

Epidemiological research methods

Part III. Randomised controlled trials (for interventions)

J. L. BOTHA, D. YACH

Once the magnitude and distribution of a health problem^{1,2} and its possible determinants have been established, attempts to prevent, treat, or control the problem by intervening on one or more of the determinants should be made. Success should be evaluated in intervention studies by investigating single steps in the natural progression of disease (Fig. 1):³ these present opportunities for primary, secondary or tertiary prevention. In intervention studies the progression from an 'initial' state to an 'end' state is investigated: a sample of individuals in either the initial or end state (in more than one group) is selected, and the groups are compared to evaluate interventions.

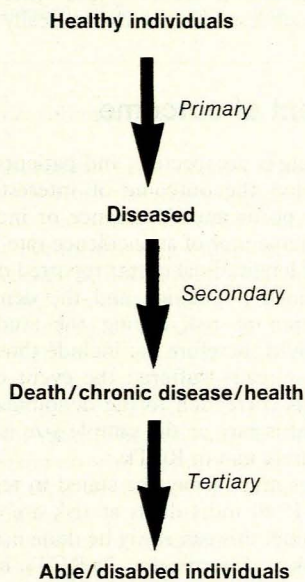


Fig. 1. Simple model of the natural history of disease showing levels of prevention.

The randomised controlled trial (RCT) is considered the most powerful technique for evaluating the effect of intervention. Outcome measurement compares the absolute or relative difference in outcome between groups that differ due to a defined intervention. The estimate of this difference can be biased by any process in the design, execution or analysis of the RCT which affects the groups differentially.

How the validity of estimated measures of effect can be affected in RCTs is discussed here.

Institute for Biostatistics of the South African Medical Research Council and Departments of Community Health, University of Stellenbosch, Parowvallei, CP, and University of Cape Town

J. L. BOTHA, M.B. CH.B., M.S.C. (CLIN. EPID.)

D. YACH, M.B. CH.B., B.S.C. HONS (EPID.), M.P.H.

What makes an RCT so powerful?

To be able to ascribe to an intervention an observed difference in outcome between two groups, one subject to that defined intervention, the two groups should differ with respect to that intervention only. They should be comparable with respect to all other variables which may be associated with the intervention and which may also determine the outcome (confounding variables or confounders⁴). In an RCT such comparability is highly likely because of random allocation of individuals to the two groups. RCTs are considered true scientific experiments, because the intervention and the allocation of individuals to groups is under the control of the investigator: allocation is determined by the study goals and not by patients' needs or characteristics. For this reason, ethical principles such as full and informed consent and the right to refuse allocation always need to be considered.⁵

RCTs are perhaps best known for assessing drug efficacy (used here to denote how well the intervention performs in those who receive the drug; efficacy can only be determined when 100% compliance can be ensured),⁶ but they can also be used to evaluate any treatment programme.⁷ As an illustration we have adapted data from a trial comparing the efficacy of three interventions: rice water (RW), rice electrolyte solution (RES) and glucose electrolyte solution (GES) in management of infantile diarrhoea.⁸ In a classic drug trial, patients suffering from a specified condition amenable to drug therapy are selected and allocated to two groups, one receiving a new drug and the other a placebo. Our example differs slightly from this classic picture in that three groups are compared and there is no clear control group receiving placebo treatment. As often happens in intervention studies, it was not ethically justified to withhold rehydration from infants with diarrhoea, so the rice-based solutions were compared with the World Health Organisation's recommended treatment, GES (identified as the control or standard treatment).

Sampling in RCTs

The target population must be clearly specified before patients are selected. Sampling from the target population then takes place in two steps: firstly, selection of a representative sample from the target population and, secondly, allocation to comparable groups.

Specification of the target population is done by defining inclusion and/or exclusion criteria. In our example, inclusion criteria were specified as 'infants under the age of 6 months admitted primarily or secondarily with acute gastro-enteritis to the paediatric wards . . . of a university-based municipal hospital in Bombay . . . from March 1983 to May 1984'. Patients requiring intravenous therapy were excluded. Specification is done for one of two reasons: to restrict the trial to patients likely to respond to or to need the intervention, or to distinguish clearly between a trial of efficacy (does the intervention work under ideal conditions?) and a trial of effectiveness (does the intervention work under field conditions?).⁹ If the target population could be restricted to patients likely to respond, the difference in response between groups would be

maximal, and fewer patients would be needed. In a trial of efficacy, compliance would have to be measured, as efficacy relates to the disappearance of diarrhoea in those who drink the rehydration fluid in adequate dosage. Effectiveness relates to the disappearance of diarrhoea in practice under field conditions, whether infants drink the rehydration fluid as prescribed or not and is of importance to clinic administrators. The specification means that, generalisation (or external validity) of the results to all infants with diarrhoea (even within Bombay) is reduced. This reduction is taken to its extreme in single-patient RCTs.^{10,11}

Sample selection is necessary if the specified target population is too large. The sampling strategies for descriptive studies² can be used here as well. The penalties for careless sampling in an RCT are similar to those in a descriptive study, but it should be remembered that sample selection is followed by allocation of sampled individuals to two (or more) groups for comparison. If too few patients are sampled, the groups will also have small sample sizes, so that the estimated measure of effect may lack precision. In addition, negative results may reflect the small sample size rather than any 'real' lack of difference between groups. The expected accrual and drop-out rates also need to be taken into account when determining sample size.

Random allocation techniques can make treatment groups comparable with respect to both known and unknown confounding variables. Simple random allocation of N individuals to two treatment groups, each with $N/2$ patients, means that each set of $N/2$ patients within N , have the same probability of being allocated to either group.¹² With simple random allocation the probability of obtaining two comparable treatment groups is high, but there is still a small, but real, probability of the groups not being comparable.

Comparability in terms of known confounders can be ensured if patients are randomly allocated to treatment groups within one (restriction) or several strata (stratification) of known confounders,¹² e.g. infants with mild or severe dehydration may be allocated to separate treatment groups. If the intake of patients into the treatment groups occurs over time, one may wish to ensure that the numbers in the groups remain approximately equal throughout the intake. This is important if there is a possibility that intake may be terminated earlier than planned (funds may run out, for example). Random allocation may take place in balanced blocks,¹² i.e. in such a way that groups have the same number of patients each ($N/2$) after every N allocations. Balanced allocation may also be important in multicentre trials to ensure at regular intervals that treatment groups have the same number of patients from various centres.

A controlled trial with random allocation to treatment groups is a powerful way of evaluating the effect of interventions, but several biases can still occur. In particular, because patients are allocated to different groups, one usually on a new (untested) drug and another on the current accepted treatment, patients have the right to refuse their allocated treatment. If the proportions of refusers differ between groups and if those who refuse differ systematically from those who consent, the estimate may be biased. Various strategies to handle this have been proposed.¹³ All those who refuse should be followed up to see if they differ systematically from those who participate in the trial and to see what influence their omission from or inclusion in the analysis has. Another strategy is to ensure that patients do not know whether they are receiving the test drug. This is sometimes impossible but should always be considered. In addition to this strategy (single-blind), the person observing the outcome should where possible be blinded to the intervention, resulting in a double-blind trial. 'Blinding' ensures that neither the patient nor the investigator can be biased by knowledge of the treatment received. In our illustration, patients were allocated to the three groups according to the

day of the week of admission. If admission to hospital is a random process, allocation would therefore have been random.

Defining the intervention

The intervention being evaluated should be defined clearly and in enough detail to enable readers to replicate the study on their own patients. Methods to measure compliance with the intervention should be described, particularly if the trial is one of efficacy. In all these trials, two other intervention-related problems need attention: those of contamination and co-intervention, both of which can result in bias. Contamination occurs if patients allocated to receive RW actually get GES, or those allocated to GES actually get RW, wilfully or not. As a result, the GES and RW groups will not differ absolutely with regard to the intervention, and its effect will be obscured. The same applies to contamination between RES and GES. Co-intervention occurs if there is a departure from the protocol-prescribed treatment and some other treatment not used in the trial, e.g. intravenous rehydration, is introduced. If the co-intervention occurs in the group on the 'new' treatment, results may appear better than they really are. On the other hand, if co-intervention occurs in the other group, the new treatment may look less effective than it really is.

Measurement of outcome

In RCTs sampling is prospective, and patients are followed up until they develop the outcome of interest. Outcomes are counted as they occur and occurrence or incidence rates are estimated. The numerator of an incidence rate has been defined as the number of longitudinal events reported during a specified period in a defined population and the denominator as the average population at risk during the study period.² The denominator should therefore not include those not at risk (i.e. those who have already suffered the event or are no longer susceptible). This correction to the denominator need not be made if the event is rare or the sample size is large, but these conditions are rarely met in RCTs.

Incidence rates must always be stated in terms of a definite period of time. If all individuals at risk are observed for the same length of time, this can easily be done using the midpoint population at risk as denominator. In RCTs, however, subjects are often observed for unequal periods of time because people die, move away or are otherwise lost to follow-up, or they come under observation after the study has been initiated. To make full use of the period of observation in each individual, a person-time unit (e.g. a person-year) is created for the denominator by totalling each person's period of time under observation.

Comparative measures of effect are of interest in RCTs. Comparison of rates can be indicated as a rate ratio, indicating how many times more or less frequently the disease occurs in one group than in another, or as a rate difference, indicating the excess rate of occurrence in the group at higher risk or the decreased rate of occurrence in the group at lower risk. In Table I these measures are shown as estimates of the incidence of diarrhoea (5 or more stools per day) on day 7 of the treatment programme. The rate ratios are dimensionless, whereas the rate differences convey the absolute magnitude of the risk. From both it is clear that the rice-based fluids are more efficacious in treating diarrhoea than GES.

Random measurement error occurs when measurements are spread around the true value in each of two groups. If events are being counted, this means that some non-events have been counted as events, and some events as non-events. Such a component of random measurement error will contribute to

TABLE I. COMPARATIVE MEASURES OF EFFECT IN AN INTERVENTION STUDY

	No.	Incidence	Diarrhoea on day 7 of intervention		
			Incidence rate/1 000	Rate* ratio	Rate* difference/1 000
GES	50	13	260	1,00	0
RW	50	1	20	0,08	240
RES	50	4	80	0,31	180

*Calculated with GES as the baseline.

TABLE II. EFFECT OF SYSTEMATIC ERROR IN MEASURING THE OUTCOME

Intervention		All groups affected			Only RES affected		
		Incidence rate/1 000	Rate ratio	Rate difference/1 000	Incidence rate/1 000	Rate ratio	Rate difference/1 000
Diarrhoea incidence in Table I underestimates real incidence by 50%	GES	520	1,00	0	260	1,00	0
	RW	40	0,08	480	20	0,08	240
	RES	160	0,31	360	160	0,62	100
Diarrhoea incidence in Table I underestimates real incidence by 10 cases	GES	460	1,00	0	260	1,00	0
	RW	220	0,48	240	20	0,08	140
	RES	280	0,61	180	280	1,08	20

imprecision of measurement and can be quantified using confidence limits.²

Systematic measurement error occurs when measurements deviate systematically from the true value. If events are being counted, this means that some non-events have been counted as events, or some events as non-events, but not both. If such systematic error occurs in both groups, the comparative estimate may be unbiased, a similar shift in scale occurred in both groups, so that the difference remains valid. This depends on which estimate of effect is used. If such systematic error occurs in one group only, the comparative estimate will always be biased; a shift in scale occurred in one group only, so that the difference is not valid. The precision of the estimate is not affected by systematic error of measurement. These effects are illustrated in Table II, if we assume that some data sheets with stool counts were mislaid.

The criteria for measuring outcome(s) should be stated in sufficient detail to permit their application elsewhere, as identical methods for, and identical application of those methods will be necessary for the success of other trials. 'Blinding' of patients and/or observers has already been mentioned. It is also important to specify beforehand how certain results (particularly unexpected ones) should be attributed or counted.¹⁴ In addition, all morbidity or mortality should be reported, to indicate whether the effect is specific to a particular disease or not.

Analysis

RCTs are examples of prospective studies in that they compare

outcomes in different groups and at different times. In that sense they are identical to other prospective studies and can be analysed using survival analysis (e.g. life-table or person-time techniques).¹⁵ In those, and simpler analyses, the measures of effect are relative or absolute differences in incidence rates, which can be biased by confounding variables. Random allocation as an attempt to control confounding may not always be successful and the analysis should always investigate this possibility.¹⁶

In our example, the severity of dehydration may be considered a confounder, being associated with both the intervention and the outcome. For illustrative purposes mildly and severely dehydrated infants are analysed separately (Table III), as if 28% of infants receiving GES and 36% of infants receiving RES had mild dehydration. When all infants were analysed together, the greater effect of RES in reducing diarrhoea in severely dehydrated infants was obscured.

Applications of intervention studies

Clinical trials, field trials, community trials, before-and-after studies, and case-control studies are all different types of intervention studies. In the first four designs listed, prospective sampling can be used; individuals in the initial state are sampled and followed up to measure the incidence of the end state (outcome measurement). The temporal sequence is clear: intervention precedes outcome. In a case-control study individuals are sampled in the end state and attempts are then made to evaluate the intervention they were exposed to. This design is more commonly used as an analytical study.

TABLE III. CONFOUNDING EFFECT OF DEHYDRATION STATUS ON THE REDUCTION IN DIARRHOEA INCIDENCE RATE

	No.	Mild dehydration			Severe dehydration			
		Incidence	Incidence rate/1 000	Rate ratio	No.	Incidence	Incidence rate/1 000	Rate ratio
GES	14	4	286	1,00	36	9	250	1,00
RES	18	3	167	0,58	32	1	31	0,13

Clinical, field and community trials are considered true scientific experiments, because the intervention (or allocation to groups) is under the control of the investigator as allocation is determined by the study goals and not by patients' needs. Others are subexperimental designs since intervention is not under the control of the investigator and may be determined by the patients' needs. These three trials are all examples of controlled trials, in that they have concurrent control groups. In clinical trials, the subjects are usually patients with disease, and the intervention aims to prevent sequelae of the initial disease.^{8,17-19} These complications may occur with high probability in a relatively short period of time (as in our example). In field trials, the subjects are usually individuals who have not yet become diseased, and the intervention aims at prevention.^{20,21} The risk of disease occurring in these subjects is usually small, therefore more subjects may be needed for a field trial than for a clinical trial, and disease-free people in the community have to be visited. Both these factors make field trials more expensive than clinical trials, usually done at hospitals. A community trial is an extension of a field trial, where intervention occurs in a community (not an individual).²² In these studies with a contemporaneous control group, group comparability is enhanced with respect to confounding variables related to the passage of time during the study.

In a before-and-after study, a group of individuals is selected in the initial state and the outcome measured. Intervention then occurs and the outcome is measured again in the same individuals, or in a different group from the same community. Again, the temporal sequence is clear, but other changes may have taken place with time, decreasing comparability of the groups. RCTs are regarded as the 'gold standard' for intervention and analytical studies. Not all RCTs are necessarily of good quality,²³⁻²⁵ and neither can RCTs be used in all circumstances.²⁶ In the next two articles of the series alternative subexperimental designs will be discussed.

REFERENCES

1. Yach D, Botha JL. Epidemiological research methods: I. Why epidemiology? *S Afr Med J* 1986; **70**: 267-270.
2. Botha JL, Yach D. Epidemiological research methods: II. Descriptive studies. *S Afr Med J* 1986; **70**: 766-767.
3. Feinstein AR. *Clinical Epidemiology: The Architecture of Clinical Research*. Philadelphia: WB Saunders, 1985.
4. Miettinen OS, Cook EF. Confounding: essence and detection. *Am J Epidemiol* 1981; **114**: 593-603.
5. Altman DG. Statistics and ethics in medical research: study design. *Br Med J* 1981; **281**: 1267-1269.
6. Meinert CL, Tonascia S, Higgins K. Content of reports on clinical trials: a critical review. *Controlled Clin Trials* 1984; **5**: 328-347.
7. Spitzer WO, Sackett DL, Sibley JC *et al*. The Burlington randomized trial of the nurse practitioner. *N Engl J Med* 1974; **290**: 251-256.
8. Mehta MN, Subramaniam S. Comparison of rice water, rice electrolyte solution and glucose electrolyte solution in the management of infantile diarrhoea. *Lancet* 1986; **i**: 843-845.
9. Sackett DL. The competing objectives of randomized trials. *N Engl J Med* 1980; **303**: 1059-1060.
10. McLeod RS, Cohen Z, Taylor DW, Cullen JB. Single-patient randomized clinical trial: use in determining optimum treatment for patient with inflammation of Kock continent ileostomy reservoir. *Lancet* 1986; **i**: 726-728.
11. Guyatt G, Sackett D, Taylor DW, Chong J, Roberts R, Pugsley S. Determining optimal therapy — randomized trials in individual patients. *N Engl J Med* 1986; **314**: 889-892.
12. Zelen M. The randomization and stratification of patients to clinical trials. *J Chronic Dis* 1974; **27**: 365-375.
13. Zelen M. A new design for randomization clinical trials. *N Engl J Med* 1979; **300**: 1242-1245.
14. Sackett DL, Gent M. Controversy in counting and attributing events in clinical trials. *N Engl J Med* 1979; **300**: 410-412.
15. Yach D, Botha JL. Epidemiological research methods: V. Follow-up studies. *S Afr Med J* 1987 (in press).
16. Rothman KJ. *Modern Epidemiology*. Boston: Little, Brown, 1986.
17. The Canadian Cooperative Study Group. A randomized trial of aspirin and sulfapyrazone in threatened stroke. *N Engl J Med* 1978; **299**: 53-59.
18. The Canadian Multicenter Transplant Study Group. A randomized clinical trial of cyclosporine in cadaveric renal transplantation: analysis at three years. *N Engl J Med* 1986; **314**: 1219-1225.
19. Macular photocoagulation study group. Argon laser photocoagulation for senile macular degeneration: results of a randomized clinical trial. *Arch Ophthalmol* 1982; **100**: 912-918.
20. Szmuness W. Hepatitis B vaccine: demonstration of efficacy in a controlled clinical trial in a high-risk population in the United States. *N Engl J Med* 1980; **303**: 833-841.
21. Multiple Risk Factor Intervention Trial Research Group. Multiple risk factor intervention trial: risk factor changes and mortality results. *JAMA* 1982; **248**: 1465-1477.
22. McKay AM, Rothman KJ. The incidence and severity of burn injuries following Project Burn Prevention. *Am J Public Health* 1982; **72**: 248-252.
23. Mosteller F, Gilber JP, McPeck B. Reporting standards and research strategies for controlled trials. *Controlled Clin Trials* 1980; **1**: 37-58.
24. Chalmers TC, Smith H, Blackburn B *et al*. A method for assessing the quality of a randomized control trial. *Controlled Clin Trials* 1981; **2**: 31-49.
25. Botha JL. How not to do clinical trials. *S Afr J Hosp Med* 1982; **8**: 122-126.
26. Greenland S, Watson E, Neutra RR. The case-control method in medical care evaluation. *Med Care* 1981; **19**: 872-878.

Nuus en Kommentaar/News and Comment

Liver transplantation

An excellent review of liver transplantation appeared in the *British Journal of Hospital Medicine* (1986; **36**: 410). K. E. F. Hobbs and J. Terblanche (on sabbatical from Cape Town) point out that during the 3 years since an international consensus conference held in 1983 there has been an explosive increase in the number of liver transplant centres throughout the world, with many more about to start. Hobbs and Terblanche estimate that about 30 patients per million population annually will need a new liver, and that any patient with untreatable progressive liver or biliary tract disease should be considered a possible transplant patient, although some can be excluded immediately because of associated problems that make the operation impossible.

They point out that the liver must be removed from a brain-dead donor with an intact circulation free from replicating hepatitis and human immunodeficiency virus and with no evidence of previous liver disease, septicaemia, excessive alcohol intake or a period of prolonged hypotension or hypoglycaemia. The supply of donors in the UK is probably sufficient to support the three

transplant units in Cambridge, Birmingham and London, but as demand increases this situation will not continue.

After discussion of the operation itself and the results, they note that a significant number of patients world-wide have been restored to a healthy productive life after liver transplantation and that children who had a miserable existence with primary biliary atresia have been greatly helped and are currently surviving into their fifth year following transplantation. It remains to be seen how long the grafted livers will continue to function in these children. Adults have a very good chance of continuing to survive if they manage to survive the first year after operation. The longest survival in the world is now 17 years. However, figures for transplantation for malignancy are more depressing.

The authors say: 'Society must decide how much money it is prepared to invest in this field. All processes in which the frontiers of science are being advanced are expensive but the experience gained can be of value in other spheres and this alone may justify the investment.'