Putting trials on trial—the costs and consequences of small trials in depression: a systematic review of methodology

Matthew Hotopf, Glyn Lewis, Charles Normand

Abstract

Study objective—To determine why, despite 122 randomised controlled trials, there is no consensus about whether the selective serotonin reuptake inhibitors or tricyclic and related antidepressants should be used as first line treatment of depression.

Design—Systematic review of all RCTs comparing selective serotonin reuptake inhibitors and tricyclic or heterocyclic antidepressants.

Main results—The shortcomings identified in the 122 trials were as follows: (1) there was inadequate description of randomisation, (2) the outcomes used were mainly observer rated measurements of depression, and studies failed to use quality of life measures or perform economic evaluations, (3) doses of tricyclic antidepressants were inadequate, (4) generalisability of studies was poor (including a reliance on secondary care settings and inadequate follow up), and (5) there were statistical shortcomings such as low statistical power, failure to use intention to treat analyses, and the tendency to make multiple comparisons.

Conclusions—Future RCTs should be designed to inform policy makers and address these methodological shortcomings.

(*J Epidemiol Community Health* 1997;51:354-358)

Department of Psychological Medicine, Institute of Psychiatry, 103 Denmark Hill, London SE5 8AZ
M Hotopf

Epidemiology Unit, London School of Hygiene and Tropical Medicine and Section of Epidemiology and General Practice, London
G Lewis

Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, London
C Normand

Correspondence to:
Dr M Hotopf.

Accepted for publication February 1997

New treatments in medicine are often expensive. In the treatment of common chronic diseases the licensing of a new treatment often creates difficult choices for policy makers and clinicians. How widely available should the new treatment be? This question is especially difficult to answer when the new treatment has only modest advantages over placebo or the existing treatment, or is extremely expensive. Recent examples of this dilemma include interferon beta-1B in multiple sclerosis,[12] and lipid lowering agents in primary prevention of coronary heart disease.[3] This systematic review examines one new treatment in psychiatry—the selective serotonin reuptake inhibitors (SSRIs). It tries to discover why, despite intensive research, doctors are uncertain whether these or tricyclics should be used as first line treatment in depression.

The debate over the use of SSRIs as first line treatment of depression remains unresolved, despite at least six meta-analyses.[4-9] SSRIs are considerably more expensive than tricyclics,[10] and it is estimated that the NHS bill for antidepressants would rise from £88 m to £250 m if they were prescribed first line (1993 costs).[3] Those recommending wider prescribing of SSRIs claim that the costs would be offset by the advantages of these drugs. The debate in the main has concentrated on differences in how well the two classes of drug are tolerated rather than on efficacy per se. There is some evidence that the SSRIs are better tolerated and this is reflected by slightly lower attrition rates in clinical trials,[6] and possibly better compliance in clinical practice. In addition SSRIs are safer in overdose. Policy makers need to know the cost effectiveness of SSRIs in primary care before recommending a prescribing policy for the primary care physicians who carry out most treatment of depression. We aimed to review the methodology of current research in the light of recent recommendations,[12] and to identify methodological weaknesses which may have contributed to current uncertainty.

Methods

LITERATURE SEARCH

We performed a literature search which was validated against an independent search performed by another group (see acknowledgements). The target randomised controlled trials (RCTs) were those comparing four SSRIs (sertraline, fluoxetine, fluvoxamine, and paroxetine) with tricyclic and heterocyclic antidepressants. Our search strategy was to use all RCTs cited in five previous meta-analyses,[4-8] a literature search on *Medline* using a search strategy which used the drug names as key words, and a hand search in two journals—*International Clinical Psychopharmacology* and *Acta Psychiatrica Scandinavica*. The independent search concentrated on *Medline* and *Embase*, using drug names as text words and, where available, using medical subject headings (*Medline*) or drug names for searches across all basic index fields (*Embase*). Reference lists of relevant papers and previous systematic reviews were scanned for published reports and citations of unpublished research. The current review does not include any unpublished data used for submission to licensing authorities. This review is part of an ongoing systematic review and meta-analysis now submitted to the Cochrane Collaboration. Studies were excluded if (1) they duplicated results of pre-

Table 1 Summary of main findings

Randomisation	Only one trial (0.8%) had an adequate description of randomisation.
Outcomes	All trials used unstructured assessments of depression. 92% used the Hamilton depression rating scale. Two trials (1.6%) used quality of life measures. One study (0.8%) performed an economic analysis.
Tricyclic dosage	102 studies used tricyclic drugs as comparison treatment. In 22 (22%) the final dose attained was not given. In 26 (25%) the final dose attained was inadequate.
Generalisability of research	Setting: Nine studies (7%) were performed exclusively in primary care. 39 studies (32%) usedhospital inpatients. Duration: 78 trials (64%) compared treatment over six weeks. Four studies (3%) had a follow up of more than eight weeks.
Statistical problems	Median sample size was 64 subjects (interquartile range 42–120). Only 11% of studies have "adequate" statistical power. 69% of studies made multiple comparisons.

KEY POINTS

● There is no consensus on whether selective serotonin reuptake inhibitors (SSRI) or tricyclic antidepressants should be used as the first line treatment of depression.

● Review of all 122 randomised controlled trials (RCTs) of these treatments showed several methodological shortcomings.

● Shortcomings included an inadequate description of randomisation; observer rated measures of depression, no quality of life measures or economic evaluation; inadequate dosage of tricyclics; poor generalisability; and statistical problems.

● Future RCTs should address these problems and aim to inform policy makers.

viously reported RCTs and (2) if the main hypothesis under study was unrelated to depression (for example, some studies specifically examined outcomes such as electrocardiographic changes,[13] sleep changes,[14 15] cognitive status,[16] or psychomotor performance.[17]

ASSESSMENT OF STUDIES

Individual RCTs were assessed on the following criteria:

● Concealment of allocation;

● The outcome measures used (specifically the use of structured assessments, quality of life measures, and economic analyses)

● Dosage of tricyclic medication

● Generalisability of the study, including its setting and duration of follow up;

● Statistical concerns, especially the power of the study and conclusions drawn, use of intention to treat analyses and uncorrected multiple statistical testing (defined as testing multiple outcomes, testing the same outcome over several points in time or performing subgroup analyses).

Results

Table 1 gives a summary of the main findings of the study.

LITERATURE SEARCH

Altogether 133 papers reporting 135 RCTs were identified. Eight of the 133 were excluded as they were duplicate publications and five because their main aim was not the treatment of depression. This left 122 RCTs. (A full list of publications is available from authors.) Most of these studies were published in peer reviewed journals. Thirty one (25%) of the studies appeared in journal supplements often devoted to research on the new drug.

QUALITY OF STUDIES

Concealment of allocation

There is evidence that the quality of RCTs may be judged on the adequacy of their description of the randomisation.[18] We used Cochrane Collaboration guidelines[19] to assess how adequately the randomisation had been described. These guidelines classify description of randomisation

as "adequate", "intermediate", or "inadequate". All but one,[20] of the RCTs reviewed here fell into the "intermediate" category since they gave no information other than stating that the trial was randomised. This is probably due to poor descriptions of the process rather than a serious problem in the conduct of the studies, but it does suggest that the overall quality of trials may not be especially high.

Outcomes

The measurement of depression. Most (115, 92%) of the RCTs we reviewed used the Hamilton rating scale for depression (HRSD)[21 22] as their main outcome. The remainder used the Montgomery-Åsberg scale.[23] These are unstructured assessments which rely on the assessor (who should be a clinician) interviewing the subject and rating symptoms of depression according to specified criteria. The HRSD represented an advance when it was first reported 36 years ago, but there are now alternatives such as semistructured[24] and structured interviews.[25 26]

There are several reasons why such alternatives are preferable. Firstly, observer based assessments such as the HRSD depend heavily on proper blinding of the assessors. All but one of these studies were double blind, but the maintenance of blinding was not reported in any of them. Furthermore, only three RCTs used separate investigators to monitor side effects and assess severity. Since the side effects between the two treatments differ, this increases the likelihood that the investigator rating outcome will not be properly blinded. The unstructured nature of the HRSD introduces the possibility of observer bias especially when there is inadequate blinding,[18] since there is inevitably interest and optimism that the new treatment will be beneficial. Secondly, many of the studies we reviewed were multi-centred. This makes the use of the HRSD problematic since its inter-rater reliability has been questioned.[27 28] Thirdly, the HRSD requires a clinician to assess the severity of depression and relies on an interview lasting 30 minutes. Since most studies measured depression at weekly intervals this increases the cost of using

the assessment with the consequence that large studies using this design would be prohibitively expensive. Many of these problems would be overcome if RCTs used structured interviews, especially those available in computerised format, such as the revised clinical interview schedule or the diagnostic interview schedule.[29][30]

Quality of life. Quality of life (QoL)[31-33] is a measure of an illness's impact on functioning and measures of QoL have repeatedly been recommended in the setting of RCTs.[12][31][34] Both depression and antidepressants impinge on many aspects of an individual's life, affecting social or occupational roles.[35] For example, antidepressants may stop the patient driving. Only two RCTs attempted to measure QoL.[20][36]

There are compelling reasons to measure QoL. Firstly, since they are global measures of functioning, they are capable of making useful comparisons between side effects of drugs. For example, tricyclic antidepressants are often sedating, whereas SSRIs may cause nausea. Unless a global measure of functioning is used, it is impossible to say which of these side effects is more troublesome. Secondly, QoL measures allow a useful assessment of cost effectiveness to be made. For example, direct costs of treatment may be compared to improvements in QoL.

Measuring QoL in depression is not without its problems. Most QoL measures (such as the widely used short form 36 (SF-36)[37] and sickness impact profile (SIP)[38] were designed to monitor QoL in physical illness. These scales include some questions which assess the limitations the illness imposes on activities of daily living and occupational or social roles. They also include many items related to general well-being such as fatigue and mood states. Clearly, the latter items will overlap with direct measures of depression, and this makes the results of QoL measures harder to interpret in depression. Further, as discussed below, most studies are of short duration and it may take longer for these more distal measures of well-being to change. Despite these problems, the failure of any RCT to measure QoL is striking and some of these difficulties could be overcome by using some of the subscales from generic scales such as the SIP. Using these subscales would allow some assessment of social or occupational functioning, which may even be more useful outcomes than changes in scores on depression rating scales.

Economic analysis. There will always be budgetary constraints on any health service and it is necessary to determine the relative cost effectiveness of treatments.[12][39] In depression, the outcomes for an economic analysis might include future health service use, occupational status, impact of the illness on family members, and the use of resources such as social work and voluntary services.

Only one of the RCTs performed an economic analysis.[20] This lack of direct costing has not prevented the publication of economic models based on selected samples of the available data.[40-43] Unfortunately these analyses are flawed because they use non-random selections of clinical trials and tend to over estimate the costs of treatment failure.[44] There is no substitute to performing an economic analysis within an RCT if valid conclusions about cost effectiveness are to be drawn.

TRICYCLIC DOSAGES
There is a consensus of opinion in the UK that tricyclics should be prescribed at doses of more than 100 mg.[45] Many studies only reported a range of dosages attained. In general, RCTs from North America used higher doses of tricyclics than those from Europe. It has previously been noted that the doses of tricyclics used were too low and this might have led to a spuriously low estimate of the drop out rate among subjects taking tricyclics.[5] We used the dose of 125 mg as the minimum satisfactory tricyclic dose. Twenty studies used non-tricyclic comparison treatments. Of the remaining 102 RCTs, the final dosage was unclear in 22, and the average dose was less than 125 mg in 26 (25%). Thus, in only 54 studies (53%) could one be sure that most subjects were receiving adequate doses of the comparison drug.

GENERALISABILITY OF RESEARCH
Setting of research
In the UK, 90% of depressed patients are treated in primary care.[46] This is not reflected in these RCTs: only 9 (7%) studies were performed exclusively in primary care. The remainder were conducted in secondary care and many came from teaching hospitals. Thirty nine studies included hospital inpatients with depression. This is a rare group which forms only 1/1000 cases of depression in UK.[46]

Apart from the need for RCTs to reflect clinical practice, there are specific reasons why the setting of research is relevant. There are important differences between patients with depression in primary and secondary care.[47] Compliance is notoriously poor in primary care.[48] If SSRIs are better tolerated, this advantage should be most obvious in primary care based studies. Tricyclic antidepressants are more widely prescribed in primary care, so it is likely that many secondary care patients will have failed to respond to tricyclics. This could act as a systematic bias against tricyclics. Some trials make treatment resistance an exclusion criterion which may remove this bias. Even so, treatment resistance is usually defined as resistance to a full dose of antidepressant prescribed over a period of four to six weeks, a definition too restrictive to remove this potential bias with confidence. For example, in one trial[49] 65% of the patients entered had been treated previously in the same episode.

Duration of research
Depression runs a chronic course[50] and the usual recommendation is that antidepressants should be prescribed for a minimum of six months.[45] This consensus is based on research demonstrating the superiority of tricyclics[51] and SSRIs[52] over placebo when prescribed for pro-

longed periods. Only four of the trials (3%) had a follow up of more than 8 weeks and most compared treatments over 6 weeks (n = 78, 64%).

Differences in costs and benefits of antidepressant therapy may be particularly marked after the acute illness. For example, patients may tolerate side effects during the acute phases of the illness when they perceive a benefit in taking the medication. In contrast, most of the expense of antidepressant prescribing is a result of the longer maintenance period

STATISTICAL CONCERNS
Sample size and power
Most RCTs were small. The median sample size was 64 subjects randomised to receive active treatment (interquartile range, 42–120). The sample size refers to the number of patients entered into a study. The mean drop out in most meta-analyses[4] has approached one third, therefore the effective sample size in the studies was much less. This glut of small RCTs is partly a consequence of the range of possible comparisons between each of the four SSRIs and numerous tricyclic drugs, but this does not help the policy maker decide which class of drugs should be used as first line treatment.

Small studies lead to type 2 errors. If the median sample size is 32 subjects per treatment group, and one third of subjects drop out of the study, the completer analysis will only include 21 subjects per group. These studies do not have the power to detect even quite important differences between the two treatments. For example, it is widely accepted that 60% of patients treated with a tricyclic antidepressant will recover within six weeks of treatment. A recovery rate of 80% in the SSRIs would be a very important finding, with major clinical significance. To detect such a difference at 80% power and 95% confidence. 91 subjects would be needed in each treatment group. Only 13 (10.6%) of RCTs reviewed here would be able to detect such a difference.

Many of the studies we reviewed did not comment on their low power. One problem of a study comparing two active treatments is the interpretation of a "negative" finding, where no statistically significant difference between treatments is detected. Many of the RCTs we have reviewed reported the failure to find a difference between the two treatments as evidence that they had comparable clinical efficacy, rather than comment on low power. Of 86 studies with fewer than 100 randomised subjects, 51 interpreted a lack of difference in outcome as evidence of comparable efficacy of the two treatments, without mentioning problems of statistical power.

In principle, meta-analysis should overcome the power problems of small trials. However, most trials did not produce elementary statistical information such as means and standard errors of HRSD scores. Many expressed recovery solely in terms of graphical representation and p values. Reporting the size of the observed effect together with confidence intervals was rare. Only two meta-analyses[4][8]

have attempted to compare efficacy ratings on the HRSD. In both, data from only a minority of the identified RCTs were used due to this problem.

Statistical analysis
It was often difficult to determine the method of statistical analysis in these RCTs. In particular there were frequently ambiguities over the use of intention to treat analysis, in which data from subjects who drop out is collected or missing data is substituted with previous values. Forty four studies did not attempt any form of intention to treat analysis and simply analysed data from those who completed the study. In a further 25 studies it was difficult to be sure exactly what form of analysis had been used. A commonly used method (38 studies) was an analysis of data from the last visit, among subjects who had been evaluated since receiving some randomised treatment. This is not a true intention to treat analysis, since it excludes early drop outs.

The importance of an intention to treat analysis is that it maintains the original random allocation upon which the validity of the RCT relies. It is also more relevant to the health economic evaluation of treatments as it takes account of treatments which are effective but poorly tolerated and thereby provides a better picture of the overall effect of the treatment under study.

Another common statistical fault with these studies is their tendency to perform multiple comparisons which lead to type I errors. According to our definition (see above) 84 studies (69%) made some form of multiple comparison, the most common being multiple endpoints, such as comparing lists of side effects or using various subscales of the HRSD.

Discussion
Most RCTs which compare SSRIs with tricyclic antidepressants are underpowered, use inadequate outcome measures over inadequate follow up periods, are based in settings which limit generalisability to primary care, and do not report an economic analysis. It is perhaps not surprising, therefore, that they have so far been unable to address the concerns of clinicians and policy makers eager to know which drugs should be used as first line treatment of depression.

Why is it that so many studies failed to answer these important questions? The simplest answer is that they were not designed to do so. The emphasis on most of these trials is more to do with comparing efficacy and side effects than determining wider prescribing policy. While this is clearly vital information, it is less easy to justify the number of inadequately powered studies. It is remarkable how uninformative many RCTs have been in determining the direction of future prescribing.

These limitations of RCTs may be relevant to other areas of clinical practice where a costly new treatment is introduced for a chronic disease. If the benefits of such a new treatment

are modest there is a sound utilitarian argument that prescribing guidelines should not alter unless there is clear evidence of cost effectiveness. The only sound way to measure cost effectiveness is in the context of a well designed RCT.

1 Anonymous. Interferon beta-1B-hope or hype? *Drug Ther Bull* 1996;34:9–11.

2 The IFNB Multiple Sclerosis Study Group. Interferon beta-1b in the treatment of multiple sclerosis: final outcome of the randomised controlled trial. *Neurology* 1995;45: 1277–85.

3 Kristiansen IV, Eggen AE, Thelle DS. Cost effectiveness of incremental programmes for lowering serum cholesterol concentration: is individual intervention worth while? *BMJ* 1991;302:1119–22.

4 Song F, Freemantle N, Sheldon TA, *et al.* Selective serotonin reuptake inhibitors: meta-analysis of efficacy and acceptability. *BMJ* 1993;306:683–87.

5 Montgomery SA, Henry J, McDonald G, *et al.* Selective serotonin reuptake inhibitors: meta-analysis of discontinuation rates. *Int Clin Psychopharmacol* 1994;9:47–53.

6 Anderson IM, Tomenson BM. Treatment discontinuation with selective serotonin reuptake inhibitors compared with tricyclic antidepressants: a meta-analysis. *BMJ* 1995;310: 1433–38.

7 Montgomery SA, Kasper S. Comparison of compliance between serotonin reuptake inhibitors and tricyclic antidepressants: a meta-analysis. *Int Clin Psychopharmacol* 1995;9(suppl 4):33–40.

8 Anderson IM, Tomenson BM. The efficacy of selective serotonin re-uptake inhibitors in depression: a meta-analysis of studies against tricyclic antidepressants. *J Psychopharmacol* 1995;8:238–49.

9 Hotopf M, Hardy R, Lewis G. Discontinuation rates of SSRIs and Tricyclic antidepressants: a meta-analysis and investigation of heterogeneity. *Br J Psychiatry* 1997;170: 120–27.

10 Freemantle N, House A, Song F, Mason JM, Sheldon TA. Prescribing selective serotonin reuptake inhibitors as a strategy for prevention of suicide. *BMJ* 1994;309:249–53.

11 Effective Health Care. The treatment of depression in primary care. *Effective Health Care Bulletin* 1993;5:

12 Department of Health. *Assessing the effects of health technologies.* London: Department of Health, 1992;

13 Upward JW, Edwards JG, Goldie A, Waller DG. Comparative effects of fluoxetine and amitriptyline on cardiac function. *Br J Clin Pharmacol* 1988;26:399–402.

14 Staner L, Kerhofs M, Detroux D, Leyman S, Linkowski P, Mendlewicz J. Acute, subchronic and withdrawal sleep EEG changes during treatment with paroxetine and amitriptyline: a double-blind randomized trial with major depression. *Sleep* 1995;18:470–77.

15 Kasper S, Voll, Viera A, Kick H. Response to total sleep deprivation before and during treatment with fluvoxamine or maprotiline in patients with major depression—results of a double blind study. *Pharmocopsychiatry* 1990;23:135–42.

16 Fudge JL, Perry PJ, Garvey MJ, Kelly MW. A comparison of the effect of fluoxetine and trazodone on the cognitive functioning of depressed outpatients. *J Affect Disord* 1990; 18:275–80.

17 Ravindram AV, Teehan Md, Bakish D, *et al.* The impact of sertraline, desipramine and placebo on psychomotor functioning in depression. *Human Psychopharmacol* 1995; 10:273–81.

18 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. *JAMA* 1995;273:408–12.

19 Sackett DL, Oxman AD. *The Cochrane Collaboration handbook.* Oxford: Cochrane Collaboration, 1994;

20 Schwartz JA, Speed NM, Brunberg JA, Brewer TL, Brown M, Greden JF. Depression in stroke rehabilitation. *Biol Psychiatry* 1993;33:694–99.

21 Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960;23:56–62.

22 Hamilton M. Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol* 1967;6:278–96.

23 Montgomery SA, Åsberg MA. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979;134: 382–89.

24 Wing JK, Cooper JE, Sartorius N. *The measurement and classification of psychiatric symptoms.* Cambridge: Cambridge University Press, 1974;

25 Robin LN, Helzer JE, Croughan J, Ratcliff KS. National Institute of Mental Health diagnostic interview schedule; its history, characteristics and validity. *Arch Gen Psychiatry* 1981;38:381–89.

26 Lewis G, Pelosi AJ, Araya R, Dunn G. Measuring psychiatric disorder in the community: a standardised assessment for lay interviewers. *Psychol Med* 1992;22:465–86.

27 Cicchetti DV, Prussoff BA. Reliability of depression and associated clinical symptoms. *Arch Gen Psychiatry* 1983; 40:987–90.

28 Rabkin JG, Klein DF. The clinical measurement of depressive disorders. In: Marsella AJ, Hirschfeld RMA, Katz MM, eds. *The measurement of depression.* New York: Guilford, 1987;30–83.

29 Lewis G. Assessing psychiatric disorder with a human interviewer or a computer. *J Epidemiol Community Health* 1994;48:207–10.

30 Greist JH, Klein MH, Erdman HP, Bires JK, Machtinger PE, Kresge DG. Comparison of computer- and interviewer-administered versions of the diagnostic interview schedule. *Hospital and Community Psychiatry* 1987;38:1304–11.

31 Fitzpatrick R, Fletcher A, Gore S, Jones D, Spiegelhalter D, Cox D. Quality of life measures in health care: I: applications and issues in assessments. *BMJ* 1992;305: 1074–77.

32 Fletcher A, Gore S, Jones D, Fitzpatrick R, Spiegelhalter D, Cox D. Quality of life measures in health care. II: design, analysis and interpretation. *BMJ* 1992;305:1145–48.

33 Spiegelhalter D, Gore S, Fletcher A, Jones D, Cox D. Quality of life measures in health care. III resource allocation. *BMJ* 1992;305:1205–09.

34 Pocock SJ. A perspective on the role of quality of life assessment in clinical trials. *Controlled Clinical Trials* 1991; 12(suppl 4):257s–65s.

35 Wells KB, Stewart A, Hays RD, *et al.* The functioning and well-being of depressed patients: results from the medical outcomes study. *JAMA* 1989;262:914–19.

36 La Pia S, Giorgio D, Ciriello R, et al. Double blind controlled study to evaluate the effectiveness and tolerability of fluoxetine versus mianserin in the treatment of depressive disorders among the elderly and their effects on cognitive behavioural parameters. *New Trends in Experimental and Clinical Psychiatry* 1992;8:139–46.

37 Stewart AD, Hays RD, Ware JE. The MOS short-form general health survey. *Med Care* 1988;26:724–32.

38 Bergner M, Bobbitt RA, Cartner WB, Gilson BS. The sickness impact profile: development and final revision of a health status measure. *Med Care* 1981;19:787–805.

39 Drummond MF, Stoddard GL, Torrance GW. *Methods for the economic evaluation of heath care.* Oxford: Oxford University Press, 1987;

40 Jonsson B, Bebbington PE. What Price Depression? The cost of depression and the cost-effectiveness of pharmacological treatment. *Br J Psychiatry* 1994;164:665–73.

41 Beuzen JN, Ravily VF, Souetre EJ, Thomander L. Impact of fluoxetine on work loss in depression. *Int J Psychopharmacol* 1993;8:319–21.

42 Boyer WF, Feighner JP. The financial implications of starting treatment with a selective serotonin reuptake inhibitor or tricyclic antidepressant in drug-naive depressed patients. In: Jonsson B, Rosenbaum J, eds. *Health economics of depression.* London: John Wiley & Son Ltd, 1993; 65–75.

43 Le Pen C, Levy E, Ravily V, Beuzen JN, Meurgey F. The cost of treatment dropout in depression. A cost benefit analysis of fluoxetine vs tricyclics. *J Affect Disord* 1994;31: 1–18.

44 Hotopf MH, Lewis G, Normand C. Are SSRIs a cost effective alternative to tricyclics? *Br J Psychiatry* 1996;168: 404–9.

45 Paykel ES, Priest RG. Recognition and management of depression in general practice: consensus statement. *BMJ* 1992;305:1198–1202.

46 Paykel E. The background: extent and nature of the disorder. In: Herbst K, Paykel E, eds. *Depression: an integrative approach.* Oxford: Heinemann, 1989;3–17.

47 Blacker CVR, Clare AW. Depressive disorder in primary care. *Br J Psychiatry* 1987;150:737–51.

48 Johnson DAW. Treatment compliance in general practice. *Acta Psychiatr Scand* 1981;63 (suppl 209):447–53.

49 Kuhs H, Rudolf GAE. A double-blind study of the comparative antidepressant effect of paroxetine and amitriptyline. *Acta Psychiatr Scand* 1989;80 (suppl 350): 145–146.

50 Piccinelli M, Wilkinson G. Outcome of depression in psychiatric settings. *Br J Psychiatry* 1994;164:297–304.

51 Loonen AJ, Peer PG, Zwanikken GJ. Continuation and maintenance therapy with antidepressive agents. Meta-analysis of research (see comment). *Pharm World Sci* 1991; 13:167–75.

52 Montgomery SA, Duphour H, Brion S, *et al.* The prophylactic efficacy of fluoxetine in unipolar depression. *Br J Psychiatry* 1988;153 (suppl 3):69–76.

53 Hotopf MH, Lewis G, Normand C. *The treatment of depression: evaluation and cost-effectiveness.* London: London School of Hygiene and Tropical Medicine, 1996.