Saturday 18 April 1998

BMJ

Routine episiotomy in developing countries

Time to change a harmful practice

More women in developing countries are delivering their babies in hospitals. In Latin America institutional births account for 70% of all deliveries; in Africa, 36%; and in developing countries overall some 40%.[1] What is becoming apparent is that in some countries virtually all the women delivering in hospital will be surgically cut. If they miss out on a caesarean section they will have an episiotomy. For example, Brazil has caesarean section rates of greater than 30%, and Argentina has episiotomy rates of greater than 80% for vaginal births.[2 3] Questions about high caesarean sections rates have been raised in the past, but unnecessary episiotomies have not been widely debated.

Obstetricians in the tropics continue to instruct health staff to apply a policy of "avoid tears-do episiotomies" routinely. They may be acting in good faith, but the evidence shows that they are wrong.[4] Aiming surgically to cut all women delivering vaginally has no demonstrable benefit for the infant or mother but causes the woman unnecessary pain and adverse psychological effects and may cause death.[5]

In England episiotomies were performed on over half of all women delivering in 1980, falling to 37% in 1985. Recently released figures for 1994-5 indicate a further dramatic fall to about 20%.[6 7] Although the older figures may not be strictly comparable with those from 1994-5, the overall trend downwards is clear, and local data support this. For example, in Liverpool Women's Hospital, in the first half of 1997 episiotomies were performed in 16% of all deliveries and 5% of normal births (J Neilson, personal communication).

Is this the trend in the world's poorer countries? We conducted a straw poll of 10 midwives from Zambia, Malawi, Nigeria, Ghana, Kenya, and Nepal attending courses in Liverpool. Our respondents had not considered whether policies of routine episiotomy could do more harm than good and found the review by Carroli et al enlightening.[4] Most indicated that health professionals performed episiotomies routinely on primigravidas to prevent third degree perineal tears. Some midwives reported that some were performed to allow midwifery and medical students the opportunity to practise the procedure.

We sought to document this anecdotal evidence of high episiotomy rates in developing countries, but data are sparse. A systematic search of Medline and contact with the Royal College of Midwives revealed very little quantitative data. We found a study in Botswana, where 1 in 3 mothers having a normal delivery had an episiotomy.[8] Another study in Burkina Faso showed that, in primary care facilities, 43% of primigravidas received episiotomies—in a health system that frequently ran out of sutures and antibiotics.[9] What is particularly worrying is that when health care resources are short episiotomy is more likely to result in complications. This increases the harm done by the procedure, in people who are least able to cope with the increased pain and suffering and least able to afford the prolonged hospitalisation.

The World Health Organisation has taken a clear stand against routine episiotomy, in line with the best available evidence.[10] Convincing obstetricians may be more problematical. Yet this is an important ethical issue for doctors and patients alike. In the West the procedure is usually discussed with women at antenatal clinics. In our experience in developing countries this does not happen. When the procedure is routine it therefore becomes a premeditated surgical procedure carried out without consent from the woman.

It is important that we rapidly compare episiotomy rates between facilities and countries. Such data will guide more informed discussion about the level of unnecessary interventions. It will then be obvious to obstetricians, midwives, and the public whether obstetric practice is based on doing what is best for women, or persisting with policies that do more harm than good.

A Maduma-Butshe *Maternal and child manager*

Organisation for Rural Programmes for Women, Khumalo, Bulawayo, Zimbabwe

Adele Dyall *Formerly senior nursing officer*

Keewatin Regional Health Board, Government of the Northwest Territories, Canada

Paul Garner *Head*

International Health Division, Liverpool School of Tropical Medicine, Liverpool L3 5QA

1 Family and Reproductive Health Division. *Coverage of maternity care: a listing of available information.* 4th ed. Geneva: World Health Organisation, 1996.
2 Barros FC, Vaughan JP, Victora CG, Huttly SRA. Epidemic of caesarean sections in Brazil. *Lancet* 1991;338:167-9.
3 Argentine Episiotomy Trial Collaboration Group. Routine *v* selective episiotomy: a randomised control trial. *Lancet* 1993;342:1517-8.
4 Carroli G, Belizan J, Stamp G. Episiotomy policies in vaginal births. In: *Pregnancy and childbirth module of the Cochrane database systematic reviews* [updated 04 March 1997]. The Cochrane Library. Oxford: Update Software, 1997.
5 Sleep J, Roberts J, Chalmers I. Care during the second stage of labour. In: Chalmers I, Enkin MW, Keirse MJNC, eds. *Effective care in pregnancy and childbirth.* Oxford: Oxford University Press, 1989:1136-41.
6 Graham ID. *Episiotomy: challenging obstetric interventions.* Oxford: Blackwell Science, 1997.

BMJ 1998;316:1179–80

7 Department of Health. *NHS maternity statistics, England: 1989-90 to 1994-5*. London: Department of Health, 1997.
8 Mamba F. The liberal use of episiotomy in normal delivery: do women in Botswana accept the practice? Guildford: University of Surrey 1994 (Masters thesis).
9 Lorenz N, Nougtara A, Garner P. Episiotomies in Burkina Faso. *Tropical Doctor* (in press).
10 Thompson A. Episiotomies should not be routine. *Safe Motherhood Newsletter* Geneva: World Health Organisation, 1997:12.

Long term pharmacotherapy of depression

Can reduce relapses and recurrences in major depression

The high rates of persistent morbidity, recurrence, and death among patients treated for depression [1–2] have highlighted the importance of long term psychological and pharmacological treatment. What evidence do we have that long term drug treatment is effective, and how should we choose between individual drugs?

Long term drug treatment comprises continuation and maintenance treatment. Continuation treatment helps to consolidate recovery from depressive episodes and prevent relapses (worsening of continuing or recently treated episodes). Maintenance or prophylactic treatment helps to prevent occurrence of new episodes and is usually recommended for patients who have had at least three depressive episodes in the preceding five years. The distinction between continuation and maintenance treatment is somewhat arbitrary and may not reflect underlying biological processes, but most researchers agree that four to six months' remission should occur before a recurrence is diagnosed.

Differences in methods between the many trials of long term pharmacological treatment[3 4] make interpretation difficult, but about 60% of patients who respond to an antidepressant and are then given a placebo remain in remission for up to two years. If instead of receiving a placebo they continue on the drug they have a 20-25% better chance of maintaining their improvement[4]—that is, twice as many relapses occur on placebo as on antidepressants (about 40% *v* 20%). Drugs also increase the time to onset of relapse or recurrence and reduce the severity of residual depressive symptoms in those who do not relapse. Nevertheless, studies have thrown little light on which patients benefit most from long term treatment, the comparative effects of different psychological and pharmacological treatments, and the optimum length of treatment. Moreover, few trials have continued for longer than two years.

The benefit of long term drug treatment has been clearly shown only in outpatients with major depression. We cannot assume that the same benefits will be achieved in the milder, heterogeneous cases of "depression" encountered in general practice. We should not prescribe long term for people with infrequent, short bouts of mild depression and those whose low mood reflects changing social circumstances unless there is convincing evidence in individual patients that they have benefited from such treatment. We should also hesitate to prescribe long term for patients whose depression is an episodic symptom of personality disorder, an effect of alcohol or drug abuse, or a phenomenon perpetuated by the desire to remain in the sick role.

Few long term comparative studies of the efficacy of different drugs have been carried out, and meta-analyses of the results of short term trials have failed to show important differences between different types of antidepressants.[5] Factors such as tolerability, unwanted effects, toxicity in overdose, and cost must therefore determine which drug to use.

A common index of tolerability is the discontinuation rate in clinical trials. Despite claims that newer antidepressants are better tolerated than older tricyclics, such as amitriptyline and imipramine, only 1-5% fewer patients receiving selective serotonin reuptake inhibitors than receiving tricyclics drop out from trials—figures of doubtful practical importance.[5] The difference between the drugs is largely attributed to fewer dropouts due to side effects, although it is often difficult to know why patients stop their treatment.[6] Furthermore, discontinuation rates from trials may not accurately represent routine clinical practice or long term treatment (when adaptation to unwanted effects may occur). Meta-analyses of efficacy and discontinuation rates have suggested that significant differences do exist between individual drugs, although the methods of these meta-analyses have been criticised.[5]

Death is more likely to result from overdoses of older tricyclic drugs than newer compounds.[7] However, only about 4% of all suicides are due to overdoses of single antidepressants, and it is not known what proportion of these overdoses are taken during treatment (when drug choice is relevant). Furthermore, a higher suicide rate among patients taking tricyclic drugs could be accounted for by doctors prescribing these drugs more often for patients prone to suicide,[8] and the overall suicide rate (by any method) among patients treated with new and old antidepressants is similar.[9]

The average net ingredient cost of an NHS prescription for a selective serotonin reuptake inhibitor in 1995 was £27.21 compared with £0.77 for amitriptyline. If all patients were prescribed serotonin reuptake inhibitors the annual cost (at 1995 prices and consumption rates) would be £350m more than if they were all prescribed amitriptyline.[6] The long term benefits purchased from this are slightly lower discontinuation rates,[5] possibly with fewer relapses and recurrences, and fewer deaths from overdose.[7] Conversely, the additional cost means there is less money available for other purchases—for example, four million psychiatric outpatient attendances or almost 22 million hours of community psychiatric nurse time.[4]

The results of cost effectiveness and cost benefit assessments depend on the model used. Recent overviews do not recommend expensive newer anti-

BMJ 1998;316:1180–1

depressants as first line treatment,[4–6 10 11] but these drugs should be prescribed both short and long term for patients who cannot tolerate older antidepressants and/or have a high risk of suicide by overdose. Because the newer antidepressants have less sedative and autonomic effects, they should also be given to patients with depressive disorders who are prone to accidents [12] or have cardiovascular disease.[4 6]

J Guy Edwards *Honorary clinical senior lecturer*

University of Southampton, Faculty of Medicine, Health and Biological Sciences, Department of Psychiatry, Royal South Hants Hospital, Southampton S014 OYG

Conflict of interest: None.

1 Kessing LV, Andersen PK, Mortensen PB, Bolwig TG. Recurrence in affective disorder. I. Case register study. *Br J Psychiatry* 1998;172:23-8.
2 Harris EC, Barraclough B. Suicide as an outcome for mental disorders. A meta-analysis. *Br J Psychiatry* 1997;170:205-28.
3 Montgomery SA. Long-term treatment of depression. *Br J Psychiatry* 1994;165 (suppl 26):31-6.
4 Edwards JG. Prevention of relapse and recurrence of depression: newer versus older antidepressants. *Adv Psych Treatment* 1997;3:52-7.
5 Anderson I. Lessons to be learnt from meta-analyses of newer versus older antidepressants. *Adv Psych Treatment* 1997;3:58-63.
6 Edwards JG. Drug choice in depression. Selective serotonin reuptake inhibitors or tricyclic antidepressants? *CNS Drugs* 1995;4:141-59.
7 Henry JA. Toxicity of newer versus older antidepressants. *Adv Psych Treatment* 1997;3:41-5.
8 Isacsson G, Redfors I, Wasserman D, Bergman U. Choice of antidepressants: questionnaire survey of psychiatrists and general practitioners in two areas of Sweden. *BMJ* 1994;309:1546-9.
9 Jick SS, Dean AD, Jick H. Antidepressants and suicide. *BMJ* 1995;310:215-8.
10 Hotopf M, Hardy R, Lewis G. Discontinuation rates of SSRIs and tricyclic antidepressants : a meta-analysis and investigation of heterogeneity. *BMJ* 1997;170:120-7.
11 Woods SW, Rizzo JA. Cost-effectiveness of antidepressant treatment reassessed. *Br J Psychiatry* 1997;170:257-63.
12 Edwards JG. Depression, antidepressants, and accidents. *BMJ* 1995; 311:887-8.

Tamoxifen for the prevention of breast cancer

Important questions remain unanswered, and existing trials should continue

News p 1187

The complexity of the effects of tamoxifen has recently been summarised: "Tamoxifen is an antioestrogen with complex pharmacology encompassing variable species-, tissue-, cell-, gene-, age- and duration of administration-specific effects, from oestrogen-like agonist actions to complete blockade of oestrogen action. This complexity is consistent with the various, and sometimes paradoxical, effects that have been associated with tamoxifen administration in animals and humans."[1] The report concluded that there was sufficient evidence that tamoxifen increased the risk of endometrial cancer and conclusive evidence that it reduced the risk of contralateral breast cancer.

Because of this complexity of effects, a consensus has existed until recently that using tamoxifen for preventing breast cancer was questionable even in clinical trials, and that only trials in women at high risk of breast cancer could be justified.[2] This month, however, investigators from the National Cancer Institute-National Surgical Adjuvant Breast and Bowel Project decided to release the initial results of the Breast Cancer Prevention Trial, based on their assessment that a reduction in breast cancer had been shown in the group receiving tamoxifen (p 1187).[3] Should we now conclude that the evidence is sufficient to warrant the widespread use of tamoxifen for preventing breast cancer in high risk women, or do the problems that precluded the preventive use of tamoxifen outside clinical trials[4] remain unanswered?

The most fundamental question is whether the reduction in incidence will translate into a comparable reduction in mortality. In fact, it would have been surprising not to see any preventive effect of tamoxifen, and the reduction in incidence observed in the trial (45%) is of the size expected from the reduction in the incidence of contralateral cancers in trials of tamoxifen used as adjuvant therapy in women operated on for breast cancer.[5] There are, however, two reasons to suspect that the prognosis of cancers diagnosed in women taking tamoxifen might be worse than that in the general population. Firstly, the biology of these cancers is likely to be different, owing to the selection of tamoxifen resistant cancers. Secondly, adjuvant tamoxifen is associated with a 25% reduction in mortality, but we don't know its efficacy in women who were already taking tamoxifen when their cancer was diagnosed. If adjuvant tamoxifen is less effective in these women, or if only well differentiated cancers are prevented by tamoxifen, the reduction in mortality from breast cancer might be less (possibly much less) than the reduction in incidence, with important bearings on the balance between risks and possible benefits of preventive tamoxifen.

A second problem is the duration of tamoxifen treatment. In the B-14 adjuvant trial of the National Surgical Adjuvant Breast and Bowel Project almost twice as many distant recurrences and deaths were observed in patients who were randomly assigned to take tamoxifen for 10 years as in those who stopped after 5 years.[6] Obviously, these findings relate to the growth of micrometastatic disease and cannot be applied directly to the preventive effect of tamoxifen, but they cast doubts on the long term effects of tamoxifen on the incidence of breast cancer.

The long term toxicity of tamoxifen is also a matter of concern: available data after 5, 10, and15 years of follow up confirm the increase in the incidence of endometrial cancer and of thromboembolic complications and provide some suggestion of ocular toxicity, but these effects are not common and might be more than balanced by the reduced risk of coronary heart disease and osteoporosis. However, with few exceptions, these data are derived from studies using short term treatments (2-5 years), and little is known about the effects of continuing tamoxifen beyond 10 years. Another issue that we know little about is the interaction of tamoxifen with other treatments, including hormone replacement therapy, given concurrently or in sequence.

BMJ 1998;316:1181–2

The point is that if we plan to put healthy women in their 60s, 50s, or even earlier (40% of the women in the Breast Cancer Prevention Trial were aged 35-49) on tamoxifen, we should try to figure out what to do 10 years later, when, hopefully, most of these women will still be healthy but still at risk of breast cancer (at higher risk, indeed, because of age). So far, women in the Breast Cancer Prevention Trial have been followed on average for four years, and, owing to the future likely contamination of the control group, we cannot expect this trial to contribute significantly to resolving this problem.

As a consequence, we need to continue the other two trials (one in Italy and one in Britain) that are currently under way. Their participants should be clearly informed that the uncertainty about the balance between the possible risks and benefits of taking tamoxifen for preventing breast cancer is not substantially modified by the early results of the Breast Cancer Prevention Trial. For the same reason, prescription of tamoxifen for preventing breast cancer outside clinical trials should continue to be discouraged.

A final problem is the definition of the level of risk of breast cancer that could justify, in the future, the start of a preventive regimen associated with costs, side effects, and long term risks. The eligibility criteria used in the Breast Cancer Prevention Trial are very inclusive, since all women aged 60 or older, and younger women with an equivalent (or greater) risk were eligible to participate. On the other hand, we don't know if tamoxifen is equally effective in preventing breast cancer in all risk groups. For instance, some evidence suggests that some women at very high risk, such as the carriers of a germline BRCA mutation in high risk families, are predisposed to develop hormone independent tumors, which are less likely to be affected by the preventive action of tamoxifen.[7] This issue might be addressed in the future, when some of the participants in the trials will be tested for germline mutations and by appropriate subgroup analyses, but at present the use of tamoxifen for prevention is not justified even in these women.

Paolo Bruzzi *Head*

Clinical Epidemiology and Trials Unit, National Institute for Cancer Research, 10 16132 Genoa, Italy

1 IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. *IARC Monographs on the evaluation of carcinogenic risks to humans*. Vol. 66. *Some pharmaceutical drugs*. Lyons: International Agency for Research on Cancer, 1996:253-366.
2 Bush TL, Helzlsouer KJ. Tamoxifen for the primary prevention of breast cancer: a review and critique of the concept and trial. *Epidemiol Rev* 1993;15:233-43.
3 Josefson D. Breast cancer trial stopped early. *BMJ* 1998;316:1187.
4 *Prevention of breast cancer*. PDQ Supportive Care/Screening/Prevention Information. Cancer Net, National Cancer Institute, April 1998 http//cancernet.nih.nci.gov/pdq.htm
5 Early Breast Cancer Trialists' Collaborative Group. Systemic treatment of early breast cancer by hormonal, cytotoxic and immune therapy. 133 randomized trials involving 31 000 recurrences and 25 000 deaths among 75 000 women. *Lancet* 1992;339:1-15, 71-85.
6 Fisher B, Dignam J, Bryant J, DeCillis A, Lawrence Wickerham D, et al. Five versus more than five years of tamoxifen therapy for breast cancer patients with negative lymph nodes and estrogen receptor-positive tumors. *J Natl Cancer Inst* 1996;88:1529-42.
7 Karp SE, Tonin PN, Begin LR, Martinez JJ, Zhang JC, Pollack MN, et al. Influence of BRCA1 mutations on nuclear grade and estrogen receptor status of breast carcinoma in Ashkenazi Jewish women. *Cancer* 1997;80:435-44.

Selegiline, or the problem of early termination of clinical trials

The clinical questions are not well answered, and probably never will be

Parkinson's disease is the second most common neurodegenerative disorder, after dementia. About 1.4% of people aged 55 years or over have Parkinson's disease,[1] and because of the aging of Western populations the absolute number of patients is rapidly increasing. Until now, treatment has been mainly symptomatic, but much effort is being put into developing neuroprotective agents that may stop progression or even cure the disease. Clearly, unrecognised adverse effects of such treatments may potentially affect large numbers of patients and any suggestion of such effects needs thorough investigation.

Selegiline has probably become the most controversial drug in Parkinson's disease during the past decade. Its presumed efficacy was initially ascribed to neuroprotection due to inhibition of monoamine oxidase-B, then to a symptomatic effect, and more recently again to neuroprotection, this time due to inhibition of apoptosis. The greatest controversy, however, occurred because selegiline caused the early termination of the intervention arms of two large multicentre studies—for completely different reasons. In the DATATOP study of the US Parkinson Study Group subjects randomised to receive selegiline did better than those randomised to placebo or tocopherol in that they reached the endpoint (start of levodopa treatment) significantly later.[2] In contrast, in the trial of the Parkinson's Disease Research Group in the United Kingdom mortality among patients receiving selegiline in addition to levodopa (plus decarboxylase inhibitor) was higher than in those taking only levodopa (plus decarboxylase inhibitor).[3]

This latter finding was totally unexpected and generated much debate,[4 5] but it did cause sales of selegiline to drop considerably.[6] In this week's *BMJ* Ben-Shlomo et al provide supplementary information from the UK trial, including further follow up data and more detail about patient characteristics and causes of death, in an attempt to explain their previous findings (p 1191).[7]

The selegiline plus levodopa arm of the UK trial was terminated at the end of September 1995 as a result of an increased risk estimate based on deaths till the end of 1993. Consequently, all end of trial analyses are biased. The more valid and unbiased estimate of

We commissioned two editorials on the paper by Ben-Shlomo et al on p 1191. The first, by Monique Breteler, discusses the clinical implications. The second, by Keith Abrams, discusses the difficulties of interpreting interim analyses of trials

BMJ 1998;316:1182–3

the risk of mortality based on the five year follow up shows a hazard ratio for these patients compared with those receiving only levodopa of 1.38 (95% confidence interval 0.95 to 2.04). Another unbiased estimate comes from the subjects who were re-randomised from the bromocriptine arm (hazard ratio 1.54; 0.83 to 2.87). Although this is a highly selected group, the authors suggest, plausibly, that because of the randomisation these data can be viewed as if they came from an independent trial.

The difference in mortality seemed to be highest in the third and fourth year of follow up, after which it diminished. This finding was previously criticised because of the lack of a biological explanation.[5] The authors dismiss the possibility that had the combined treatment arm of the trial continued, the mortality difference might have diminished even further. Their dismissal may not, however, be justified, for at least two reasons. Firstly, notwithstanding the similar findings in the bromocriptine group, the data are compatible with a "randomly high" increased mortality at 24 to 48 months. This holds in particular for the "as treated" analysis, which seems the more informative here. Secondly, one can question the implicit assumption that had the mortality ratio further decreased this would have meant that the findings of increased mortality were due to chance. Hardly any effective therapy is entirely harmless. Even if selegiline in combination with levodopa reduces mortality in the long run, the net effect on mortality might be unfavourable in situations where there is little to gain—for example, at the beginning of the disease.

The cause specific mortality rates do not suggest a specific cause for the excess deaths. An excess of deaths from Parkinson's disease was reported, but this probably reflects lack of further information since in almost half the cases where the panel could not reach a diagnosis the diagnosis on the death certificate was Parkinson's disease. Also the comparison of clinical characteristics in those who died between those who took only levodopa and those on combination treatment yielded no clear clues to explain the extra deaths in the combined group as all differences were far from significant.

What therefore can we conclude about the hazards of combination treatment with levodopa and selegiline? Firstly, the UK trial finds no significant increased mortality due to combined treatment with levodopa and selegiline. Secondly, because of the interim analyses, doubts have been raised, where previously we had no indication that selegiline had any detrimental effect whatsoever. Any new evidence incriminating selegiline will weigh heavily. The possibility that the net effect of positive and adverse effects of combination therapy with selegiline depends on background risk or severity of disease should be considered. Thirdly, the recommendation to avoid combined treatment among patients with more advanced Parkinson's disease and postural hypotension, frequent falls, confusion, and dementia is not based on unequivocal results from the trial but rather on clinicians' beliefs.

What does this mean for the treatment of patients with Parkinson's disease? Unfortunately, the early termination of the arms involving selegiline in both the US and the UK trials has limited the evidence on the long term effects of selegiline alone and in combination with levodopa and diminished the possibility that this can ever be validly obtained. Appreciation of the effect of combination treatment on morbidity is almost impossible in the UK trial, because only the results of the intention to treat analyses have been reported, whereas more than half of the study population had withdrawn from their original randomised treatment. On the basis of the limited evidence available, a cautious recommendation seems to be not to start combination treatment with selegiline and levodopa in patients with newly diagnosed Parkinson's disease. At the moment, however, there is little evidence to advise people who have been using both drugs for years and seem to be doing fine to change their treatment.

Monique M B Breteler *Assistant professor of neuroepidemiology*

Department of Epidemiology and Biostatistics, Erasmus University Medical School, 3000 DR Rotterdam, the Netherlands

1 de Rijk MC, Breteler MMB, Graveland GA, Ott A, Grobbee DE, van der Meché FGA, et al. Prevalence of Parkinson's disease in the elderly: the Rotterdam Study. *Neurology* 1995;45:2143-6.
2 Parkinson Study Group. Effect of deprenyl on the progression of disability in early Parkinson's disease. *N Engl J Med* 1989;321:1364-71.
3 Lees AJ, on behalf of the Parkinson's Disease Research Group of the United Kingdom. Comparison of therapeutic effects and mortality data of levodopa and levodopa combined with selegiline in patients with early, mild Parkinson's disease. *BMJ* 1995;311:1602-7.
4 Lees AJ, on behalf of the Parkinson's Disease Research Group of the United Kingdom. Effect of adding selegiline to levodopa in early, mild Parkinson's disease. *BMJ* 1996;312:702-5.
5 Olanow CW, Fahn S, Langston JW, Godbold J. Selegiline and mortality in Parkinson's disease. *Ann Neurol* 1996;40:841-5.
6 Silva MT, Schapira AHV, Jenner P. Unexpected findings of study of selegiline have not been treated with caution its authors advised. *BMJ* 1997;315:370.
7 Ben-Shlomo Y, Churchyard A, Head J, Hurwitz B, Overstall P, Ockelford J, et al. Investigation by Parkinson's Disease Research Group of United Kingdom into excess mortality seen with combined levodopa and selegiline treatment in patients with early, mild Parkinson's disease: further results of randomised trial and confidential inquiry. *BMJ* 1998;316:1191-6.

Monitoring randomised controlled trials

Parkinson's disease trial illustrates the dangers of stopping early

Papers p 1191

The trial of the Parkinson's Disease Research Group reported by Ben-Shlomo et al in this issue (p 1191), which updates the results of a previously curtailed randomised controlled trial,[1] raises several methodological issues. The current results relate to an initial three arm trial in which 782 patients with early stage Parkinson's disease were randomised to treatment with either levodopa alone (arm 1), levodopa and selegiline in combination (arm 2), or bromocriptine (arm 3). The first trial report, based on follow up to December 1991, showed no significant differences at the 5% level between arms 1 and

BMJ 1998;316:1183–4

2 in disability levels, but both arms showed significant improvements over baseline.[2] At this stage there were too few deaths to assess differences in mortality. The second report was based on follow up to December 1993, resulting in an average 5.6 years follow up.[3] As with the earlier report, there continued to be no significant differences between arms 1 and 2 in terms of disability levels. However, based on 44 deaths in 249 patients in arm 1 and 76 in 271 patients in arm 2, a significant difference in all cause mortality was observed, yielding a hazard ratio of 1.57 (95% confidence interval 1.09 to 2.30) and a P value of 0.015. At this point the trial was terminated, and patients in arm 2 were advised to switch to levodopa alone, but follow up continued. The current analysis, based on more complete follow up to September 1995, when the trial was terminated, shows that on the basis of 73 deaths in arm 1 and 103 in arm 2, the difference in all cause mortality is no longer significant at the 5% level, yielding a hazard ratio of 1.32 (0.98 to 1.79) on an intention to treat analysis.

These results raise several questions, paramount being whether the trial should have been stopped when it was. The essential question is whether the eventual result could have shown a clinical benefit in favour of combination therapy if the trial had been continued. The original sample size calculation assumed a 30% reduction in all cause mortality in favour of combination therapy at 10 years. Such a reduction equates to a hazard ratio of 0.74. The figure of 30% was based on an earlier uncontrolled, retrospective survey,[4] and it could be argued that such a minimum clinically worthwhile difference is over-optimistic.

Several approaches for estimating the minimum clinically worthwhile difference exist, including the use of elicitation techniques.[5 6] These techniques require that the beliefs and demands of clinicians—both those taking part in the trial and those not—about a possible treatment effect are elicited; they can serve two functions. The first is to ensure that the trial is ethical in terms of equipoise—that is, given trial participants' beliefs and demands, genuine uncertainty exists about the optimal treatment.[7] The second is that the demands may be used to design the trial and monitor it, as they represent the treatment difference required for clinical practice to change. If such an exercise had been conducted before the Parkinson's disease trial, would the average reduction in mortality demanded have been as high as 30%?

Obviously such an exercise cannot be conducted retrospectively, but it may be instructive to explore the implications of a demand less than 30%. When the trial was stopped the lower 95% confidence limit was 1.09, considerably greater than 0.74, and as such the possibility that continuing the trial would result in a 95% confidence interval that contained 0.74 would be unlikely. However, had a more modest 10% reduction been used, the corresponding hazard ratio would have been 0.90, and the argument for stopping would not be quite so persuasive.

Though such an exercise can be enlightening, it does not provide a quantitative summary of the plausibility of observing a clinically worthwhile difference if the trial continues. The Parkinson's disease trial did not use a formal stopping guideline to adjust the significance levels for the fact that interim analyses were being performed. Various classical methods have been advocated for such adjustments.[8] For example, the trial anticipated 10 annual interim analyses, so a simple adjustment would have been to use 0.01 rather than 0.05 at each analysis so that the overall significance level did not exceed 0.05.

An alternative strategy for assessing the level of evidence at each interim analyses would have been to adopt a Bayesian approach.[9] This approach enables the accumulating information obtained at successive interim analyses to be summarised and probability statements to be made about future results. Information can be summarised by a credibility interval, in which a quantity of interest lies within a specified probability.[10] The Parkinson's disease trial was designed to follow up patients over 10 years, during which about 260 deaths could be assumed to occur. Thus at the interim analysis to December 1993,[2] when 120 deaths had been observed, it could be predicted that the 95% credibility interval at the end of the trial would be 0.95 to 2.58. While this interval contains neither 0.74 nor 0.90, it does contain unity—that is, it suggests no treatment difference. Analogously, if a similar analysis was performed on the more complete follow up to September 1995 then the corresponding 95% credibility interval would be 0.78 to 2.25, indicating a wide range of plausible outcomes.

Though only statistical aspects of the monitoring of the UK Parkinson's disease trial have been considered, any decision to stop a trial is a complex one. The decision should not rely solely on statistical arguments, Bayesian or otherwise, but must be placed within a wider context, for example, by taking into account the balance between individual and collective ethics. This should be done by an independent data monitoring committee, who can assess all the available evidence relating to a trial, both internal and external.[6 8]

Keith R Abrams *Senior lecturer in medical statistics*

Department of Epidemiology and Public Health, University of Leicester, Leicster LE1 6TP

1 Ben-Shlomo Y, Churchyard A, Head J, Hurwitz B, Overstall P, Ockelford J, et al. Investigation by Parkinson's Disease Research Group of the United Kingdom into excess mortality seen with combined levodopa and selegiline in patients with early, mild Parkinson's disease. *BMJ* 1998;316:1191-6.
2 Parkinson's Disease Research Group in the United Kingdom. Comparisons of therapeutic effects of levodopa, levodopa and selegiline, and bromocriptine in patients with early, mild Parkinson's disease: three year interim report. *BMJ* 1993;307:469-72.
3 Lees AJ, on behalf of the Parkinson's Disease Research Group of the United Kingdom. Comparison of therapeutic effects and mortality data of levodopa and levodopa combined with selegiline in patients with early, mild Parkinson's disease. *BMJ* 1995;311:1602-7.
4 Birkmayer W, Knoll J, Riederer P, Youdim MBH, Hars V, Marton J. Increased life expectancy resulting from addition of l-deprenyl to Madopar treatment in Parkinson's disease: a long term study. *J Neural Transm* 1985;64:113-27.
5 Freedman LS, Spiegelhalter DJ. The assessment of subjective opinion and its use in relation to stopping rules for clinical trials. *Statistician* 1983;32:153-60.
6 Abrams KR, Jones DR. Bayesian interim analysis of randomised trials. *BMJ* 1997;314:1911-2.
7 Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med* 1987;317:141-5.
8 Pocock SJ. When to stop a trial. *BMJ* 1992;305:235-40.
9 Fayers PM, Ashby D, Parmar MKB. Tutorial in biostatistics: Bayesian data monitoring in clinical trials. *Stat Med* 1997;16:1413-30.
10 Lilford RJ, Braunholtz D. The statistical basis of public policy: a paradigm shift is overdue. *BMJ* 1996;313:603-7.