

Randomised controlled trials – almost the best available evidence for practice

Vivien Coates

INTRODUCTION

The first paper in this series discussed the growing need for evidence based practice and used a hierarchy of evidence developed by Muir Grey (1997) to illustrate that evidence tends to be associated with quantitative research designs which are most robust and reproducible. According to this hierarchy, systematic reviews based upon randomised controlled trials are reported to offer the strongest evidence, and the next best level of evidence is generated by properly designed randomised controlled trials.

In this paper randomised controlled trials (RCTs) will be discussed. The strengths and weaknesses of this design will be considered, some of the terms and phrases associated with RCTs will be explained and their applicability to nursing will be commented upon. This research design will be illustrated using a study reported by Davies et al (2001). This study was conducted to test whether the provision of a DSN could be proven to deliver a more effective service to patients with diabetes in hospital than usual care, which did not include input from a DSN. Finally, the applicability of the evidence to be derived from this study will be appraised using a schedule developed by Muir Grey (1997).

Strength of evidence

The hierarchy of evidence developed by Muir Grey (1997) has been widely used as a scheme for assessing the strength of evidence for informing practice and was described in a previous paper in this series (Coates, 2004).

After systematic reviews the next best level of evidence is that generated by 'at least one properly designed randomised controlled trial of appropriate size'. The properties which are required for a properly designed RCT and what is meant by appropriate size will be discussed below.

Randomised controlled trials

Quantitative research is concerned with 'precise measurement, replicability,

prediction and control (Powers and Knapp, 1990). It includes random sampling, random assignment of patients to groups and testing for significant differences. It often involves testing hypotheses, it is very structured and uses numbers to present results. Research experiments are one type of quantitative research and are characterised by the researcher, systematically and rigorously studying cause-and-effect relationships between variables and ensuring that the results obtained (the effect) can only be attributed to the intervention (the cause; Parahoo, 1997). Experiments tend to be associated with the biological sciences, conducted in laboratory settings, in which the particular phenomenon of interest (the cause) can be isolated, controlled and measured to test whether it has had an effect. However, RCTs are being used more often in nursing research, partly because of the drive for more robust evidence upon which to base practice.

Research experiments can be designed in different ways, however, RCTs, which are sometimes referred to as clinical trials, are considered one of the best ways of determining whether a cause has had an effect or not (Getliffe, 1998). Parahoo (1997) describes a RCT as 'an experiment in which subjects are randomly allocated to one or more control groups and to one or more experimental groups, depending on the number of interventions'.

RCTs are defined by three key features: an intervention, a control and

LEARNING POINTS

- Randomised controlled trials (RCTs) are a type of experimental research which are being more widely used in nursing.
- RCTs can generate powerful evidence which can be used by policy makers, service managers, clinicians and patients.
- The key features of are RCT are an intervention, a control and randomisation.
- The RCT of Davies et al (2001) has generated evidence to illustrate that a DSN service can make a tangible, quantifiable difference to inpatients with diabetes.

Vivien Coates is a Senior Lecturer in Nursing at the University of Ulster.

LEARNING POINTS

● Trials always involve testing a clinical treatment; this is usually referred to as the intervention.

● A control group is vital in experiments as the new treatment has to be compared with something.

● People in the sample (subjects) are randomly allocated to either the intervention group or the control group.

● In clinical trials all the steps to be taken during the study are developed in advance and are written down in a protocol that must be followed.

randomisation. These properties will be considered below.

Intervention

Trials always involve testing a clinical treatment; this is usually referred to as the intervention. However, in some cases the term manipulation may be used if the researcher does something to some of the people in the study. The new treatment or aspect of care is only offered to patients in the experimental group. The experiment can be designed in several ways. It may be that half the patients are given a new medication and half are not. At the end of the study the effect of the drug is measured on all patients; this is a post-test only design. If the drug is effective there should be a difference between the results of those in the intervention and the control groups. In contrast a pre-post test or before-after design could be used. In which case, the blood levels of all patients are measured at the start of the study, then those in the intervention group receive the new medication while those in the control group do not. The blood levels of all patients are measured at the end of the study and are compared.

In its simplest form there will be only one intervention group. In some studies several experimental interventions will be offered, each one in a different ‘arm’ of the study. For example, if you wanted to know the best way to teach patients how to monitor their blood glucose you could have one intervention group (arm) in which you gave individual teaching, in another arm you could have group teaching, whilst in another arm you give patients a video to watch at home. All other care is the same. The control group receives usual care. The ability of patients to monitor blood glucose would then be assessed in all groups to see which group appeared to learn most effectively.

Control

A control group is vital in experiments as the new treatment has to be compared with something. This is usually standard treatment or care but it could also be against other treatments. By using a control group it is possible to check if the results of the study could have happened by chance.

In the control group, patients will be treated exactly the same way as the patients in the intervention group except that they will not receive the intervention.

Randomisation

People in the sample (subjects) are randomly allocated to either the intervention group or the control group. In experiments, patients involved in a research study are termed subjects and each subject must have an equal chance of being in the intervention or the control group. The purpose of randomisation is to ensure that there is no bias in the allocation of patients to the control or intervention groups. For example, if testing an intervention to increase patients’ knowledge the researchers must resist the temptation to put those they perceive to be the ‘brightest’ patients into the experimental group. As the researcher is often the person who has developed the intervention and would like it to be effective, the temptation to influence the randomisation process may be great. Tables of random numbers are often used to ensure patients are randomly allocated, and further details on using the tables are available elsewhere (Polit et al, 2001).

Once allocated to a group, patients are asked to remain in it until the end of the study. For example, in an experiment on the effects of blood glucose monitoring, those in the intervention group who are to monitor their blood glucose cannot elect to change over to the control group mid study because they have grown tired of monitoring. (Similarly, those who were not monitoring cannot choose to change to the monitoring group mid study). However, ethically we cannot force patients to belong to one group or another so if they change their mind or disagree with the way they have been allocated to a group their preferences must be respected, but it is at the expense of withdrawing them from the trial.

Protocols

In clinical trials all the steps to be taken during the study are developed in advance and are written down in a protocol that must be followed. The protocol specifies all aspects of the study. For example, the

target population, which is the group of people from whom the research subjects will be drawn, must be stated. It may be large, for example, all those with type 2 diabetes in the UK or it may be smaller, such as patients registered with a particular practice. However, after sampling it should be possible to generalise the results from the sample to the entire larger population from which they were drawn. This is known as the generalisability of the results.

The protocol will specify how patients are to be recruited to the study and which inclusion and exclusion characteristics are relevant. When recruiting patients the researcher should not be in doubt as to whether someone is eligible to take part or not. Ethical issues and gaining informed consent must be included. The randomisation process should be explained and all patients should be randomised using the same process. The delivery of the intervention should be clear, exactly what is to be done, for how long and by whom. Similarly what constitutes normal treatment or a placebo for the control group must be made clear.

The importance of the protocol is evident if you imagine a large multicentre RCT such as the UKPDS (Turner et al, 1998). It was essential that the patients in the intervention group receiving intensive treatment were all treated according to one protocol. If intensive treatment was interpreted to mean different things in different centres then the results could not be compared. Similarly, for the control group, usual care had to be comparable from one centre to the next. All the research techniques and instruments must be identified in advance. All the variables of interest in the study, whether they are physiological, such as blood results or psychosocial ones, such as patient satisfaction, must be defined in the protocol. The means of measuring them is specified prior to starting the study. The purpose of all the specifications and the protocol is to eliminate bias as far as possible.

Strengths and limitations of experimental studies

As the hierarchy of evidence indicates, RCTs offer one of the best means of producing evidence for clinical practice.

Indeed, according to Sibbald and Roland (1998) RCTs are 'the most rigorous way of determining whether a cause effect relationship exists between treatment and outcome and for assessing the cost-effectiveness of a treatment'.

However, there are limitations to RCTs which are not always acknowledged. Firstly, nursing care often focuses upon psychosocial and behavioural issues which do not lend themselves to experimentation. These issues can be difficult to isolate, control and manipulate and therefore they may be better investigated using other methods, although these other methods may not be very robust. For some aspects of care experiments may be too impractical to be feasible.

In isolating and controlling aspects of healthcare experimental research can be criticised for being reductionist. To conduct the experiment the researcher must focus upon only a few defined variables, when in reality there are too many variables to control. In many ways this research approach is more suited to medicine which is based on a biomedical model of care than to nursing which strives for a more holistic approach to care.

Experiments have also been criticised for being 'artificial'. Care of patients taking part in an experiment is often not the same as in real life. If we think back to the publication of the results of the Diabetes Control and Complications Trial (DCCT group, 1993) we welcomed the outcomes of the patients in the intensive treatment group, but were concerned about the chances of replicating the standard of control in regular care.

The Hawthorne effect is also known to play havoc with experimental results. This occurs when behaviour is changed simply because people are aware that they are being studied, even if they do not know precisely what is being investigated. To overcome this effect double-blind research designs can be used, in which neither the patient nor the staff are aware of who is in the intervention or the control group. This can be achieved in drug trials in which one group gets the new drug while those in the other group take a placebo tablet. However, this is not possible for many types of nursing

LEARNING POINTS

- The protocol will specify how patients are to be recruited to the study and which inclusion and exclusion characteristics are relevant.
- All the research techniques and instruments must be identified in advance. All the variables of interest in the study, whether they are physiological, such as blood results or psychosocial ones, such as patient satisfaction, must be defined in the protocol.
- There are limitations to RCTs which are not always acknowledged. Firstly, nursing care often focuses upon psychosocial and behavioural issues which do not lend themselves to experimentation.

LEARNING POINTS

● Despite great developments in the role of DSNs there has been a dearth of evidence regarding the impact that they make on patient care.

● This is partly because they work in multidisciplinary healthcare teams, so it is difficult to isolate the contribution of DSNs. However, in terms of providing evidence, experimental studies to prove DSNs do make a difference are urgently required.

care in which the intervention cannot be disguised or approximated through a placebo. A fuller critique of the limitations of RCTs can be seen in Hicks (1998) and Watson et al (2004).

Evaluating a DSN service: an example of an RCT

Despite great developments in the role of DSNs there has been a dearth of evidence regarding the impact that they make on patient care (Loveman et al, 2004). This is partly because they work in multidisciplinary healthcare teams, so it is difficult to isolate the contribution of DSNs. However, in terms of providing evidence, experimental studies to prove DSNs do make a difference are urgently required.

Davies et al (2001) set out to 'evaluate the effectiveness and cost implications of a hospital diabetes specialist nursing service' using a prospective RCT. A prospective study is one in which a current phenomenon is studied over time. In contrast to a retrospective study in which a current phenomenon is studied by seeking information from the past, such as patient records or nurses diaries (Parahoo, 1997).

Population

The target population was all patients with diabetes who were referred to a DSN service within a single hospital. Patients were invited to participate in the study and with their informed permission were randomised to either the control or the intervention groups. Those in the control group received usual care from all relevant healthcare professionals except for the DSN. Those allocated to the intervention group received usual care plus input from the DSN service. The extent of the DSN care included individual structured patient education appropriate to the needs of each patient, practical advice about the management of their diabetes and feedback to the rest of the clinical team. To reduce the chance of any differences between the groups being due to the personality of the DSN rather than the role per se, four DSNs were rotated into the service.

Outcome measures

The primary outcome measures were length of stay in hospital, frequency of re-admission

(within 12 months) and time in days to first admission. Secondary outcome measures were the use of community resources post discharge, patient knowledge and quality of life. Patient satisfaction was also measured a week after discharge. In this study the input of the DSNs is the cause and reduced hospital stay, frequency of readmission and time to readmission are the primary effects. Patients were asked to complete these instruments at the start of the study and one week post discharge, therefore this is a pre-test post-test design.

In this study the researchers used data gathering instruments, such as questionnaires and scales which had been previously developed and which had reported validity and reliability. Quality of life was measured using the Audit of Diabetes-Dependent Quality of Life (ADDQoL) an instrument previously developed by Todd et al (1993). Knowledge was measured by a questionnaire developed by Dunn et al (1984) and patient satisfaction was assessed using the Diabetes Clinic Satisfaction Questionnaire developed by Wilson & Home (1993). It is important to specify the instruments used to gather data, and by using previously developed ones with reported validity and reliability the credibility of the data is increased. It would undermine the quality of the study if the authors had used their own 'home made' instruments or failed to give references in which details about the development of the instruments could be found. Validity and reliability are often overlooked in nursing research but are vital if the results are to be trusted. In this study abstract concepts, such as quality of life, knowledge and satisfaction were to be measured. As they are intangible it is particularly difficult to be sure that they were actually measuring these variables rather than something else. They can only be measured indirectly using a questionnaire or a scale. If the scale is not measuring what it is supposed to measure then the results are not accurate. Similarly, the variables must be reliably measured. This means that the instrument needs to give consistent results each time it is used. The tangible outcomes, such as length of hospital stay can be accurately recorded as long as the

hospital records are accurate. The reliability and validity of biochemical and physiological measures are often not reported as it is assumed they are accurate. However, this may not always be so, for example, blood pressure results will only be accurate if the correct measurement techniques are used.

In addition, patients were followed up after discharge to gather data regarding attendance at outpatients, contacts with primary and social care, and time away from normal activities a month post discharge using a postal questionnaire. However, this data was used as secondary outcomes, it was not part of the principle testing of the effect of the DSN upon care.

Sample size

Sample size is crucially important in experimental design. Prior to starting the study these authors estimated the number of people who would be required to be involved in the study to ensure that if a difference did occur between the two groups that it could be detected. It is important that the number of patients required is known from the outset, it is not acceptable to get as many patients as possible during a certain time span and then check if there are enough. If too few patients are involved, then important differences may not be detected. If too many people are used it is a waste of resources. It is best if just enough patients are involved to enable any differences to be determined.

Power is the probability that a statistical test will detect a significant difference that exists (Burns and Grove, 2003). A type II error (false negative) is the failure to detect a clinically important difference which does exist. Type II errors are often caused by using too small a sample to enable changes to be detected with statistical certainty or where measurement instruments are not sensitive enough to detect small changes (Getliffe, 1998).

A power analysis was conducted prior to the study and indicated that a sample size of 140 patients in each group would be needed and there would be an 80% chance that any differences between the groups would be detected. Recruitment needed to continue until this number was achieved.

Results

All the tests used to analyse the data are

included in the report so that readers know how the results were derived. The authors found that although 508 people were eligible to participate in the study only 300 agreed to take part. However, the basic requirement of 140 people in each group was achieved.

The primary outcomes indicated that the people in the intervention group had a median (the exact middle score across the range of scores) length of stay of 8.0 days whereas those in the control group were admitted for 11.0 days. This result is reported as ($p < 0.01$) which indicates the results were statistically significant. Statistical significance is used to determine whether results could have happened by chance. The letter 'p' indicates the probability of a chance occurrence. Significance at the $p < 0.01$ level indicates the probability that a difference in length of stay of 8 days rather than 11 days would be found by chance is less than one in a hundred. Thus it was extremely unlikely that this was a chance finding and we can conclude that the input of the DSN did cause the length of stay to be shorter for those in the intervention group. There was no evidence of a difference in readmission frequency or to time of readmission. From the secondary outcomes, the intervention group was more satisfied with their care than the control group ($p < 0.001$). This significance level indicates that there would be only one chance in a thousand of this result happening by chance. The knowledge levels between the two groups were no different at the start of the study but were found to be significantly greater for the intervention group ($p < 0.001$) after the study. Quality of life scores were not significantly different between the groups at either the start or the end of the study.

Evidence-based practice

In the first paper in this series (Coates, 2004) the importance of appraising the suitability of research which might be used as evidence to inform practice was discussed. The questions developed by Muir Grey (1997) were included as a way in which evidence can be appraised. These questions will be considered below.

Is this the best type of research method for this question?

As the researchers wanted evidence

LEARNING POINTS

● Sample size is crucially important in experimental design. It is best if just enough patients are involved to enable any differences to be determined.

● All the tests used to analyse the data are included in the report so that readers know how the results were derived.

● Statistical significance is used to determine whether results could have happened by chance. The letter 'p' indicates the probability of a chance occurrence.

regarding the effect that the inclusion of a DSN in the inpatient healthcare team has upon aspects of diabetes care, then yes, this is an appropriate design.

Is the research of adequate quality?

Yes, this study is of good quality. The RCT has been carefully designed and a reasonably detailed account of it is presented. The three key features of an RCT are evident, the variables relevant to the study have been identified and instruments with satisfactory validity and reliability were used to measure them. The required sample size was calculated and sufficient patients were recruited. The data analysis tests were specified and the results are clearly presented. These points all add to the quality of this research.

What is the size of the beneficial effect and of the adverse effect?

The results of this study illustrate that there are statistically significant benefits of being treated by a DSN (thankfully!). Furthermore, there were no adverse effects, e.g. early discharge did not result in faster readmission. It is important that this study included enough patients to enable significance testing to occur as non-significant differences do not provide evidence for practice. The actual size (if any) of the beneficial and adverse effects of the intervention can be calculated using specific statistical analyses, but they were not calculated as part of this study.

Is the research generalisable to the whole population from which the research sample was drawn?

The results of this study could be generalised to the whole diabetes population on the medical and surgical wards in the hospital in which the study was conducted. The way in which the sample was selected, randomised and the fact that where possible the researchers checked to see if there were important differences between those who agreed to take part in the study and those who did not, contribute to the generalisability of these results.

Are the results applicable to the 'local' population?

As this study is only one experiment, in one

locality, using one clinical team it is probably not possible to say that the results are applicable to your own local population. This is why RCTs are taken as almost the best evidence, whilst systematic reviews draw together the results of a range of RCTs.

Are the results applicable to your patients?

The results might be applicable to other patients. However, as the results cannot be widely generalised they cannot be said to be applicable to patients beyond the study site. Nurses need to be aware of the power of evidence produced by a RCT but still be mindful that the results may not identify the most suitable intervention for an individual patient in their care.

Conclusion

The key features of an experiment have been explained and their application to practice illustrated using the study by Davies et al (2001). The focus of this paper was to consider RCTs as a means of generating evidence for practice and some of the strengths and weaknesses of this design were discussed. While terms frequently used in experimental research have been explained it is not possible to include all issues relating to RCTs. In particular the ethical and research governance issues relating to RCTs have not been explored. Similarly, not all aspects of the research by Davies et al (2001) were commented upon, rather it was used as a worked example of an RCT. For those seeking evidence to support the impact that DSNs can make to diabetes care the full paper should be consulted. ■

- Burns N, Grove SK (2003) *Understanding nursing research*. Philadelphia: Saunders
- Coates V (2004) Evidence, research and diabetes nursing. *Journal of Diabetes Nursing* 8(1): 8–11
- Davies M, Dixon S, Currie CJ, Davis RE, Peters JR (2001) Evaluation of a hospital diabetes specialist nursing service: a randomised controlled trial. *Diabetic Medicine* 18: 301–07
- Diabetes Control and Complications Trial Research Group (DCCT) (1993) The effects of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *New Engl J Med* 329(14): 977–86
- Dunn SM, Bryson JM, Hoskins PL, Alford JB, Handelsman DJ, Turtle JR (1984) Development of the Diabetes Knowledge (DKN) Scales: forms DKNA, DKNB & DKNC. *Diabetes Care* 7: 36–41
- Getliffe K (1998) Developing a protocol for a randomised

controlled trial: factors to consider. *Nurse Researcher* 6(1): 5–17

Hicks C (1998) The randomised controlled trial: a critique. *Nurse Researcher* 6(1): 19–31

Loveman E, Royle P, Waugh N (2004) *Specialist nurses in diabetes mellitus* (Cochrane Review). In: The Cochrane Library, Issue 1, 2004. Chichester, UK. John Wiley & Sons, Ltd

Muir Grey JA (1997) *Evidence-based health care .how to make health policy and mangement decisions*. Edinburgh: Churchill Livingstone.

Parahoo K (1997) *Nursing research. Principles, process and issues*. London: Macmillan Press Ltd

Polit DF, Beck, CT, Hungler, BP (2001) *Essentials of nursing research: methods, appraisal and utilization*. Philadelphia; Lippincott

Powers BA, Knapp TR (1990) *A dictionary of nursing theory and research*. London: Sage Publications.

Sibbald B, Roland M (1998) Why are randomised controlled trials important? *British Medical Journal* 316: 201

Todd C, Bradley C, Symonds E (1993) Psycho-social measurement in the audit of diabetes services. *Audit Trends* 1. 141–43

Turner RC, Mills H, Neil HAVV, Stratton IM, Manley SE, Mathews DR et al for the United Kingdom Prospective Diabetes Study Group (1998) Risk factors for coronary artery disease in non-insulin dependent diabetes mellitus: United Kingdom Prospective Diabetes Study (UKPDS:23) *British Medical Journal* 316: 823–28

Watson B, Proctor S, Cochrane W. (2004) Using randomised controlled trials (RCTs) to test service interventions: issues of standardisation, selection and generalisability. *Nurse Researcher* 11(3): 28–42

Wilson AE, Home PD for the Royal college of Physicians and the British Diabetic Association (1993) A dataset to allow exchange of information for monitoring continuing diabetes care. *Diabetic Medicine* 10; 378–90