous factors that could confound the assessment of the relative effectiveness of one service over another can be approximately balanced particularly in large trials. An important fact not widely appreciated is that the effects of any maldistributions that do occur as a result of chance are automatically included in the statistical tests of the likelihood that chance is responsible for the overall difference in outcome between the randomly assigned groups.

Confounding is a potential pitfall in all types of observational designs and may be partially controlled in either the design or analytical phase of the study where the aim is to make the groups as similar as possible with respect to the confounder. This obviously depends on the confounder(s) being identified, usually from previous studies or because it is biologically plausible. In the design phase the patient groups may be restricted. Here the level of the confounder is kept constant in the groups by selection of the subjects. This improves the internal validity of the study but limits its generalisabilty. Alternatively subjects may be matched, which entails selecting for each case one or more controls with the same value of the confounder. The ability to generalise is preserved and the power of the study improved but at a price. Matching is expensive, unmatched subjects are excluded, and the effect of the matched variable on outcome cannot be studied. Matching should be used sparingly as it is irreversible and should be used to control for very strong confounders or confounders that are more easily matched than measured.

In the analysis phase, provided the requisite data are collected, (this must be considered in the design phase) an attempt can be made to adjust for confounding. The following methods can be applied to both RCTs and observational studies. *Stratification* is where subjects with similar characteristics or similar levels of a confounder are grouped and the relation between exposure and outcome within each stratum estimated separately. However, the number of confounders that can be controlled for simultaneously is limited as the number of strata multiply leaving small numbers in some strata. *Statistical modelling* using multiple linear and logistic regression enables adjustment for several confounders simultaneously and although software packages have made this easier, inappropriate use of these methods is more likely.

The effects of confounders in observational studies do not diminish with increasing sample size in contrast with RCTs, which may be made sufficently large to reduce the possibility of chance imbalance (see references Datta,[4] Grisso,[5] and Brennan and Croft[6] for more detail on confounders).

CHANCE

Chance can create very different groups even though they have been randomised. For example in a trial including 36 patients even when we have 18 in each group any characteristic that is present in half the subjects has a 6% chance of being at least twice as common in one treatment group as another.[7] Such imbalance of confounders could have a pronounced effect on the results. This can lead to an imprecise estimate of effect but the imprecision can be statistically described. The effect of chance can be diminished by recruiting larger numbers or by adjustment for known confounders in the analysis.

STUDY DESIGN OVERVIEW

The broad categories of study design are shown in figure 2. Descriptive studies though useful in terms of generating hypotheses cannot test them and are not the focus of this paper.

The aim of evaluative studies is to determine the existence and strength of a possible association between an exposure/intervention and an outcome. The essence of evaluative studies is comparison, which is crucial in reaching conclusions about what is normal or abnormal or in determining whether a treatment improves the course of the disease.

In experimental studies the investigator makes some change to the study population and collects data on the outcomes of that change. This is the primary way of studying new treatments. RCTs are seen as the gold standard and are dealt with in detail by Kendall in this series.[8] However, they are expensive and may lack generalisabilty because of exclusion criteria. They can also be somewhat artificial being explanatory (that is, providing evidence of what can be achieved in the most favourable circumstances) rather than pragmatic, reflecting the real world of health care.

Experimental non-RCTs/community trials are studies where the experimental units (typically large general practices) are assigned non-randomly to different types of services. They are often externally valid but because of bias and confounding not internally valid. The North Staffordshire Trauma centre trial would be an example of this type of study design.[9]

The treatment of RCTs as the gold standard for the evaluation of treatments when they are appropriate, practical, and ethical has led to the denigration of non-experimental methods. However RCTs may not always be the best method and observational studies are needed to evaluate the parts RCTs cannot reach.[10] RCTs may be:

1 *Unnecessary* when the effect of an intervention is dramatic and the likelihood of confounders small.

2 *Inappropriate*. They are rarely large enough to measure accurately infrequent outcomes or interventions designed to prevent rare events. For those events that occur in the future RCTs are also inappropriate. In those trials where the effectiveness of the intervention depends on the subjects beliefs and preferences the very act of randomisation may reduce the effectiveness of the intervention.[9]

3 *Impossible* because of the reluctance of clinicians to participate, ethically unacceptable, political issues are a bar to a trial, or contamination is a problem that would require such a large trial to overcome it that it would not be feasible.

4 *Inadequate*. Usually RCTs are internally valid but lack external validity because the health care professionals, patients, or

the setting of the trial are atypical or patients may receive better treatment simply because they are in a trial.

With observational studies the investigator does not control the treatment or exposure but attempts to make valid comparisons between people with or without diseases or between those naturally exposed or unexposed to a factor of interest. These types of study are comprehensively covered by Mann in this series[11] but some features bear repetition. Cohort studies entail observing two or more groups differing in exposure to a potential cause of disease overtime to compare the incidence of disease in each group. This type of study is prospective with the exposure measured before outcome, enabling the direction of events to be established that is necessary if we want to say anything about causation. Only one risk factor can be assessed for each study but multiple outcomes can be measured. Typically these studies require large numbers, especially if the incidence of the disease/outcome is low and a long time scale. Hence they are likely to be expensive.

With a case-control study the comparison is made between people with and without the disease in order to identify differences in previous exposures in an attempt to identify the cause of the event. Such studies are retrospective, multiple exposures can be assessed, and they are the primary method of studying new or unusual outcomes. Although they are quick and involve small numbers, defining the control group can be difficult. There is a risk of recall bias by the patient and measurement bias by the investigator if he is aware of the outcome.

The principal problem of observational studies is that although externally valid their internal validity may be undermined by previously unrecognised confounding factors that may not be evenly distributed between intervention groups.[12] Thus patients receiving differing treatments may differ systematically with respect to any number of known and unknown factors that affect prognosis.These include the severity of the main and accompanying disease(s), clinical setting, and clinician. Although statistical adjustments may be made in an attempt to exclude the effects of these confounders (and thus isolate any differences attributable solely to the treatment) this assumes both a complete knowledge of confounding variables and their comprehensive and accurate measurement. Neither is likely to be possible and a least a modicum of bias will remain. As most common treatments that interest us will probably have only a moderate sized effect the ability to exclude moderate effects of confounding is vital. Thus the RCT is the preferred design providing it is ethical, practical, and appropriate.

Qualitative research methods[13][14] aim to develop concepts that help us understand social phenomena in natural settings giving due emphasis to the meanings, experiences, and views of all participants. It is concerned not with the "how often" question but with why something is happening, how does it work, and what do people think, believe, or do in order to understand the meaning and interepration of human social arrangements such as hospitals, forms of management, and decision making. It uses methods of observation, interviews, focus groups, and consensus methods. In contrast with quantitative methods the approach is mainly inductive with the hypothesis developed during the research rather than a priori with analysis for the most part being narrative rather than in a numeric form. It has been criticised for being at risk from researcher bias and lacking reproducability and validity. However, this method should be approached with the same rigor as quantitative methods. Usually the samples are small, the objective being not to establish a random sample from the population but rather to identify specific groups of people who either possess characteristics or live in circumstances relevant to the social phenomena studied. It is more useful to see qualitative research as complementary to quantitative methods rather than the antithesis of them.
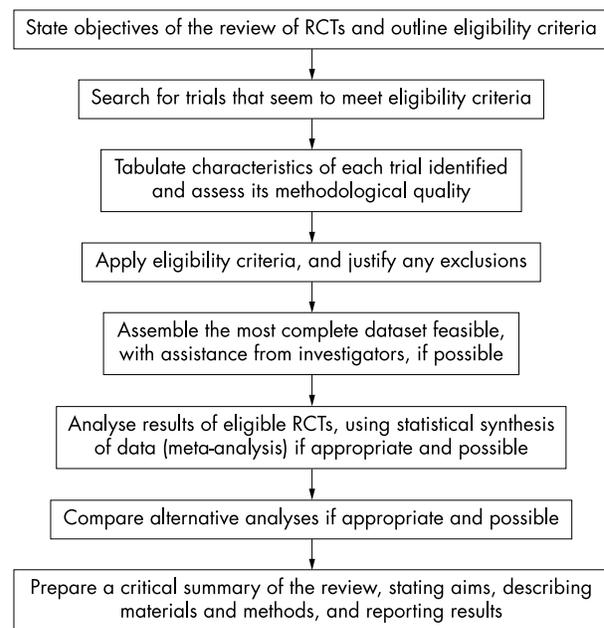


```
┌─────────────────────────────────────────────────────────────┐
│ State objectives of the review of RCTs and outline eligibility criteria │
└─────────────────────────────────────────────────────────────┘
                            ↓
┌─────────────────────────────────────────────────────────────┐
│ Search for trials that seem to meet eligibility criteria        │
└─────────────────────────────────────────────────────────────┘
                            ↓
┌─────────────────────────────────────────────────────────────┐
│ Tabulate characteristics of each trial identified               │
│ and assess its methodological quality                           │
└─────────────────────────────────────────────────────────────┘
                            ↓
┌─────────────────────────────────────────────────────────────┐
│ Apply eligibility criteria, and justify any exclusions          │
└─────────────────────────────────────────────────────────────┘
                            ↓
┌─────────────────────────────────────────────────────────────┐
│ Assemble the most complete dataset feasible,                    │
│ with assistance from investigators, if possible                 │
└─────────────────────────────────────────────────────────────┘
                            ↓
┌─────────────────────────────────────────────────────────────┐
│ Analyse results of eligible RCTs, using statistical synthesis   │
│ of data (meta-analysis) if appropriate and possible             │
└─────────────────────────────────────────────────────────────┘
                            ↓
┌─────────────────────────────────────────────────────────────┐
│ Compare alternative analyses if appropriate and possible        │
└─────────────────────────────────────────────────────────────┘
                            ↓
┌─────────────────────────────────────────────────────────────┐
│ Prepare a critical summary of the review, stating aims, describing │
│ materials and methods, and reporting results                    │
└─────────────────────────────────────────────────────────────┘
```

Figure 3   Reprinted with permission from *BMJ* 1997;**315**:672.

Specifically they may be useful:

1 In the preliminary stages of research into a new area before quantitative research can begin, providing a description and understanding of behaviour.

2 To supplement quantitative methods. They can improve the accuracy and relevance of quantitative studies by increasing our knowledge of the generation of data and the identification of the appropriate variables to be measured.

3 Provide explanations of the unexpected or unexplained findings in quantitative data. Specifically to explore complex phenomena or areas not amenable to quantitative research. This is usually in complex situations where the relevant variables associated with an outcome are not apparent. It aims to increase our understanding of what is going on.

STUDIES OF STUDIES
So far the unit of research has been the patient or a population. However, trials themselves may be the unit of analysis.The systematic review is a scientific tool that can be used to summarise, appraise, and communicate the results and implications of otherwise unmanageable numbers of trials. Crucially, the systematic review will contain an explicit statement of the objectives, materials, and methods and will have been conducted according to an explicit and reproducible methodology. It is especially valuable in bringing together separately conducted studies and synthesising their results. The steps in a systematic review are shown in figure 3.

It is useful to define a review, overview, and meta-analysis.[16]

Review
The general term for all attempts to synthesise the results and conclusions of two or more publications on a given topic.

Overview/systematic literature review
This is when a review strives to comprehensively identify and track down all the literature on a given topic.

Meta-analysis
When an overview incorporates a specific statistical strategy for assembling the results of several studies into a single summary estimate.

Box 1 (reproduced with permission for *BMJ* 1997;315:673)

Advantages of systematic reviews
- Explicit methods limit bias in identifying and rejecting studies
- Conclusions are more reliable and acurate because of methods used
- Large amounts of information can be assimilated quickly by healthcare providers, researchers, and policymakers
- Delay between research discoveries and implementation of effective diagnostic and therapeutic strategies may be reduced
- Results of different studies can be formally compared to establish generalisability of findings and consistency (lack of heterogeneity) of results
- Reasons for heterogeneity (inconsistency in results across studies) can be identified and new hypotheses generated about particular subgroups
- Quantitative systematic reviews (meta-analyses) increase the precision of the overall result

Why are systematic reviews important? A traditional review maybe no more than a subjective assessment by an expert using a select group of materials to support their conclusion. In contrast, a systematic review attempts to be systematic in both identification and evaluation of material, objective in its intereptation, and reproducible in its conclusions. The advantages of systematic reviews are summarised in box 1.

Are meta-analyses the gold standard of evidence? Their pre-eminence based on such trials as the use of thrombolytics in myocardial infarction and the suppression of post-infarction arrhythmias by lignocaine (lidocaine) has been challenged by the failure of large RCTs to confirm the findings of earlier meta-analyses. The most infamous example was the meta-analysis that showed that giving intravenous magnesium to people with myocardial infarction was beneficial. A subsequent megatrial involving 58 000 patients (ISIS-4) failed to demonstrate any benefit and the misleading nature of the meta-analysis was attributed to publication bias and the weaker smaller trials.[17]

The aim of this introductory paper has been to give the reader an insight into the fundamental problem of research (dealing with bias, chance, and confounding) and the strengths and weaknesses of the different types of design available. The general reading list at the end will be of interest to those looking to improve their research designs.

General reading

Hulley SB, Cummings SR. *Designing clinical research.* Baltimore: Wilkins and Wilkins, 1988.

Altman DG. *Practical statistics for medical research.* London: Chapman and Hall, 1991.

Sackett DL, Haynes RB, Guyatt GH, *et al. Clinical epidemiology. A basic science for clinical medicine.* 2nd edn. Boston: Little Brown, 1991.

Lowe D. *Planning for medical research. A practical guide to research methods.* Cardiff: Astraglobe, 1993.

Crombie IK, Davies HTO. *Research in health care.* Chichester: Wiley, 1996.

Mays N, Pope C. *Qualitative research in healthcare.* London: BMJ Publishing Group, 1996.

REFERENCES

1 **Sitthi-amorn C**, Poashyachinda V. Bias. *Lancet* 1993;**342**:286–8.
2 **Sackett DL**. Bias in analytic research. *J Chron Dis* 1979;**32**:51–63.
3 **Lecky F**, Driscoll P. The clinical relevance of observational research. *J Accid Emerg Med* 1998;**15**:142–6.
4 **Datta M**. You cannot exclude the explanation you have not considered. *Lancet* 1993;**342**:345–7.
5 **Grisso JA**. Making comparisons. *Lancet* 1993;**342**:157–9.
6 **Brennan P**,Croft P. Interepting the results of observational research:chance is not such a fine thing. *BMJ* 1994;**309**:727–30.
7 **Altman DG**. *Statistics applied to medical research.*London: Chapman Hall, 1991.
8 **Kendall JM**. Designing a research project: randomised controlled trials and their principles. *Emerg Med J* (in press).
9 **Nicholl J**, Turner J.The effectiveness of a regional trauma system in reducing mortality from major trauma. *BMJ* 1997;**315**:1349–54.
10 **Black N**. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;**312**:1215–18.
11 **Mann C**. Cohort, cross sectional and case control studies. *Emerg Med J* (in press).
12 **Sheldon T**. Please bypass the PORT. *BMJ* 1994;**309**:142.
13 **Mays N**, Pope C. *Qualitative research in health care.* London: BMJ Publishing, 1996.
14 **Black N**. Why we need qualtitative research. *J Epidemiol Community Health* 1994;**48**:425–6.
15 **Greenhalgh T**. Papers that summarise other papers. *BMJ* 1997;**315**:672–5.
16 **ScHARR**. *Introduction to systematic reviews.* Sheffield: Critical Reviews Advisory Group, 1996.
17 **Egger M**, Davey Smith G.Misleading meta-analysis:lessons from "an effective, safe, simple" intervention that wasn't. *BMJ* 1995;**310**:752–4.