Statistics in Question

SHEILA M GORE

ASSESSING CLINICAL TRIALS—FIRST STEPS

Leading articles,¹⁻⁴ correspondence,⁵⁻⁹ research papers,^{10,11} and elementary series¹²⁻¹⁴ in the *British Medical Journal* make up a valuable introduction to scientific method and clinical trials. Statistics in Question is an opportunity for revision. The series has a question and answer format with illustrations, and it is in two parts. Assessing clinical trials includes material on medical ethics, consultation, trial size, design, randomisation, observer variability, trial procedures, record sheets, and data checking. The second part deals with statistical methods commonly reported in medical journals and answers questions about their application.

Medical ethics

(1) *Why should the publication of selected series be discouraged if randomised comparison is appropriate?*

—non-randomised studies do not necessarily get the correct answer because of selection bias

—to avoid ethical conflicts randomisation should begin with the first patient¹⁵

—non-randomised studies lead to rash verdicts

—they do not convince the sceptics

COMMENT

Because of selection bias non-randomised studies can give falsely optimistic^{16,17} or even wrong¹⁸ reports of new treatments. Adrenalectomy for essential hypertension—introduced without proper evaluation—was once widely practised but later abandoned because it was of no value.

Again, jejunioleal bypass was introduced without a randomised clinical trial and was widely adopted. Because published reports and their own experience suggested a serious complication rate,¹⁹ Danish surgeons carried out a randomised trial between jejunioleal bypass surgery and medical treatment for patients with morbid obesity.²⁰ The patients who had bypass surgery

lost more weight and had a much improved quality of life, though there were complications, some severe. Non-randomised studies had not made their point.

This trial was criticised^{22,23} because patients had been deceived—the patients who were randomised to medical treatment were told that the results of liver biopsy contradicted surgery. The Danes had faced a dilemma—whether to continue using a treatment about which they had serious doubts or run a randomised trial in which they had to deceive

Dilemma

- 1 Only a randomised trial could convince the Danes
- 2 Uncritical acclaim by the media had persuaded patients to seek jejunioleal bypass
- 3 'Informed' consent to medical treatment would probably have been refused
- 4 Failure to request informed consent is unethical

Treatments introduced without proper evaluation

- 1 Adrenalectomy for essential hypertension¹⁸
- 2 Adenotonsillectomy in children and young adults²¹
- 3 Jejunioleal bypass in morbid obesity²⁰

patients. Ethical conflicts could be avoided in future if all procedures, including operations, were evaluated in randomised controlled trials before they were generally adopted.

Smithells *et al*²⁴ gave periconceptional vitamin supplements to women at high risk of having a baby with a neural-tube defect. One out of 178 infants or fetuses (0.6%) had a neural-tube defect. But among the infants or fetuses of high-risk

Vitamins, neural-tube defects, and ethics committees

women who were already pregnant or who had refused supplements, 13 out of 260 (5%) had a defect. The study was not randomised because ethics committees at three of the five study centres had refused permission for high-risk women to be assigned randomly to placebo.²⁵ The poor design of the trial meant that the results were difficult to interpret.²⁶⁻²⁹ Selection bias or accidental bias, or both, might account for the different outcomes in the treated and control mothers.

Two of the authors wrote²⁵: "In our paper we subscribe to no belief, reach no conclusions, and offer four possible interpretations. . . ." A less cautious editorial verdict³⁰ is reason enough to discourage publication of results from selected series when randomisation is appropriate. The pity is that further studies are needed for doctors to know whether or not they can reassure high-risk patients that periconceptional vitamin supplements lessen the risk of neural-tube defect.³¹

(2) *When a clinical trial protocol is accepted by the hospital ethics committee is this a sufficient guarantee of scientific merit?*

—no, protocol review is outside the remit of most ethics committees

—ethics committees usually are only concerned with what is done to the patient

COMMENT

Acceptance of a clinical trial protocol by the hospital ethics committee is not sufficient evidence of scientific value, though it is necessary evidence. Ethical approval was given for the investigation by Smithells *et al* of the possible prevention of neural-tube defects.²⁴ The control group was self-selected because of either being pregnant or refusing to take periconceptional vitamin supplements. Selection bias could not be dismissed as an explanation for the results, and thus the study had limited scientific value.

Altman has discussed the ethical implications of inferior design.³¹ But ethics committees are usually only concerned with what is done to the patient,³² and they are not set up to inquire into the design or size of clinical trials³³; it is unusual for a statistician to be on an ethics committee. Protocol review is clearly outside the remit of most ethics committees.

Consultation

(3) *What are the five questions a statistician is most likely to ask before advising a doctor on the design of a clinical trial?*

What is the precise aim of the investigation?

What is the clinical background?

Are resources limited?

What is the response distribution?

What would be regarded as a clinically important difference between treatments?

COMMENT

The first consultation between the doctor and the statistician should establish the precise aim of the investigation. Trial design is then tailored to fit the right problem. The statistician

"An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem."

John Tukey

will need information about the clinical background, and may ask the doctor for references to recent research articles. Limitation of resources imposes important constraints on the design of a clinical trial. Resources include hospital staff and equipment, time, data processing and secretarial help, finance, and of course patients. The availability of patients is commonly, if unwittingly, overestimated by doctors. The statistician should

Lasagna's law: the availability of patients is commonly overestimated, even by a factor of 10.

find out how many suitable patients were referred in, say, the previous six months. There is little point in starting a clinical trial whose duration extends beyond the time for which clinical and statistical supervision can be guaranteed.

To determine the number of patients in a clinical trial the statistician must learn something about the response distribution. This could be binary, Gaussian (normal), lognormal, exponential, Poisson, or some other distribution. The doctor may

provide this information or it can be deduced from previous research, abstracted from case notes, or obtained in a pilot study. The doctor must assess the difference between treatments that is clinically important before the statistician can calculate how large the trial should be. Trial size is determined so that if a worthwhile difference exists between treatments the difference will probably be evident, and also statistically significant, in the results.

Pilot studies

(4) Why are pilot studies done?

—to estimate the rate of patient accrual or the variability of treatment outcome

—to test questionnaire, interview, or assessment

—to check that the trial runs smoothly

COMMENT

Pilot studies are devised to estimate patient accrual and the variability of response if this information, which is needed to calculate the size and likely duration of a clinical trial, is not otherwise available (see question 3).

A different type of pilot study is called for when a questionnaire, interview, or assessment scale is used to record information about patients. Form design³⁴ is not a trivial exercise—content, length, wording, appearance, and layout must be carefully considered. The draft questionnaire should then be tried out (piloted) on a sample of patients to identify language that is difficult to understand or questions that are frankly ambiguous. Do the answers accord with expectation? For example, is height recorded in inches or centimetres?

Startling replies may be elicited by an interviewer. Timing is important in interviews, and it is wise not to rush patients. When a bronchitic patient was asked if he brought up sputum in the morning, he answered, "No," paused, and added as an afterthought, "I swallow it again."

To measure stroke rehabilitation Garraway *et al*³⁵ devised a rehabilitation scale. In a series of pilot studies they checked that observers could agree on the interpretation and use of the scale.

Pilot studies: Clinical assessment of stroke

The classic reason for doing a pilot study is to accustom laboratories, hospital staff, and patients to the trial procedures³⁶ and thus ensure that the experiment runs smoothly. The pilot study mirrors the actual trial in every detail, so that if no amendment is needed the results from patients who have been randomised in the pilot study can be analysed along with the main results.

I am grateful to the editor of *Age and Ageing* for permission to publish the figure on clinical assessment of stroke.

References

- Anonymous. Statistical errors. *Br Med J* 1977;ii:66.
- Anonymous. Randomised clinical trials. *Br Med J* 1977;ii:1238-9.
- Anonymous. Interpreting clinical trials. *Br Med J* 1978;ii:1318.
- Anonymous. The case-control study. *Br Med J* 1979;ii:884-5.
- Pocock SJ. Randomised clinical trials. *Br Med J* 1977;ii:1661.
- Peto R, Doll R. When is significant not significant? *Br Med J* 1977;ii:259.
- Anderson TW, Ashley MJ, Clarke EA. Not so double-blind? *Br Med J* 1976;ii:457-8.
- Newell DJ. Type II errors and ethics. *Br Med J* 1978;ii:1789.
- Cochrane AL, St Leger AS, Sweetnam P. Emotion and empiricism. *Br Med J* 1979;ii:486.
- Gore SM, Jones IG, Rytter EC. Misuse of statistical methods: critical assessment of articles in *BMJ* from January to March 1976. *Br Med J* 1977;ii:85-7.
- Lionel NDW, Herxheimer A. Assessing reports of therapeutic trials. *Br Med J* 1970;iii:637-9.
- Swinscow TDV. *Statistics at square one*. London: British Medical Association, 1977.
- Rose G, Barker DJP. *Epidemiology for the uninitiated*. London: British Medical Association, 1979.
- Altman DG. Statistics and ethics in medical research. *Br Med J* 1980;281:1182-4.
- Chalmers TC. When should randomisation begin? *Lancet* 1968;ii:858.
- Doll R, Peto R. Randomised controlled trials and retrospective controls. *Br Med J* 1980;280:44.
- Colton T. *Statistics in medicine*. Boston: Little, Brown, 1974.
- Ederer F. The randomised clinical trial. In: Phillips CI, Wolfe JN, eds. *Clinical practice and economics*. London: Pitman, 1977.
- Anderson B. Research ethics and deception. *Lancet* 1980;ii:772.
- The Danish Obesity Project. Randomised trial of jejunoileal bypass versus medical treatment in morbid obesity. *Lancet* 1979;ii:1255-7.
- Roydhouse N. A controlled study of adenotonsillectomy. *Lancet* 1969;iii:931-2.
- Anonymous. Bypassing obesity. *Lancet* 1979;ii:1275-6.
- Berenbaum MC. Research ethics and deception. *Lancet* 1980;ii:772.
- Smithells RW, Sheppard S, Seller MJ, *et al*. Possible prevention of neural-tube defects by periconceptional vitamin supplementation. *Lancet* 1980;ii:339-40.
- Smithells RW, Sheppard S. Possible prevention of neural-tube defects by periconceptional vitamin supplementation. *Lancet* 1980;ii:647.
- Kirke PN. Vitamins, neural tube defects, and ethics committees. *Lancet* 1980;ii:1300-1.
- Freed DLJ. Vitamins, neural tube defects, and ethics committees. *Lancet* 1980;ii:1301.
- Raab GM, Gore SM. Vitamins, neural tube defects, and ethics committees. *Lancet* 1980;ii:1301.
- Elwood JH. Possible prevention of neural-tube defects by periconceptional vitamin supplementation. *Lancet* 1980;ii:648.
- Anonymous. Vitamins, neural-tube defects and ethics committees. *Lancet* 1980;ii:1601-2.
- Altman DG. Statistics and ethics in medical research. Study design. *Br Med J* 1980;281:1267-9.
- Thomson IE, French K, Melia KM, *et al*. Research ethical committees in Scotland. *Br Med J* 1981;282:718-20.
- Altman DH. Statistics and ethics in medical research. III How large a sample? *Br Med J* 1980;281:1336-8.
- Wright P, Haybittle J. Design of forms for clinical trials (1-3). *Br Med J* 1979;ii:529-30, 590-2, 650-1.
- Garraway WM, Akhtar AJ, Gore SM, Prescott RJ, Smith RG. Observer variation in the clinical assessment of stroke. *Age Ageing* 1976;5:233-40.
- National Study by the Royal College of Radiologists. A study of the utilisation of skull radiography in 9 accident-and-emergency units in the UK. *Lancet* 1980;ii:1234-6.

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

This is the first of a series of articles on assessing clinical trials and statistical methods.

After a normal birth, what is the risk to a baby whose mother was subsequently found to have had an active Trichomonas vaginalis infection during the delivery?

Maternal *Trichomonas vaginalis* infection may cause neonatal vaginitis¹ or urinary tract infection.²

¹ Al-Salihi FL, Curran JP, Wang J-S. Neonatal trichomonas vaginalis: report of three cases and review of the literature. *Pediatrics* 1974;53:196-200.

² Crowther IA. Trichomonas vaginitis in infancy. *Lancet* 1962;ii:1074.

³ Littlewood JM, Kohler HG. Urinary tract infection by *Trichomonas vaginalis* in a newborn baby. *Arch Dis Child* 1966;41:693-5.