to remind ourselves of what procedures such a doctor is carrying out and that dementia and alcoholism are, unfortunately, not new among practising doctors.

The central element of managing infected doctors and protecting their patients is to create an environment in which this management is possible. If, for example, it was considered desirable to dissuade infected surgeons from operating how would they be identified? Surely not through obligatory screening of all surgeons; this makes no more sense than screening all patients to "protect" surgeons. Total confidentiality must be preserved so that doctors can consult their colleagues if they are fearful of having become infected. Lack of confidentiality will discourage such individuals from coming forward and must be judged as detrimental to their health care. Surely the court case was about protecting the principles of confidentiality and was not an effort at a cover up. Without the certainty of confidentiality infected doctors and all other patients will not consult, and the public health will thus be compromised. If infected doctors are driven underground they will not be aware of their antibody state and thus will not be offered care, assessment, and counselling. This might lead to ignorance of their deteriorating health and performance and dementia and failure to become educated about adequate infection control procedures. Doctors with HIV infection and AIDS can be best helped by creating a non-Draconian, confidential, and sympathetic environment that encourages them to seek help. Ultimately this will be the best way for patients and doctors to be protected.

MICHAEL W ADLER

Professor of Genitourinary Medicine,
University College and Middlesex School of Medicine,
James Pringle House,
Middlesex Hospital,
London W1N 8AA

Regular Review

Clinical experimentation in obstetrics

R J LILFORD

Clinical decisions should be based on well conducted clinical experiments. Control patients must be randomly chosen in order to avoid bias (inaccuracy) and measures must be taken to ensure that the important variables are equally distributed between treatment and control groups.[1] Unfortunately, obstetricians have been found wanting in commitment to follow these principles.[2]

Following these principles in perinatal medicine is, however, particularly difficult. Enormous samples are needed to show the small changes that are hoped for in outcomes such as stillbirth, handicap, and congenital abnormality rates because most pregnancies are successful. In contrast, physicians deal mostly with chronic symptoms, where trials on one patient may be possible,[3] or with conditions carrying a high mortality. Thus, though a trial with about 1000 patients in each arm is large enough to show a 5% reduction in a common event such as death after myocardial infarction, a trial 100 times larger is needed to show the same percentage reduction in the intrapartum stillbirth rate at the same level of significance. Small trials lack power—the ability to detect a true difference—and where statistical significance is present the magnitude of the difference is likely to be exaggerated.[4,5] The difficulty of a clinical trial in obstetrics will depend largely on the number of patients required and the ease with which they can be recruited. Some trials can realistically be mounted within a short time and carried out in a single hospital unit, others require multicentre collaboration, while in some cases the requisite population size renders the question unanswerable.

Trials feasible in a single unit

Trials that can be carried out in a single unit are usually those where the endpoint is a continuous variable or ordinal scale rather than a proportion (rate). Thus trials with fewer than 100 patients in each arm will detect a 5% drop in mean blood pressure, a 0·05 change in pH in fetal blood, or a gain of two weeks before delivery. (These figures are based on a power of 85%, a significance level of 5%, and reasonable estimates for standard deviation.) Sometimes these are important measures in themselves—for example, gestational age—but usually they are surrogate measures for other more important endpoints.

Lactic acid concentration, pH of fetal blood, or Apgar score at birth may be used as surrogate measures for stillbirth or asphyxial brain damage; the assumption is that any measure with a favourable effect on these variables will also favourably change certain more important endpoints. This is a dangerous extrapolation because variables such as the mean fetal blood pH at birth may be measurably improved by a small effect on many people while a larger, deleterious effect on a rare endpoint, such as stillbirth, will go undetected.

Trials requiring larger populations

Differences in proportions (rates) are usually much more difficult to detect. Occasionally in obstetrics we are interested in common events such as the incidence of dyspareunia after

perineal suture or the forceps rate with alternate birth positions, but usually we are concerned with rare endpoints.

When we are studying routine or screening measures the population is large but mortality and major morbidity are rare. In my hospital we have started several measures that are intended to reduce perinatal disasters—a fetal assessment laboratory, computerised data management, and routine ultrasonographic assessment of fetal growth. We should test these technologies in a randomised comparison with existing methods to establish that they produce benefits—for example, a 10% improvement in our perinatal mortality rate of 8·5 for every 1000 births. It would, however, be necessary to randomise over 32 000 patients to have a modest (80%) chance of showing this reduction at 5% significance.

Sometimes both the condition under investigation and the outcome are uncommon—for example, the effect of caesarean section in managing fetuses presenting in the breech position. Reports of uncontrolled trials suggest that the increased perinatal mortality of a trial of selected vaginal breech delivery is only 2·5 in 1000 births.[6] To show that elective caesarean section could reduce intrapartum and neonatal death rates from, say, 5 to 2·5 in every 1000 births a trial would need nearly 10 000 patients in each arm (significance 5%; power 80%). This number of subjects would accrue from about one million births, which would mean randomising all eligible breech presentations in the British Isles over two years.

Number of pregnancies that would generate enough patients to show a 20% change in perinatal mortality with different management policies; power 80%, significance level 5%

Hypothesis	Perinatal mortality in condition (per 1000 births)	Study size: power 80% p<0·05†	Approximate prevalence of condition in obstetric population (%)	Total numbers of pregnancies required for study
Glucose tolerance test for all pregnant women or in selected patients	8·5	83 000	100	92 000
Home *v* hospital delivery for "low risk" women	4	176 000	25	704 000
Induction of labour *v* conservative management for postmaturity	20*	35 000	5	394 000
Caesarean section for all babies between 24 and 28 weeks' gestation *v* trial of vaginal delivery	400	1 100	0·1	1 100 100

*This figure is a guess as there are no reliable recent data.
†There is nothing sacred in the 5% significance level, and I believe this should be relaxed where the consequences of erroneous rejection of the null hypothesis are no worse than those of failure to detect a true difference.

Larger differences in mortality and neurological morbidity might be expected from different treatment options in problems such as the premature fetus in the breech presentation, the very premature infant (less than 28 weeks' gestation), or the second twin. The difficulties here are that these conditions are even rarer, and individual departments will encounter very few cases each year. This, of course, is a problem for all doctors who study rare conditions.[7] Obstetricians, however, often find themselves in a catch 22 position—either the condition is common but perinatal death or damage is rare, or perinatal death or damage is common but the condition is rare (table).

Implications for trial design

These observations have important implications for the design of trials and reporting of results. Small trials can detect changes in surrogate measures or refute extravagant claims, but sample sizes sufficient to detect the small differences in mortality or morbidity rates that would nevertheless influence clinical practice are much larger than investigators realise.

An estimate (or better a measure) of prospective parents' values should be stated when the trial is designed and reported. The sample size might thus depend on the reduction in neurological morbidity or perinatal mortality that patients will choose in preference to, say, the danger and pain of caesarean section or the inconvenience, worry, and expense of intensive antenatal fetal monitoring. In a hypothetical study of value trade offs we found that most women would subject themselves to unpleasant procedures, such as operative delivery, to reduce fetal risk by one in many thousands.[8] Elegant techniques for assessing human values and trade offs have been developed by economists and psychologists.[9]

The results of negative trials should not lead to firm recommendations on clinical practice unless they were capable of detecting clinically important effects; this obvious requirement is often violated in reports in perinatal medicine.[8 10] In some cases, however, all treatment options are so safe (with fetal risks of about 1 in 10 000) that trials to show changes in mortality and major morbidity might be not only impossible but also unnecessary. Many obstetric decisions—for example, the management of ruptured membranes without contractions in a woman at term—seem to carry this order of risk, and considerations other than safety will have the greatest influence on management decisions.

Other problems for the obstetrician-scientist

The clinical scientist faces many other problems in perinatal medicine. As in all surgical specialties it is often necessary to randomise operators and patients because of varying degrees of skill. There are difficulties in ensuring compliance with randomisation, especially when other members of the team, such as midwives, are less enthusiastic about the study.

When doctors have no reason to prefer one treatment method over another, they are said to be in collective "equipose"[11] and a trial would be entirely ethical because the control group, while denied any benefit of the "new" treatment, would be saved from possible harm. The ethics of randomisation may be less clear when the strength of belief in the superiority of a particular treatment is intermediate between complete uncertainty and complete certainty.[12] This objection was raised in the Medical Research Council trial of multivitamins for preventing recurrent neural tube defects; previous evidence for a net beneficial effect was strong but inconclusive.[13 14] The method by which consent should be obtained for trials in which doctors, collectively or individually, have a preference for one method of treatment is controversial.[15-18] Epidemiologists are more enthusiastic about such trials than clinicians as they are trained to be sceptical and are more conscious of the duty to provide the community with accurate data.[19 20]

Finally, trials may distort the difference in outcome between groups since few obstetric trials can be double blind.

This "Hawthorne effect" may explain the generally better outcome for both the study and the control groups during clinical trials. It may also, however, bring about a greater effect in one group rather than the other. This is particularly likely in trials of fetal monitoring, where the conditions of the experiment are likely to alter vigilance, particularly in the control group.[21]

Many of these problems can be solved with patience, staff training, imagination, pragmatic cluster allocation, methodical procedure, and repeated experimentation. The most intractable problem is that of sample size.

Confronting the difficulties

Randomised trials have gained widespread acceptance among obstetricians in the past few years, and this trend should be strongly encouraged to limit haphazard diffusion of new treatments.[22] Though some management options are already so safe that further reductions in hard endpoints will be practically impossible to show, others can be evaluated by multicentre trials. The National Perinatal Epidemiology Unit in Oxford has already made an immense contribution by coordinating collaborative efforts and giving advice to individual investigators. A useful register of randomised and quasirandomised trials had been created by this unit and this can be used to amalgamate the results of individual studies (meta-analysis).[23]

The logistic difficulties of mounting large multicentre studies are considerable and add greatly to the expense, but computers may help. Most obstetric departments are acquiring obstetric information systems. These should be adapted to automatically stratify and randomise patients on entry into a trial and to provide much of the information necessary for subsequent analysis. Electronic linkage of computers in collaborating centres will enable this information to be transferred, without identification details, to a central resource for final analysis. The use of computers will not only save time, money, and effort but also will prevent cheating at randomisation and bias due to missing data. This will be much more valuable than the present passive archival functions of most departmental computers.

Finally, randomised trials do not necessarily show the best management option, although they do provide the best information for clinical decisions. Thus, if a trial could show that planned caesarean section is slightly safer for delivering the fetus in the breech position this would not of itself confirm that this should be recommended to all mothers. Other factors must contribute to the final decision—such as the relative maternal morbidity and mortality of vaginal delivery and elective and intrapartum caesarean section, the failure rate of trials of vaginal delivery, the subsequent effects of a uterine scar, and psychological factors. These factors could be combined intuitively or analysed formally and mathematically in a clinical decision analysis—an exciting subject about which we shall hear much more in the next few years.

R J LILFORD

Professor of Obstetrics and Gynaecology,
St James's University Hospital,
Leeds LS9 7TF

1 Taves DR. Minimization: A new method of assigning patients to treatment and control groups. *Clin Pharmacol Ther* 1974;15:443-53.

2 Chalmers I, Sinclair JC. Promoting perinatal health: is it time for a change in emphasis in research? *Early Hum Dev* 1985;10:171-91.

3 Guyatt GH, Sackett D, Wayne Taylor D, *et al.* Determining optimal therapy—Randomized trials in individual patients. *N Engl J Med* 1986;314:889-92.

4 Williams NR, Lilford RJ, Osborne J. A basic program to perform power calculations. *Computers in Patient Care Survey* 1987;4:4-7.

5 Altman DG. Size of clinical trials. *Br Med J* 1983;286:1842-3.

6 Bingham P, Lilford RJ. The management of the selected term breech presentation: assessment of the risks of selected vaginal delivery versus caesarean section for all cases. *Obstet Gynecol* 1987;69:965-78.

7 Schaffner F. The difficulties of controlled trials. *Br Med J* 1985;290:1353.

8 Barrett JFR, Tyrrell S, Lilford RJ. Prolonged pregnancy. *Br Med J* 1987;294:56-7.

9 Torrance GW. Measurement of health state utilities for economic appraisal. A review. *Journal of Health Economics* 1986;5:1-30.

10 Cardozo L, Fysh J, Pearce JM. Prolonged pregnancy: the management debate. *Br Med J* 1986;293:1059-63.

11 Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med* 1987;317:141-5.

12 Lockwood M, Anscombe GEM. Sins of omission? The non-treatment of controls in clinical trials. *Aristotelian Society* 1983;57:207-22.

13 Smithells RW, Sheppard S, Schorah CJ, *et al.* Vitamin deficiences and neural tube defects. *Arch Dis Child* 1976;51:944-50.

14 Smithells RW, Seller MJ, Harris R, *et al.* Further experience of vitamin supplementation for prevention of neural tube defect recurrences. *Lancet* 1983;i:1027-31.

15 Ingelfinger FJ. Informed (but uneducated) consent. *N Engl J Med* 1972;287:465-6.

16 Simes RJ, Tattersall MHN, Coates AS, Raghavan D, Solomon HJ, Smartt H. Randomised comparison of procedures for obtaining informed consent in clinical trials of treatment for cancer. *Br Med J* 1986;293:1065-8.

17 Cancer Research Campaign Working Party in Breast Conservation. Informed consent: ethical, legal, and medical implications for doctors and patients who participate in randomised clinical trials. *Br Med J* 1983;286:1117-21.

18 Cassileth BR, Zupkis RV, Sutton-Smith K, March V. Informed consent. Why are its goals imperfectly realized? *N Engl J Med* 1980;302:896-900.

19 Challah S, Mays NB. The randomised controlled trial in the evaluation of new technology: a case study. *Br Med J* 1986;292:877-8.

20 Hellman S. Randomized clinical trials and the doctor-patient relationship. *Cancer Clinical Trials* 1979;2:189-93.

21 MacDonald D, Grant A, Sheridan-Pereira M, Boylan P, Chalmers I. The Dublin randomized controlled trial of intrapartum fetal heart rate monitoring. *Am J Obstet Gynecol* 1985;152:524-39.

22 Tarnow-Mordi W, Wilkinson A. Mechanical ventilation of the newborn. *Br Med J* 1986;292:575.

23 Chalmers I, Hetherington J, Newdick M, *et al.* The Oxford database of Perinatal trials: developing a register of published reports of controlled trials. *Controlled Clin Trials* 1986;7:306-24.