# Statistics in Question

SHEILA M GORE

# ASSESSING CLINICAL TRIALS—DESIGN II

At opposite ends of a range are fixed sample size designs and full sequential designs. In the former sample size is determined, or fixed, before the trial begins. In a full sequential design sample size depends in a particular way on the results as they accumulate, and the trial is stopped as soon as a significant difference in treatment is established. Group sequential designs[1] are a compromise between these two approaches. Analyses are planned to coincide with regular trial meetings (at yearly intervals, say). A maximum number of analyses is allowed. At each analysis the decision to stop the trial or to continue is based on correctly applying repeated significance tests to the accumulated data. Doctors should consult a statistician about sequential trials.

Historical comparison alone usually gives insufficient evidence. Non-randomised studies of treatment with anticoagulant drugs during the 1950s and 1960s showed an apparent protective effect (pooled data): the prevention of 53% of deaths after myocardial infarction. The evidence was not sufficient, however, to persuade doctors generally to manage myocardial infarction with anticoagulant drugs. Later, in randomised trials the apparent protective effect of anticoagulant drugs was estimated as preventing 20% of deaths. Note that the bias in the historically controlled studies was of the same order of magnitude as the treatment effect—a grave warning of the errors inherent in historical control series.[2]

In research programmes a treatment is often carried forward from one trial to act as standard or control treatment in the next. Could the patients who have been assigned to this treatment in the earlier study form a historical control group in the later trial? The answer is a qualified yes. Minimum criteria[3] for combining historical and prospective randomised control groups are shown below.

Historical comparison should not take the place of a prospective randomised control group, however.

## Sequential trials

(10) *What ethical argument is there against fixed sample size?*

**—no patient should receive a treatment that has been established as inferior on the basis of substantial accumulated evidence**

**—planning interim analyses is therefore a sensible precaution**

COMMENT

An ethical argument against fixed sample size is that no patient should receive a treatment that has been found inferior


by a substantial accumulation of evidence. A controlled randomised trial[4] of active immunotherapy for stage IIB malignant melanoma was stopped after one year, when only 15 patients had been admitted, because four deaths had occurred in the vaccinated group, three of them after early widespread recurrence of the disease, compared with no deaths in the control group. Unless a formal stopping rule has been defined there may be disagreement about what is sufficient evidence of harm. For this and other reasons there was controversy over the University Group Diabetes Program[5 6] in which treatment by tolbutamide was discontinued.

In the Norwegian multicentre study[7] of mortality after timolol treatment for patients who had had acute myocardial infarction ethical provision was made for confidential and independent interim review of the study information.

It is a sensible precaution to plan interim analyses.

Appropriate adjustment has to be made to the nominal significance level to account for repeated looks at the data. (What do I mean by "nominal significance level"? At each analysis the difference between treatments is tested and if significant at a specified level—the nominal significance level—the trial ends. The nominal significance level is used as a stopping rule, therefore. The rules are worked out in advance of the trial so that overall the sequential plan is associated with an acceptably low risk of a false-positive—that is, of claiming a treatment difference when none exists.) Nominal significance levels are more stringent when there are repeated opportunities for stopping the trial early, so that overall the risk of falsely claiming that one treatment is better than another or better than no treatment remains fixed.

(11) *Why is a fully sequential design often impracticable?*

**—treatment results are known too late to limit patient entry**

**—most fully sequential plans compare only two treatments and assume well-behaved response distributions**

**—difficult to allow for more than one response variable or to adjust for prognostic factors**

**—onus on doctors, data-processing staff, and statistician**

COMMENT

A fully sequential design is often impracticable when there is a long delay between starting a treatment and its outcome; when there is more than one response variable; when more than two treatments are compared; or when patients have very different prognoses. Even when the outcome is known soon after starting treatment, analysing as soon as the result is known for each patient or patient-pair requires (*a*) prompt reporting of treatment results to the trial co-ordinator; (*b*) efficient data processing; (*c*) the availability of the statistician to update the analysis (unless the doctor enters results on a prepared chart, in which case the treatment code has to be broken).

The doctor, the data-processing staff, and the statistician have other responsibilities; they cannot spend all their time on a single clinical trial. Together, these reasons explain why fully sequential designs are uncommon despite their ethical appeal.

(12) *If you repeatedly test accumulating data*[8] *at, say, the 1% nominal significance level, in what way do you alter your chance overall of finding a significant difference?*

**—overall the chance of finding a spurious significant difference is increased**

COMMENT

If accumulating data are tested repeatedly at a nominal 1% significance level the chance of finding a spurious significant difference is *increased.*[8] For example, if you test for a difference between two treatments after every 12th patient and intend to stop the trial as soon as the difference is nominally significant at the 1% level, then by the 10th test the effective (type I) error is about 5%—that is, the chance is about 1 in 20 (not 1 in 100 as might be supposed), that a significant difference will be declared even when there is actually no difference between treatments.

Statistical advice should be sought about planning and analysing group sequential trials.

McPherson K Statistics: the problem of examining accumulating data more than once
N Engl J Med 1974; 290: 501-2

Pocock S J The size of cancer trials and stopping rules
Br J Cancer 1978; 38: 757-66

## Historical comparison

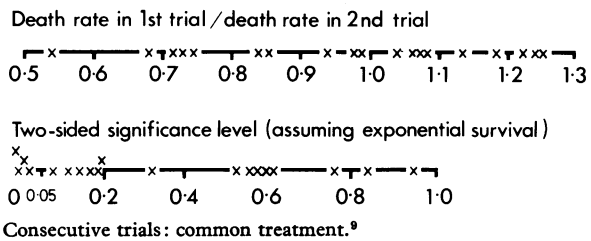
(13) *Why is the answer "a qualified yes" to the question of whether historical and prospective randomised control groups can be combined?*

**—a historical control group—acceptable a priori—is an embarrassment later if comparison of results between historical and prospective randomised control groups approaches significance**

COMMENT

If there is an acceptable historical control group the proportion of patients who are randomised to the control arm in the new trial can be reduced. But a historical control group that is acceptable a priori may be an embarrassment later if a comparison of the results between historical and prospective control groups even approaches significance, let alone exceeds conventional levels. A surprisingly large difference ($p < 0.01$) emerged[2] when a treatment group was carried over as control from the fifth Medical Research Council acute myeloid leukaemia trial to the sixth. In such instances the historical control group must be abandoned, with loss of precision owing to reduced numbers of patients. Pocock[9] reported on 19 (unselected) pairs of consecutive trials in which a combination of historical and prospective randomised groups could have been considered.

Pocock S J The combination of randomised and historical controls in clinical trials
J Chron Dis 1976; 29: 175-88

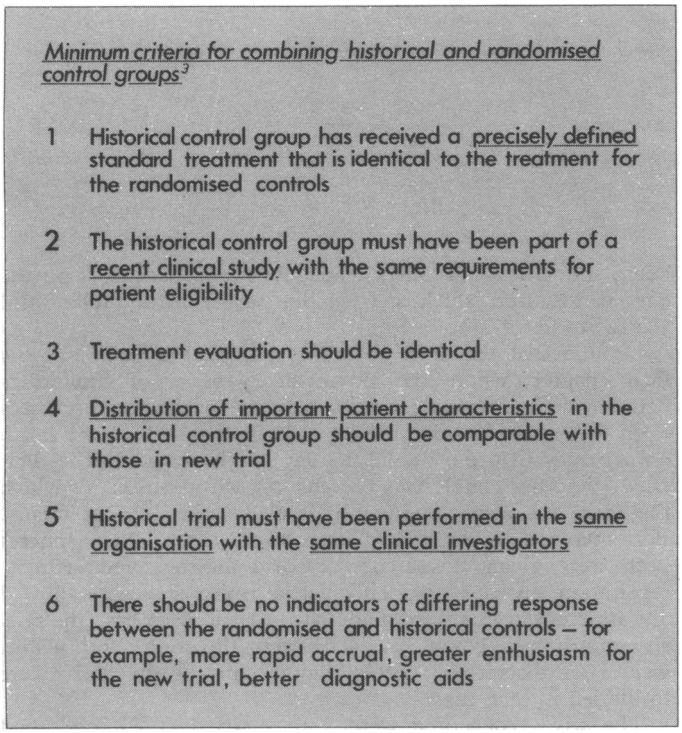


Consecutive trials: common treatment.[9]

Of the 19 significance levels that compared survival on the same treatment in matched trials, four were embarrassingly significant at the 2% level. Failure to establish bias is not proof that there is no bias.

Thus only qualified approval can be given to the combination of historical and prospective randomised control groups. Seek statistical advice.

(14) *Justify criteria 2, 4, and 5 for combining historical and prospective randomised control groups.*




COMMENT

**2** The group must have been part of a *recent clinical study* with the same requirements for patient eligibility. *Recent* because there may be a time trend in responses or disease characteristics. *Clinical trial membership* because there is a tendency for patients to respond better to treatment in a clinical trial than ordinarily.[10]

**4** *The distribution of important patient characteristics* in the historical control group should be comparable with the distribution of those characteristics among patients in the new trial. Otherwise the groups vary obviously in aspects that may determine response. They may be samples from subtly different patient populations.

**5** The historical trial must have been performed in the *same organisation* with the *same clinical investigators*—simply because otherwise there may be response differences attributable to organisation.

I am grateful to the editor of the *Journal of Chronic Diseases* for permission to reproduce under question 14 a modified version of the criteria for combining historical and randomised control groups.

## References

1 Pocock SJ. The size of cancer clinical trials and stopping rules. *Br J Cancer* 1978;**38**:757-66.
2 Peto R. Clinical trial methodology. *Biomedicine Special Issue* 1978;**28**: 24-36.
3 Pocock SJ. The combination of randomized and historical controls in clinical trials. *J Chron Dis* 1976;**29**:175-88.
4 McIllmurray MB, Embleton MJ, Reeves WG, Langman MJS, Dean M. Controlled trial of active immunotherapy in management of stage IIB malignant melanoma. *Br Med J* 1977;i:540-2.
5 Anonymous. Are antidiabetic drugs dangerous? *Br Med J* 1970;iv:444.
6 Cornfield J. The University Group Diabetes Program. A further statistical analysis of the mortality findings. *JAMA* 1971;**217**:1676-86.
7 The Norwegian Multicenter Study Group. Timolol-induced reduction in mortality and reinfarction in patients surviving acute myocardial infarction. *N Engl J Med* 1981;**304**:801-7.
8 McPherson K. Statistics: the problems of examining accumulating data more than once. *N Engl J Med* 1974;**290**:501-2.
9 Pocock SJ. Randomised clinical trials. *Br Med J* 1977;i:1661.
10 Lennox EL, Stiller CA, Morris Jones PH, Kinnier Wilson LM. Nephroblastoma: treatment during 1970-3 and the effect on survival of inclusion in the first MRC trial. *Br Med J* 1979;ii:567-9.

Sheila M Gore, MA, is a statistician in the MRC Biostatistics Unit, Medical Research Council Centre, Hills Road, Cambridge CB2 2QH.

*No reprints will be available from the author.*

---

*A 21-year-old man had a clear cell carcinoma of the sweat glands on the left scapular region. The four-inch (10 cm) tumour was widely excised, the axilla dissected, and the skin defect closed with a graft. The tumour had spread to two lymph nodes in the scapular region, deep to the tumour and to seven more in the axilla. The highest lymph node was clear, but unfortunately there was transcapular spread of tumour in the lower axilla with invasion to local fat and perineural spaces. Do radiotherapy and chemotherapy have a place in treating this patient? What is the future prognosis?*

Clear cell carcinoma of the sweat glands usually occurs in the sixth and seventh decades, although rarely they may occur in young adults and have been reported in adolescence. From the description given one cannot tell whether the tumour has, in fact, been excised en bloc with adequate clearance of all affected tissues. If it has no further treatment is required now. The five-year survival in such conditions, with lymph-node spread, is about 25%. If, however, the surgeon thinks that excision was inadequate or the pathology report shows that the line of excision contains tumour cells then it would be reasonable to prescribe radiotherapy in an attempt to "sterilise" the tumour bed. Chemotherapy has little place in the treatment of cutaneous cancers, and certainly there is no indication for its use at this stage.

*Is it necessary to give oral polio vaccine to parents when their children are given it?*

No, but theoretically it might be argued that it would be safer; in practice it would be difficult, at least in a child health clinic, because nearly always only the mother comes with a child who is to be immunised, and if one insisted that the father came too he would probably have to miss work. Paralytic poliomyelitis in contacts of vaccinated children has been reported in rare instances—presumably as a result of a wild strain of polio virus.[1] As a result, in 1978 the DHSS advised that "when infants are given oral polio vaccine, the vaccine should be offered to the parents if they are unimmunised, in order to protect them from the very small risk of contact-vaccine-associated poliomyelitis."

1 Blattner RJ. Paralytic poliomyelitis: contacts of vaccinated children. *J Pediatr* 1967;**71**:759-62.

*Nitroglycerin can be obtained as a paste for treating angina. What base is used in the preparation of this paste, and has it any advantage over oral isosorbide dinitrate?*

Nitroglycerin ointment is not marketed in Britain, but an American preparation (Nitrol, which contains 2% nitroglycerin in a lanolin-petroleum base) may be obtained from Jaffe Chemists, 42 Charlotte Street, London W1. Both isosorbide dinitrate and nitroglycerin ointment improve exercise tolerance in patients with angina, and their effects appear to be comparable.[1] Both have a duration of effect of up to about five hours. The effective dose range for isosorbide dinitrate is wide, and there is no clearly defined dose range for nitroglycerin ointment, so that the correct dose for a particular patient has to be individually determined for both preparations. An excessive dose of nitroglycerin ointment can be removed with alcohol.

1 Abrams J. Usefulness of long-acting nitrates in cardiovascular disease. *Am J Med* 1978;**64**:183-6.