Statistics in Question

SHEILA M GORE

ASSESSING CLINICAL TRIALS— TRIAL SIZE

The size of a clinical trial is decided by balancing statistical considerations—especially the power of the investigation to identify a difference that is clinically important—against practical considerations, such as limitation on time or the availability of patients. In general, the smaller the difference that is judged clinically significant the greater the number of patients needed for a sensible clinical trial.

(5) *The report on a clinical trial concludes that “the difference between treatments is not statistically significant, $p > 0.05$.” Why is it wrong to interpret this as meaning “there is no clinically important difference between the treatments”?*

—**inadequate sample size can be the reason for not recognising treatment superiority**

—**“not statistically significant” does not condemn**

—**correct interpretation of trial results calls for complete reporting of the statistical methods used to determine trial size, and estimation of confidence intervals**

COMMENT

Significance testing is not good statistics. What is needed is estimation. As well as reporting significance levels ($p < 0.01$ versus “not statistically significant”), investigators should quote an interval that they are confident covers, or includes, the real treatment difference.^{1,2} This is no more difficult to do than significance testing, and is more informative. Regrettably, in the medical league confidence intervals have not passed first base.

Reporting only that the difference between treatments is “not statistically significant” may jeopardise correct interpretation of the clinical trial. Freiman *et al*³ reviewed 71 “negative” binomial trials (a binomial trial is one in which the outcome for an individual patient is categorical, and there are two categories of outcome: treatment success or failure) to determine whether the investigators had studied large enough samples. Forty-five of the

Power function: the confidence with which the investigator can claim that a specified treatment benefit has not been overlooked

trials had less than a 70% chance (POWER) of detecting even a substantial improvement (50%) in cure rate. Estimates of 90% confidence intervals for the true improvement showed that in 34 out of the 71 studies the results were compatible with a 50% difference between cure rates. The authors warned that many treatments labelled as “no different from control” in trials that

Aim for 80% power to detect what is regarded a priori as a realistic, worthwhile treatment difference. Of course, the trial has less power to detect correspondingly smaller differences

How imperfectly does clinical trial outcome measure up to reality?*

Wrong decisions in columns 1 and 2 are not too serious, but it is important not to miss a worthwhile treatment difference (column 3).

	①	②	③
REALITY	No difference between treatments	Small difference (clinically insignificant)	Worthwhile treatment difference
CLINICAL TRIAL OUTCOME	5% of trials conclude treatments differ significantly, $p < 0.05$ Direction of difference is fortuitous	A proportion of trials (more than 5% but less than 80%) declare a statistically significant treatment difference	20% of trials find that the difference between treatments is not statistically significant, $p > 0.05$ Direction of treatment difference is probably reliable nevertheless, and so compute confidence interval

*Illustrated for significance level 5%, and trial size chosen to give 80% power for detecting a worthwhile treatment difference.

use inadequate sample size have not received a fair test. It is wrong to conclude that a non-significant difference rules out the possibility of a worthwhile, or even substantial, clinical difference between treatments.⁴

To interpret a clinical trial correctly⁵ you need to know the power of the trial to do the job it is intended to do, which is identification of treatment gain, and the trial results need to be reported completely. Complete reporting entails the estimation of confidence intervals. In a survey of 38 clinical trials published from July to December 1977 as original articles in a leading general medical journal, there was no mention of how trial size was determined. What applied in one journal probably applies in others. A survey⁶ on the size of randomised controlled trials registered with Union Internationale Contre le Cancer gave more encouraging results. Most investigators used statistical methods to determine trial size. When statistical methods are used it is helpful at least to refer to these in published reports so that readers are aware that statistical considerations weigh heavily in determining trial size. Doctors should be able to absorb good principles in the design of experiments from the very journals they read to keep informed about clinical research and practice.

(6) Fifteen per cent of patients die within one year of admission to hospital for suspected myocardial infarction. Preventing one-third of those deaths would be a major advance. Roughly how many patients are needed for a clinical trial if the doctors want to be 90% sure that a difference between treatments as large as the prevention of one-third of deaths will not be missed (significance level, 5%)?

—about 2000 patients

COMMENT

About 2000 patients are needed to compare a new treatment against a standard treatment when the critical outcome is prevention of death within a year of admission for suspected myocardial infarction. Only then can doctors be at least 90% sure that a difference between treatments as large as the prevention of one-third of deaths will not be missed.

A randomised double-blind trial compared three drugs—propranolol, atenolol, placebo—in the immediate treatment of suspected myocardial infarction.⁷ Only 388 patients were randomised; 259 were given beta-blockers, of whom 36 (14%) died within one year of admission. Of the 129 patients who had received placebo, 19 (15%) died.

The observed difference in death rates at one year was -1% ,

a point estimate for the true difference. But the estimator has a large standard error because relatively few patients entered the trial.⁸ The 95% confidence interval for the true difference in death rates is surprisingly wide, from -8% to $+7\%$, and does not exclude the possibility that one treatment or the other prevents one-third of deaths. The interval does not include -10% or $+10\%$, so that prevention of two-thirds of deaths—a dramatic treatment difference—is unlikely. A sufficient

Immediate treatment of suspected myocardial infarction. Beta-blockers versus placebo: significance level 5%.

number of patients were studied to give a reasonable guarantee against failure to detect only very large treatment differences, such as prevention of two-thirds of deaths. But so large a benefit from beta-blockers, or any other drugs, used in the immediate treatment of suspected myocardial infarction is a priori unlikely.

The published report by Wilcox *et al*⁷ was extremely lucid, except in three aspects. There was no statement of the magnitude of treatment effect which the authors considered worth while, no information on how trial size was determined, and no confidence intervals were given. In these omissions the authors are not alone.

Elwood and Sweetnam,¹⁰ by comparison, reported on a trial of aspirin after recent myocardial infarction which was designed to detect a 25% reduction in secondary mortality (significance

level, 5%). They observed a 17% reduction, which was not statistically significant. Secondary mortality rates were 12.3% and 14.8% in the aspirin and placebo groups respectively; sample size was 1682 patients. For the reduction achieved in the trial to have been statistically significant the number of patients (to give 70% power, see figure above) would have had to have been almost double. This puts the power of the trial in perspective. Correctly, the Norwegian report on timolol-induced reduction in mortality after acute myocardial infarction (trial size, 1884 patients) noted that few studies had entered a sufficient number of patients to exclude the possibility of a beneficial effect from beta-blockade being overlooked.¹¹

In clinical trials in which the critical outcome is time of death (or time to some other event, such as metastasis) the power of the trial to identify the merits of two treatments depends on how many patients die (or experience the event) rather than on the number of patients entered.¹²

Peto R Clinical trial methodology
Biomedicine Special Issue 1978; 28: 24-36

Peto¹³ has complained that a proliferation of small trials has made it difficult for doctors to know how best to manage suspected myocardial infarction. The review by Chalmers *et al*¹⁴ of 32 comparative studies of anticoagulant treatment, on which Peto comments, did much to clarify the management.

Pocock¹⁵ has given guidelines for calculating sample size in binomial trials, together with worked examples.

References

- Altman DG. Statistics and ethics in medical research. VI Presentation of results. *Br Med J* 1980;**281**:1542-4.
- Welff HR. Confidence limits in evaluating controlled therapeutic trials. *Lancet* 1973;iii:969-70.
- Freiman JA, Chalmers TC, Smith H, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomised control trial. Survey of 71 "negative" trials. *N Engl J Med* 1978;**299**:690-4.
- Altman DG. Statistics and ethics in medical research. III How large a sample? *Br Med J* 1980;**281**:1336-8.
- Anonymous. Interpreting clinical trials. *Br Med J* 1978;ii:1318.
- Pocock SJ. Size of cancer clinical trials and stopping rules. *Br J Cancer* 1978;**38**:757-66.
- Wilcox RG, Roland JM, Banks DC, Hampton JR, Mitchell JRA. Randomised trial comparing propranolol with atenolol in immediate treatment of suspected myocardial infarction. *Br Med J* 1980;**280**:885-8.
- Baber NS, Lewis JA. Beta-blockers in the treatment of myocardial infarction. *Br Med J* 1980;**281**:59.
- Rose G. Beta-blockers in immediate treatment of myocardial infarction. *Br Med J* 1980;**280**:1088.
- Elwood PC, Sweetnam PM. Aspirin and secondary mortality after myocardial infarction. *Lancet* 1979;ii:1313-5.
- 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.
- Peto R, Pike MC, Armitage P, *et al*. Design and analysis of randomised clinical trials requiring prolonged observation of each patient. I Introduction and design. *Br J Cancer* 1976;**34**:585-612.
- Peto R. Clinical trial methodology. *Biomedicine Special Issue* 1978;**28**:24-36.
- Chalmers TC, Matta RJ, Smith H, Kunzler AM. Evidence favouring the use of anticoagulants in the hospital phase of acute myocardial infarction. *N Engl J Med* 1977;**297**:1091-6.
- Pocock SJ. Statistical methods in the design and analysis of long-term clinical trials. *Drug treatment and prevention in cerebrovascular disorders*. Amsterdam: Elsevier-North Holland Biomedical Press, 1979.

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

A child of 2 has had mild infantile eczema. There is a paternal history of hay fever and a maternal one of migraine. Is it wise for him to be inoculated against measles and pertussis? (He has already had DPT inoculations.)

The presence of a mild infantile eczema, and a family history of hay fever and migraine, are not contraindications for immunisation against measles and pertussis. The history of migraine is totally irrelevant; as for the allergy, there is possibly theoretically a very slightly increased risk of an anaphylactoid reaction to the injection; I have fortunately never seen it happen, but one should never begin an immunisation session without ensuring that adrenaline and suitable syringe are ready. This is my opinion: I am aware that various bodies regard allergy as a contraindication to immunisation, because, I presume, of fear of litigation. If all the so-called contraindications for immunisation, repeated from book to book without discoverable supporting evidence, are adhered to, thousands of children will suffer for it, and some will die.

What is the sacroiliac syndrome?

Sacroiliac pain may be referred pain from the lumbar spine, as an instance of disease in the lumbar spine or pain referred from one component of the lumbar discs, or it may also be pain referred from a fracture of the upper lumbar spine, where the fracture has resulted in unilateral tilting. Apart from these obvious causes of sacroiliac pain, however, there is a set of symptoms that leads to what I know as the sacroiliac syndrome—namely, acute pain that comes on in the sacroiliac region in many cases related to sudden acute strain, and in others of gradual onset. This syndrome manifests itself in pain and tenderness in the sacroiliac region and on straining the joint. The important factor is that there are no neurological features and, in contradistinction to a disc derangement, straight-leg raising is full and free on both sides. That this is confined to the sacroiliac region may be confirmed

by injecting the joint with local anaesthetic, and often following this with manipulation will cure the condition. There are those who say it is such a close-bound joint that it does not move and therefore should not cause any problems, but undoubtedly there is a syndrome where pain is localised to this joint with no neurological features, no referred features, and normal straight-leg raising, which does respond to manipulation of that joint.

What is the rationale for starving a 1-year-old child with acute diarrhoea?

This has been the subject of much debate, and there are good discussions of the matter in paediatric publications.¹⁻⁴ For a mild diarrhoea there may not be a strong case for excluding food; but for severe cases it is normal, partly because of the danger of high solute foods, such as milk, causing hypernatraemia. There seems little doubt that milk feeds increase the volume of diarrhoeic stools; but even so absorption of nitrogen, fat, sodium, potassium, calcium, and chloride may be increased.^{2,3}

No one would dispute, however, that plain fluids (electrolyte preparations) should be given by mouth unless there is severe vomiting. Hydration is more important than anything else in treatment. I quote up-to-date advice.⁵ "The bowel is put at rest by cutting back on the dietary load and offering clear fluids or a glucose electrolyte solution. This allows mucosal healing and eliminates compounding factors resulting from the effect of unabsorbed osmotically active substances in the bowel lumen." The author goes on to advise against pectin, opiates, diphenoxylate, or antibiotics.

¹ O'Keefe FS. Enteral treatment of diarrhoea in infancy, including epidemic diarrhoea of the newborn. *Am J Dis Child* 1948;**76**:616-25.

² Chung AW. The effect of oral feeding at different levels on the absorption of food-stuff in infantile diarrhoea. *J Pediatr* 1948;**33**:1-13.

³ Chung AW, Yiscorova B. The effect of early oral feeding versus early oral starvation on the course of infantile diarrhoea. *J Pediatr* 1948;**33**:14-22.

⁴ Holt LE. Oral feeding in diarrhoea and hypertonic dehydration. *Pediatrics* 1958;**22**:1024.

⁵ Green M, Haggerty RJ. *Ambulatory pediatrics*. Philadelphia: Saunders, 1977.