EDUCATION & DEBATE

When to stop a clinical trial

Stuart J Pocock

Most randomised clinical trials require periodic monitoring of the accumulating data. While the efficiency of trial management is enhanced by data monitoring, ethical reasons should primarily dictate the need to terminate or change a trial in response to interim findings.

This article focuses on the ethical dilemma of when to stop a clinical trial and places statistical stopping rules in the context of such ethical decision making. Other issues include the organisation of data monitoring committees and the problems of premature publication and exaggerated estimation in trials that stop early. Several topical examples are used to convey the relevance of these issues to current practice.

The ethical dilemma

The basic ethical conflict in monitoring trial results is to balance the interests of patients within the trial— that is, the individual ethics of randomising the next patient—and the longer term interest of obtaining reliable conclusions on sufficient data—that is, the collective ethics of making appropriate treatment policies for future patients.

One common misperception is to concentrate exclusively on individual ethics. For instance, suppose a trial has randomised two patients, one to treatment A who died and one to treatment B who was still alive. Extreme use of individual ethics requires that the next patient should receive full information, including these findings. If there was no other information to distinguish between treatments such "ethical" revelation could make it difficult to randomise the next patient ("I personally would like treatment B, please"). However, it is impossible to reach reliable conclusions with only two patients (both statistical significance and clinical common sense would be lacking). Such full pursuit of individual ethics to the exclusion of collective ethics seriously conflicts with the conduct of randomised controlled trials. Unbiased and precise comparison of treatments would become impossible, and the development of new treatments would be chaotic and unscientific.

The real ethical conflict is faced when rather more information has accrued. For instance, in a randomised trial comparing extracorporeal membrane oxygenation (ECMO) with conventional medical treatment in newborns with persistent pulmonary hypertension[1][2] a decision was taken to halt randomisation when the data disclosed four deaths among 10 infants receiving conventional medical treatment compared with none among nine infants having extracorporeal membrane oxygenation (p=0·054). Full details are debated elsewhere,[2] but the main issue is that randomisation stopped early on the basis of a fairly small amount of data, all subsequent patients being allocated to extracorporeal membrane oxygenation. The investigators were sensitive to the individual ethics of seeking parental consent and randomisation for the next newborn infant in the light of the four

---

Ethical dilemma

Individual ethics—needs of next eligible patient
Collective ethics—correct policy for the future
Never randomise to an established inferior treatment
Need evidence sufficient to change clinical practice for the better
Failure to achieve a clear result is unethical?

---

versus nil deaths. However, with only 19 patients this does not represent strong evidence of the superiority of extracorporeal membrane oxygenation and provides little scope for making reliable judgments on the benefits of this treatment for universal use in such newborn infants in the future. Thus collective ethics may have been compromised by such early stopping.

Debate continues on the merits of extracorporeal membrane oxygenation, and there is the prospect of a subsequent British randomised trial to resolve the issue. The more serious the disease, especially in infants, the more highlighted the ethical dilemma becomes. In this case the lack of an independent data monitoring committee may have prevented a more balanced resolution of the ethical conflict, leading to the investigators' understandable reliance on individual ethics. However, if extracorporeal membrane oxygenation really is effective the prolonged uncertainties maintained by lack of really substantial evidence may well have led to fewer newborn infants worldwide receiving it than would have been the case had the trial continued longer.

In contrast, we can argue that the second international study of infarct survival (ISIS-2) illustrates the other extreme, collective ethics being very much to the fore.[3] In that trial 17 187 patients with myocardial infarction were randomised to streptokinase or placebo (the additional aspirin comparison need not concern us here) and the five week death rates were 9·2% in the streptokinase treated group and 12·0% in the placebo group, p<0·00000001. Clearly, with such extreme significance substantial evidence of streptokinase efficacy existed earlier. However, the data monitoring committee believed its "ethical responsibility was to report the results when they would be likely to change medical practice in the future."[4] The estimated mortality odds reduction due to streptokinase was 25% with a fairly narrow 95% confidence interval of 18% to 32%. Note that the absolute mortality reduction was 2·8%, such estimates of absolute benefit or risk per patient treated being of considerable relevance to decision making. If the trial had stopped sooner these estimates would have been less precise and the impact on clinical practice would have been less. Also the ability to report subgroup findings—for example, by time since onset—would have been diminished.

The background here illustrates the difficulty of getting results of clinical trials to change routine clinical practice. A previous overview of 22 small trials

Medical Statistics Unit, London School of Hygiene and Tropical Medicine, London WC1E 7HT
Stuart J Pocock, *professor of medical statistics*

produced similar evidence,[5] but thrombolytic treatment was still rarely used. It is argued that only by such megatrials as ISIS-2 and GISSI (the Italian trial of streptokinase in myocardial infarction)[6] achieving "proof beyond reasonable doubt," with evidence well beyond conventional levels of statistical significance, can we convince clinicians to change their treatment policies. The eventual success of this approach is without question, but concern has been expressed about individual ethics, as investigators were still allowed to randomise patients to placebo even when the overall benefits of streptokinase were established. One riposte is to emphasise the duality of ethical standards, whereby trialists have to defend their ethical position while those in routine clinical practice can pursue misguided policies (like not using thrombolytic treatment) without ethical accountability. Given this haphazard reality of clinical practice, a reasonable case can be made for trials based on collective ethics in order to exert appropriate pressure for change, especially in situations where the expected size of effect is modest.

A relevant concept here is clinical equipoise—that is, a state of "genuine uncertainty within the expert medical community" which makes it ethical to pursue a particular clinical trial.[7] While reviewing interim results, such clinical equipoise (that is, substantial uncertainty) is required by the data monitoring committee in order for the trial to continue. However, this concept may lack clinical reality as trials need to impact on the broader realm of (non-expert) routine practice, where disagreement among clinicians is often coupled with the force of habit.

For instance, the ISIS-2 trial organisers think the ethical dilemma (whether to randomise) is primarily an individual choice for each collaborating doctor—that is, organisers can inform fully but the "uncertainty principle" rests with individual doctors. We can then claim that individual ethics at each doctor-patient contact were maintained in ISIS-2, provided that we accept transferring the ethical choice to collaborators with the trial organisers adopting an advisory rather than a decision making role. Thus several months before recruitment stopped investigators were informed of the benefits of streptokinase in patients treated within four hours of pain onset but were still allowed to randomise as they saw fit. This position is not universally accepted, and some trialists think that ISIS-2 should have stopped earlier, certainly for patients randomised within four hours of pain onset.

Statistical stopping guidelines

At each interim analysis one key to evaluating the strength of evidence for stopping is the statistical significance of the primary treatment comparison. There is extensive published work on statistical stopping "rules" (more appropriately called "guidelines"), but in practice the basic concepts are fairly simple (see box).

The plethora of stopping rules[11 12] can be somewhat confusing, but for major trials it is now recognised that early interim analyses based on limited data should require very small p values for stopping, whereas later analyses can have stopping p values somewhat nearer to conventional levels of significance.

The O'Brien and Fleming rule (see example below) is often used for this.[13] The logic here relates to (a) credibility—that is, the medical community is rightly sceptical of small trials even if highly significant; (b) realism—many trials are to evaluate moderate treatment difference, and early, "overdramatic" differences are implausible; (c) statistical properties—having lenient p values for stopping early reduces the power to detect realistic effects (assuming we preserve an appro-

priately small overall type I error—that is, risk of a "false positive" claim).

As illustration, the west of Scotland coronary prevention study is an ongoing placebo controlled trial of pravastatin in over 6000 middle aged men with raised low density lipoprotein cholesterol concentrations followed up for five years.[14] By using the O'Brien and Fleming rule, five analyses are planned at annual intervals with the following guideline p values for stopping applied to each of three primary end points (cardiac mortality, non-fatal myocardial infarction, and their combination): analysis 1, $p < 0.00000001$; analysis 2, $p < 0.0001$; analysis 3, $p < 0.001$; analysis 4, $p < 0.004$; analysis 5 (final), $p < 0.009$.

In practice the analyses will be at annual intervals, although in principle the rule was developed for equally spaced totals of events. This plan allows for an overall type I error of 0.01, the sliding scale of interim p values being sufficiently small that we can claim $p < 0.01$ overall if either one stops early or $p < 0.009$ in the final analysis. A p value of < 0.01 rather than 0.05 is advisable for such a major trial as we need small p values for stopping if collective ethics are of high priority. However, if the final analysis is reached and produces a p value above 0.009 (say, $p = 0.03$) we are still not denied the opportunity to claim some evidence of a treatment effect, perhaps better expressed by the magnitude of reduction in (say) cardiac mortality with its confidence interval.

Note that these are only stopping guidelines (not rules) for treatment efficacy, as the data monitoring committee's recommendations will also depend on other information both within the trial—for example, adverse events, total mortality data— and from other studies.

Statistical guidelines are of value in all major trials only if they relate to the full practical implications of early stopping. For instance, a non-small cell lung cancer trial compared chemotherapy plus radiation with radiation alone.[15] The trial intended to accrue 240 patients but stopped early, using an O'Brien and Fleming rule, when the survival difference in the first 155 patients reached $p = 0.007$ in favour of chemotherapy plus radiation. Subsequent discussion[16 17] questioned whether this trial provided sufficient evidence for widespread use of chemotherapy in advanced lung cancer. The treatment is toxic and expensive, and prognosis is still poor, with only a quarter of patients alive after two years.

With only 155 patients we cannot reliably estimate the magnitude of any survival differences (the confidence interval is wide). Souhami et al argue that "It is a pity that this otherwise excellent trial was not allowed to reach a size at which any difference in survival could be quantified with reasonable certainty. Then, and only then, would a discussion of cost and benefit have any validity."[17] The investigators' concern with indi-

---

Concepts of statistical stopping guidelines

- A sufficiently small p value for treatment difference on a trial's primary end point can be a guideline for when it is ethically desirable to stop a trial

- With several repeated looks at the accumulating data, we need to guard against an excessive risk of a false positive result. This requires smaller p values than the conventional 0.05 level[8 9]

- Though the early sequential methodology is related to continuous monitoring of the gradually accumulating data,[10] it is usually more acceptable to adopt a "group sequential" approach with a limited number of preplanned interim analyses before each data monitoring committee meeting[11 12]

vidual ethics is contradicted by their plea for continuation of other studies elsewhere still in progress to confirm their findings.[18]

The difficulty in finding effective treatment for non-small cell lung cancer is relevant here. Thus a pessimistic prior belief regarding chemotherapy's value makes us suspect that the true survival benefit for chemotherapy is less than that observed. Such "shrinkage" of estimates can be formalised by Bayesian methods[19] to see whether large observed differences on small amounts of data can (or cannot) overwhelm prior scepticism.

One criticism of rules like O'Brien and Fleming is their rigidity in requiring a fixed maximum number of preplanned interim analyses. Greater flexibility is possible with the Peto-Haybittle rule, which simply specifies a fixed p value (often $p < 0.001$) for stopping early.[20] For example, the European myocardial infarction amiodarone trial (EMIAT) is an ongoing study of 1500 patients at high risk after myocardial infarction comparing amiodarone (an antiarrhythmic drug) with placebo. Two year mortality is the primary end point, and the stopping guideline for efficacy is $p < 0.001$ in favour of amiodarone. Though the trial is planned for just three "efficacy" analyses (that is, substantial evidence is needed to counter the lack of such efficacy with other antiarrhythmic drugs), the same rule could have been applied with more frequent looks if desired.

Another flexible approach by Lan and DeMets requires prespecification of the rate at which the type I error is used up as the trial accumulates data,[21] and from this we can work out p values for stopping at any choice of analysis times. Such an approach was used in the cardiac arrhythmia suppression trial (CAST), which stopped early due to excessive mortality in two of the active treatment groups (see below).[22 23] This flexible approach may be distorted by adjusting the timing of future analyses in response to the observed treatment difference so far, but this seems to have little effect on its statistical properties. Also it may bring undue technicality to what should be a simple objective statistical guideline among the complexity of trial design making.

Many other statistical strategies exist, some following conventional (frequentist) methods[24] and others pursuing Bayesian methods.[25] One criticism of Bayesian methods (although useful conceptually) is their failure to allow for the fact that the more often you look at accumulating data the greater the chance of stopping a trial prematurely with a false claim of a treatment difference. However, arguably all data monitoring committees are Bayesians as they judge trial data in the context of prior evidence or belief.

What if the new treatment looks worse?

In any clinical trial we must consider not only "positive" stopping (that is, the new treatment looks better) but also "negative" stopping because either the new treatment exhibits safety problems or lacks efficacy. We cannot lay statistical plans for every potential "negative" contingency, but some logical guidelines are possible. Usually there is an asymmetry between positive and negative stopping—that is, ethically we require weaker evidence if a new treatment looks worse. Several approaches are available.

In the European myocardial infarction amiodarone trial I mentioned above the positive rule ($p < 0.001$) for mortality reduction with amiodarone. For mortality excess with amiodarone a less stringent rule ($p < 0.01$) is in operation but also with more frequent looks at the data (that is, every four months). Such asymmetry raises the issue of whether to use one sided or two sided p values. Personally, I favour universal use of two sided testing, simply because the misguided obsession with

---

Disadvantages of stopping a trial early

Lack of credibility	— small trials not convincing
Lack of realism	— dramatic treatment difference implausible
Imprecision	— wide confidence interval for treatment effect
Bias	— trial liable to stop on a "random high"
Speed	— insufficient time and information to consider overall balance of costs and benefits
Pressure	— unduly enthusiastic and extrapolated recommendations may follow
Mistakes	— risk of false positive result. Hence need extreme statistical significance to stop early for efficacy

---

$p < 0.05$ is just too easy to achieve with a one sided approach.

The PACK trial (prevention of atherosclerotic complications with ketanserin), comparing ketanserin with placebo in 3899 patients with intermittent claudication, had a similar asymmetric rule.[26] However, evidence of a harmful effect of ketanserin in patients taking potassium losing diuretics (35 deaths with ketanserin versus 15 with placebo in these patients) led to an ad hoc decision to halt the trial in this subgroup. Thus, though statistical plans are useful, data monitoring committees must also take decisions based on the unexpected, especially if related to a harmful effect. However, we should not overreact to suggestions of harm (or efficacy) in subgroup analysis.[27]

Should a trial stop before evidence reaches "significance in the wrong direction," as what matters is whether a new treatment is better than the control? For instance, a trial of neutron treatment for pelvic cancers stopped early when the relative risk for mortality after neutron therapy compared with conventional radiotherapy was 1.52, with a 95% confidence interval of 0.91 to 2.50.[28] Although the treatment difference was not statistically significant (the confidence interval included unity), the trial stopped because the confidence interval did not include the sort of mortality reduction after neutron therapy that would make it viable. It had previously been argued that a true 30% mortality reduction (relative risk = 0.7) was the minimum required to change clinical practice. Such use of interim confidence intervals for negative stopping seems well attuned to clinical reality and can be adapted further by using wider intervals to allow for repeated analyses over time.[29]

An alternative approach is to stop if, given the data so far, the chances are small of reaching statistical significance in favour of the new treatment at the end of the trial. The terms for this approach are stochastic curtailment,[30] conditional power,[31] or futility index.[32] However, I question its merit for decision making as it is too focused on the need to achieve significance, but conditional power might usefully be reported after negative early stopping. Trials which fall short of conventional significance can still be consistent with a possible positive effect and contribute to the total evidence, especially if meta-analysis methods are later to be used. It seems a dangerous pathway for trials to stop prematurely for reasons of economy and lack of significance.

The cardiac arrhythmia suppression trial is a dramatic case of negative stopping with 56 deaths or cardiac arrests with two active drugs (encainide or flecainide) compared with 22 deaths or cardiac arrests with matched placebo.[22 23] Advisory guidelines for both

negative significance and stochastic curtailment were employed, although the trial was originally designed as "one sided." Use of interim confidence intervals for mortality comparisons might have given earlier insight into the clear lack of efficacy.

Negative stopping decisions cannot simply be based on statistical guidelines. For instance, if negative interim data are in contradiction with previous, more positive evidence, then the quality, extent, and relevance of that evidence will influence one's judgment.

Independent data monitoring

It is desirable that interim clinical trial results are not known to investigators, as their ability to enter, treat, and evaluate patients according to protocol may be adversely influenced by such knowledge. In major trials it is common practice to establish a data monitoring committee of independent experts who have no other involvement in the trial. Such a committee is periodically supplied with a confidential report of the current interim results by the trial's statistical centre, and the committee's prime function is to recommend whether the trial should be stopped or altered, ethical considerations being of prime concern.
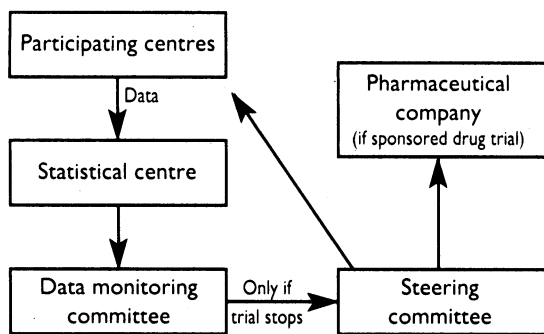
Flow of data in typical committee structure for multicentre trial. Only data monitoring committee and statistician see interim results

It is important that only the data monitoring committee and the trial statistician are aware of the interim results. As the trial is not organised by the data monitoring committee its responsibility should be confined to recommendations to the organisers (often a steering committee) rather than decisions. Usually any data monitoring committee meeting has two options: a recommendation that the trial should continue as planned, with no additional details; or a specific recommendation that the trial should be stopped or altered in some way. This second option usually necessitates full or partial release of interim results to the steering committee so that it can rapidly reach a definite decision to concur or (rarely) disagree with the data monitoring committee's recommendation.

Obviously, the data monitoring committee membership needs experts with considerable experience both of clinical trials in general and of the disease and treatments under investigation, and requires both clinicians and statisticians. The trial statistician producing the results needs to attend the data monitoring committee but is often not a voting member, in which case another statistician member is required. The data monitoring committee is often quite small (minimum three people?), and an odd number is helpful in achieving internal decisions.

Though it is vital that the primary comparative results are not released, practice varies on what lesser information can be made available to the steering committee or investigators. For instance, in a lipid lowering drug trial with coronary events and deaths as primary end points, should one release data on the totals of primary events (both groups combined) or lipid changes or adverse events by treatment group? Such knowledge satisfies curiosity and instills confidence that the trial is functioning well, but could it adversely affect continuation? Knowledge that the total of primary events was well below twice that expected in the control group might lead to (possibly false) speculation that the treatment was effective, in which case such data should not be released. However, one referee disagreed with this view, arguing that a steering committee cannot steer appropriately without access to such information.

Many trials are sponsored by pharmaceutical companies, and it is important to define clearly their participation (if any) in data monitoring.[33] The credibility of a company sponsored trial is enhanced if the company has no part in data monitoring. "If this information is available to a company, it might attempt to terminate the trial early, either because an apparently "good" result has appeared which would be commercially beneficial, or because an apparently "bad" trend suggested that money was being wasted."[33] Most companies have the best of scientific intentions to ensure that trial decisions are made ethically, but in order to avoid any suggestion that commercial reasons could have influence it is essential that the data monitoring committee does not contain any company employees, either officially or in attendance. Also, interim reports should be prepared by an independent statistical centre and not be seen by the company.

Besides these general principles other aspects of trial management will vary. At one extreme all aspects of trial conduct and data processing can be without company involvement, as in the west of Scotland study.[14] This is often impracticable as company resources (trial monitors, international coordinators, data collection facilities) are needed to achieve cost effective, quality trial management. It is then crucial to take practical steps to preserve optimal credibility. Possible options include, firstly, blinding the treatment code during all company data handling (designated company members may need access for emergency unblinding or adverse event reporting but should be committed to secrecy); secondly, having the primary end points—for example, deaths and major clinical events—processed and validated separately from the company, as in the PACK trial.[26] Of course, whatever procedures are adopted the quality and completeness of up to date interim data are of overriding importance. Also, once any trial stops and publication is under way the relations between and responsibilities of investigators, steering committee, data monitoring committee, independent statistical centre, and the company need clear definition.

Publication and interpretation

In line with the confidentiality of interim data it is strongly recommended that no presentation or publication of results should take place while patients are still being randomised. One exception was the ISIS-2 trial: a letter was published outlining the efficacy of streptokinase in reducing mortality in patients treated within four hours of onset of a myocardial infarction,[34] even though randomisation was allowed to continue at the discretion of each individual collaborator. Such a brief interim communication was in the spirit of collective ethics ("let's inform cardiologists about the value of streptokinase") but we should be wary of such a precedent leading to other trials releasing interim data in this way.

The decision to stop a trial has immediate consequences, whereas publication may take considerable time. For instance, the United States AIDS Clinical Trial Group decided in August 1989 to accept

a data monitoring committee recommendation to stop a placebo controlled trial of zidovudine for asymptomatic HIV infected patients with a CD4 count $<500\times10^6$ cells/l. The rate of progression to AIDS in the placebo group was more than double that with zidovudine (low dose and high dose) and, though the average follow up was short (12 months), it was thought unethical to continue given the magnitude and statistical significance of the observed differences. Investigators were immediately informed and a press release followed, so that the medical, patient, and general communities could all be informed of the trial's conclusions. One problem is the inevitable delay in publishing trial results in a peer reviewed medical journal—in this case eight months until April 1990[35]—during which everyone has to operate without knowing the full trial findings.

Certain principles seem pertinent here. Interim data are rarely clean and complete, so it is unwise to release any interim report for public consumption. Also the trial organisers have had no say in these confidential data monitoring committee reports, and their collective wisdom is vital to a quality final publication. An efficient trial organisation can reduce delay, but thereafter medical journals need to provide a "fast track" for major trials that stop early without compromising the need for peer review.

For any clinical trial there are usually other ongoing trials studying related therapeutic issues. Thus when one trial stops can the others ethically continue? The reactions of trialists can vary. One, keen not to affect his trial's continuation, declared: "Please don't rush to send me your results. Second class mail over Christmas, preferably." However, most trialists wish to keep abreast of others' findings, and one responsibility of any trial that stops is to inform others confidentially of their results. For instance, organisers of the Anglo-French Concorde trial of zidovudine versus placebo in asymptomatic HIV infection were informed of the AIDS Clinical Trial Group's results ahead of publication in order to make informed decisions on their own trial's future. In fact, they decided to continue, though allowing patients with two CD4 counts below 500×10^6 cells/l to take zidovudine if they wished. The Concorde trial has over 1700 patients and will undoubtedly be a major addition to knowledge on when to use zidovudine—that is, its continuation helps collective ethics considerably.

In HIV trials, especially in the United States, the push towards individual ethics at the expense of collective ethics has been detrimental to determining the most effective therapeutic policies. Decisions on which trials to start, when not to randomise, and when to stop a trial have all been adversely affected by the pressure for instant results and action. Though efficiency of research programmes can undoubtedly be improved, real gains in knowledge are at serious risk if we deviate from good clinical trial practice—for example, by stopping trials too soon.

Of concern is the manner of communicating trial findings after early stopping to both the medical profession and the general public. The United States physicians' health study[36 37] is an interesting case, in which 22 071 physicians took part in a factorial placebo controlled trial of both low dose aspirin and β carotene. The trial's aspirin component stopped because of (a) a surprisingly large and highly significant 47% reduction in myocardial infarction with aspirin, and (b) no effect of aspirin on the primary end point—cardiovascular death. This mix of both positive and negative findings raises problems concerning the role of aspirin in primary prevention, especially in view of the increasing risk of haemorrhagic stroke on aspirin and the very low death rate overall in this healthy sample of physicians. The relative reduction of 47% can be re-expressed less dramatically as an estimated absolute reduction of two infarcts for every 1000 physician years' taking aspirin.[38] The general release of conclusions on aspirin's benefits for myocardial infarction (not a predefined primary end point) led to clinical and public confusion on whether to prescribe low dose aspirin. Thus while public statements may be necessary when trials stop early, they should be cautious, without overextrapolation to other groups not studied, so that sound therapeutic practice is not destabilised with poorly documented new ideas.

When any trial stops early it is important to appreciate that the results are prone to exaggeration, because if results are on a "random high"—that is, due to chance they are above the true treatment effect—one is much more likely to stop early than if they are on a "random low." For instance, in the United States physicians' health study the first publication shows 104 versus 189 myocardial infarctions with aspirin and placebo, relative risk 0·53, while the additional data in the final publication show an extra 35 versus 50 myocardial infarctions—that is, relative risk 0·7. Thus if a trial has a surprisingly large treatment difference at interim analysis we can anticipate that estimates will shrink back from the "random high" as more data come in. However, if the trial stops early we may have to wait for evidence from other related trials (if allowed) before such shrinkage is observed.

Conclusion

This article has attempted to elucidate some of the ethical, statistical, and organisational issues that need to be considered in data monitoring and early stopping of clinical trials. The ethical dilemma faced by data monitoring committees has no easy solution, but it is important to note that premature stopping based on limited evidence can have severe consequences, either by introducing exaggerated claims so that inadequate treatments enter clinical practice or by failing to collect sufficient evidence on effective treatments in order to convince rightly sceptical clinicians of their true merits.

1 O'Rourke PP, Crone RK, Vacante JP, Ware JH, Lillikei CW, Parad RB, et al. Extracorporeal membrane oxygenation and conventional medical therapy in neonates with persistent pulmonary hypertension of the newborn: a prospective randomized study. Pediatrics 1989;84:957-63.
2 Ware JH. Investigating therapies of potentially great benefit: ECMO (with discussion). Statistical Science 1989;4:298-340.
3 ISIS-2 Collaborative Group. Randomized trial of intravenous streptokinase, oral aspirin, both, or neither among 17 187 cases of suspected acute myocardial infarction. Lancet 1988;ii:349-60.
4 Doll R. Development of controlled trials in preventive and therapeutic medicine. J Biosoc Sci 1991;23:365-78.
5 Yusuf S, Collins R, Peto R, et al. Intravenous and intracoronary fibrinolytic therapy in acute myocardial infarction: overview of results on mortality, reinfarction and side effects from 33 randomized controlled trials. Eur Heart J 1985;6:556-85.
6 GISSI. Long-term effects of intravenous thrombolysis in acute myocardial infarction: final report of the GISSI study. Lancet 1987;ii:871-4.
7 Freedman B. Equipoise and the ethics of clinical research. N Engl J Med 1987;317:141-5.
8 McPherson K. Statistics: the problem of examining accumulating data more than once. N Engl J Med 1974;290:501-2.
9 Pocock SJ. Clinical trials: a practical approach. Chichester: Wiley, 1983.
10 Armitage P. Sequential medical trials. Oxford: Blackwell, 1975.
11 Geller NL, Pocock SJ. Interim analyses in randomized clinical trials: ramifications and guidelines for practitioners. Biometrics 1987;43:213-23.
12 Fleming TR, Harrington DP, O'Brien PC. Designs for group sequential tests. Controlled Clin Trials 1984;5:348-61.
13 O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics 1979;35:549-56.
14 West of Scotland Coronary Prevention Study Group. A coronary primary prevention study of 6000 Scottish men aged 45-64 years: study design. J Clin Epidemiol (in press).
15 Dillman RO, Seagren SL, Propert K, Guerra J, Eaton WL, Perry MC, et al. A randomized trial of induction chemotherapy plus high-dose radiation versus radiation alone in stage III non-small cell lung cancer. N Engl J Med 1990;323:940-5.
16 Tannock IF, Boyes M. When is a cancer treatment worthwhile? N Engl J Med 1990;323:989-90.
17 Souhami RL, Spiro SG, Cullen M. Chemotherapy and radiation therapy as

compared with radiation therapy in stage III non-small cell lung cancer. *N Engl J Med* 1991;324:1136.

18 Propert KJ, Dillman RO, Seagren SL. Chemotherapy and radiation therapy as compared with radiation therapy in stage III non-small cell lung cancer. *N Engl J Med* 1991;324:1136-7.

19 Pocock SJ, Hughes MD. Practical problems in interim analyses, with particular regard to estimation. *Controlled Clin Trials* 1989;10:2209-15.

20 Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, *et al.* Design and analysis of randomised controlled trials requiring prolonged observation of each patient. 1. Introduction and design. *Br J Cancer* 1976;34:585-612.

21 Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika* 1983;70:659-63.

22 Cardiac Arrhythmia Suppression Trial Investigators. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *N Engl J Med* 1989;321:406-12.

23 Hallstrom A. Statistical interim monitoring of the cardiac arrhythmia suppression trial. *Stat Med* 1990;9:1081-90.

24 Whitehead J. *The design and analysis of sequential clinical trials.* Chichester: Ellis Horwood, 1983.

25 Freedman LS, Spiegelhalter DJ. Comparison of Bayesian with group sequential methods for monitoring clinical trials. *Controlled Clin Trials* 1989;10:357-67.

26 PACK Trial Group. Prevention of atherosclerotic complications: controlled trial of ketanserin. *BMJ* 1989;298:424-30.

27 Pocock SJ, Hughes MD. Estimation issues in clinical trials and overviews. *Stat Med* 1990;9:657-71.

28 Errington RD, Ashby D, Gore SM, Abrams KR, Myint S, Bonnett DE, *et al.* High energy neutron treatment for pelvic cancers: study stopped because of increased mortality. *BMJ* 1991;302:1045-51.

29 Fleming TR, Watelet LF. Approaches to monitoring clinical trials. *J Natl Cancer Inst* 1989;81:188-93.

30 Lan KK, Simon R, Halperin M. Stochastically curtailed sampling in long-term clinical trials. *Communications in Statistics—Theory and Methods* 1984;13:2339-53.

31 Henderson WG, Fisher SG, Weber L, Hammermeister KE, Sethi G. Conditional power for arbitrary survival curves to decide whether to extend a clinical trial. *Controlled Clin Trials* 1991;12:304-13.

32 Ware JH, Muller JE, Braunwald E. The futility index: an approach to the cost-effective termination of randomized clinical trials. *Am J Med* 1985;78:635-43.

33 Hampton JR, Julian DG. Role of the pharmaceutical industry in major clinical trials. *Lancet* 1987;ii:1258-9.

34 ISIS Steering Committee. Intravenous streptokinase given within 0-4 hours of onset of myocardial infarction reduced mortality in ISIS-2. *Lancet* 1987;i:502.

35 Volberding PA, Lagakos SW, Koch MA, Pettinelli C, Myers MW, Booth DK, *et al.* Zidovudine in asymptomatic HIV infection: a controlled trial in persons with fewer than 500 CD$_4$-positive cells per cubic millimeter. *N Engl J Med* 1990;322:941-9.

36 Steering Committee of the Physicians' Health Study Research Group. Preliminary report: findings from the aspirin component of the ongoing physicians' health study. *N Engl J Med* 1988;318:262-4.

37 Steering Committee of the Physicians' Health Study Research Group. Final report on the aspirin component of the ongoing physicians' health study. *N Engl J Med* 1989;321:129-35.

38 Pocock SJ, Thompson SG. Primary prevention trials in cardiovascular disease. *J Epidemiol Community Health* 1990;44:3-6.

(Accepted 29 May 1992)

Ocular complications observed in leprosy patients in Romania

T J ffytche, F Brandt, J Cerchez, S Stoicescu, P Stingl, M Filitis

St Thomas's Hospital, London SE1 7EH
T J ffytche, *consultant ophthalmic surgeon*

Munich, Germany
F Brandt, *ophthalmologist*

Tulcea, Romania
J Cerchez, *ophthalmic surgeon*
M Filitis, *dermatologist*

Eye Clinic Hospital, Bucharest, Romania
S Stoicescu, *ophthalmic surgeon*

Steingaden, Germany
P Stingl, *physician*

Correspondence to:
Mr ffytche.

The political changes that have occurred in Romania since the revolution of December 1989 have thrown to light many health care problems within the country, and the plight of the orphans, the children with AIDS, and the elderly and the mentally subnormal populations have received widespread coverage by the world press.

In the early part of 1990 reports of the existence of a leprosarium in the eastern part of the country emerged, amid considerable publicity and uninformed opinion. Leprosy patients who are ethnically European are likely to have a high prevalence of multibacillary disease and are therefore at risk from ocular complications.[1]

To assess the needs of these patients three visits to the leprosarium were arranged in 1990-1, through the help of the Romanian Ministry of Health.

Tichilesti Leprosarium

Tichilesti is a small isolated settlement lying on the Romanian border with Russia, close to the mouth of the Danube and about 40 km from Tulcea, the nearest large town. The leprosarium was established at the beginning of this century, when there were about 200 patients. The number has now diminished to 53, most of whom have been there for many years, the last new admission being in 1986. Tichilesti is thus one of several settlements of purely European patients with leprosy; others include centres in southern Russia, Spain, Portugal, Greece, and Turkey, although most of these contain an appreciable proportion of non-Europeans.

The conditions in the leprosarium are austere, although not unduly so, and sanitation is poor, but the patients seem generally content with their surroundings (fig 1). The health care is administered by a resident doctor, two nurses, and an administrator. Minor medical problems are dealt with on site, but patients requiring specialist attention or surgery are transferred to the main hospital in Tulcea or even to Bucharest.

Patients and methods

Out of a possible 53 patients 46 (87%) were examined; one man was thought not to have leprosy and was excluded from the study. All of the remaining 45 were of European extraction; 22 were women and 23 men. One patient was aged 39, 12 were aged between 40 and 59, and the remainder (32) were over 60. Most patients were believed to have had leprosy since young adulthood, and although one patient had had leprosy diagnosed fewer than 10 years previously, 84% had had the disease for over 30 years.

Medical records were difficult to assess and smear testing had been infrequent, but clinical examination suggested that 91% of patients had multibacillary disease, the remaining patients were thought to have paucibacillary disease. Two patients had positive smear test results, and although the regularity of administration of treatments was uncertain, almost all patients were taking or had taken sulphones, either as dapsone or glucosulphone, some had been given rifampicin, and several were also using chaulmoogra oil.

Ophthalmic examination—Eye examinations were carried out according to guidelines set for collecting data on ocular complications of leprosy,[2] and the

FIG 1—*Patients waiting for eye examination*