CENTRE FOR HEALTH
PROGRAM EVALUATION

Cost-Effectiveness/Utility Analysis of Two Drug Regimens in the Treatment of Depression

Harri Sintonen

Visiting Fellow, National Centre for health Program Evaluation and Professor, Department of health Policy and Management, University of Kuopio, Finland

Jouko Lönnqvist

Professor, National Public Health Institute, Helsinki, Finland

Olli Kiviruusu

Research Assistant, National Public Health Institute, Helsinki, Finland

CENTRE PROFILE

The Centre for Health Program Evaluation (CHPE) is a research and teaching organisation established in 1990 to:

- undertake academic and applied research into health programs, health systems and current policy issues;
- develop appropriate evaluation methodologies; and
- promote the teaching of health economics and health program evaluation, in order to increase the supply of trained specialists and to improve the level of understanding in the health community.

The Centre comprises two independent research units, the Health Economics Unit (HEU) which is part of the Faculty of Business and Economics at Monash University, and the Program Evaluation Unit (PEU) which is part of the Department of General Practice and Public Health at The University of Melbourne. The two units undertake their own individual work programs as well as collaborative research and teaching activities.

PUBLICATIONS

The views expressed in Centre publications are those of the author(s) and do not necessarily reflect the views of the Centre or its sponsors. Readers of publications are encouraged to contact the author(s) with comments, criticisms and suggestions.

A list of the Centre's papers is provided inside the back cover. Further information and copies of the papers may be obtained by contacting:

The Co-ordinator
Centre for Health Program Evaluation
PO Box 477
West Heidelberg Vic 3081, Australia
Telephone + 61 3 9496 4433/4434   **Facsimile** + 61 3 9496 4424
E-mail CHPE@BusEco.monash.edu.au

ACKNOWLEDGMENTS

ABSTRACT

This paper compares the cost-effectiveness/utility of drug regimens based on fluoxetine and moclobemide in the treatment of depression in Finland. The outcome data are based on a 6 week double-blind RCT (n=209) and the cost data on a cost study linked to the RCT (n=141). Quality of life changes were measured by 15D. Five different outcome measures were used.

Moclobemide showed consistently a better outcome in all outcome measures, but the differences did not generally reach the conventional limit of statistical significance (5%). The difference in the average time-weighted quality-of-life gain on a 0-1 scale was 0.02 (p=0.08). The direct costs were on average lower in the moclobemide regimen, but the difference was not significant (p=0.14), whereas the average total costs (direct and indirect) were almost the same in both regimens. These results suggest that in terms of incremental cost-effectiveness/utility the moclobemide regimen would dominate (produce a better marginal outcome at a equal or less cost), but this conclusion is surrounded by a slightly higher degree of uncertainty than what is conventionally applied. A larger study is needed to give more precision especially to the cost estimates.

Cost-Effectiveness/Utility Analysis of Two Drug Regimens in the Treatment of Depression

Introduction

Depression is an illness with a high incidence and prevalence, but it is difficult to give exact figures due to the fact that much depressive illness remains undiagnosed and untreated. Yet a survey in the USA showed that the lifetime prevalence of major depression/dysthymia is 6% and that of depressive symptoms 23% (Johnson et al. 1992). In the UK family doctors diagnose about two million cases of depression each year (West 1992). In terms of treatment costs, disability and quality of life loss depression represents a significant economic and social burden to the community (Stoudmire et al. 1986, West 1992, Kind and Sorensen 1993, Jönsson and Bebbington 1994, Fadden et al. 1987).

Yet the amount of economic evaluation of depression treatments and mental health care in general has been very scarce (O'Donnell et al. 1988). Indeed, Jönsson and Bebbington (1994) claim that their decision model-based approach serves as a first attempt at cost-effectiveness analysis in the field of antidepressant treatment. Clinical evaluation of depression drugs has focused on risks, in particular on adverse effects, and on beneficial effects in terms of disease-specific physician-administered rating scales, but has ignored the costs (Maynard 1993).

The problem is that in this narrow clinical sense, differences in effectiveness between antidepressants are difficult to demonstrate (Rudorfer and Potter 1989). Moreover, if adverse and beneficial effects are looked at separately, it is difficult to say what the net effect actually is. To establish that it is necessary to consider the probability and magnitude of beneficial effects simultaneously with the probability and severity of various kinds of adverse effect. That can only be achieved if a comprehensive and sensitive measure of health-related quality of life is used to assess the outcome.

Although it is widely accepted that the ultimate and principal aim of depression treatment is to improve quality of life and quality-of-life studies may help to identify outcome differences (e.g. Commonwealth ...1992, Henry 1993), there has been very little research into the quality-of-life effects of depressive illness and its treatment (Katon et al. 1992). In particular this applies to 'genuine' quality-of-life studies, where the patients themselves have assessed their quality of life. Obviously the study by Wells et al. (1989) was the first one to address the quality of life of depressive patients in that way.

Should even quality-of-life studies fail to reveal any differences in outcome, it may still matter which drug regimen is chosen. The drugs may, and do, differ significantly in price. Even if they did not, the treatment process associated with the use of various drugs, and thus the overall costs of the regimens, may differ. In these times of stringent budgets and cost-consciousness, the overall cost is an important consideration. The choice should be based of the relative efficiency or cost-effectiveness of the regimens.

Jönsson and Bebbington (1994) conclude that there is a crucial need for better insights in these matters from fully empirical studies. Ideally such studies should include documentation about direct and indirect costs as well as patients' functioning and quality of life. In this study a step is taken to put this ideal into practice. The purpose is to compare the short-term cost-effectiveness/utility of two drug regimens in the treatment of depression. The drugs are fluoxetine and moclobemide. Fluoxetine is one of the most widely used selective serotonin re-uptake inhibitors (SSRIs). Moclobemide is a reversible inhibitor of momoamine oxidase-A (a RIMA). The quality of life is assessed by the patients themselves in a randomised clinical trial. The estimation of direct and indirect costs is based on a cost study linked to the trial.

Material and methods

The data are based on a six-week multicentre, double-blind randomised trial, carried out in six centres in Finland (Helsinki, Turku, Tampere, Jyväskylä, Kuopio and Oulu). Particular attention was paid to the selection of the field investigators (n=27) and to their training. In the training sessions videotaped, structured interviews were used to increase inter-rater reliability. The investigators were contacted when the data needed checking.

Out of 612 consecutive depressed patients visiting psychiatric out-patient clinics or hospitals, 209 patients over 18 years of age, meeting the DMS-III-R criteria for depressive disorder with the minimum score of 16 on the 17-item Hamilton Depression Rating Scale (HDRS, Hamilton 1967) were enrolled in the trial. After randomisation the fluoxetine group (F-group or F-regimen for short) included 107 and the moclobemide group (M-group or M-regimen) 102 patients (Lönnqvist et al. 1994b).

The study drugs were supplied in identical capsules. During the first two weeks of treatment the dosages were 150 mg of moclobemide twice a day or 20 mg fluoxetine in the morning and a placebo capsule in the afternoon. After that the dosages could be increased to 450 mg or 40 mg, respectively. Use of sedatives and hypnotics was to be avoided as far as possible. Upon strong indication for sedatives, benzodiazepines were allowed.

The patients were assessed at baseline and at 1, 2, 4 and 6 weeks after starting the treatment in terms of HDRS, Montgomery-Åsberg Depression Rating Scale and the CGI Severity and Improvement subscales. Of these, only HDRS scores will be considered in this paper. The health-related quality of life was measured at weeks 0, 2 and 6 by using 15D (Sintonen and Pekurinen 1993) and SF-20 (Stewart et al. 1988). Since SF-20 produces only a score for six health measures (a profile) and does not have a value component to aggregate the profiles into single index numbers, SF-20 is not suitable for measuring quality of life changes as required by cost-utility analysis. Therefore, the results are based on 15D.

The 15D is based on a self-administered questionnaire. The questions relate to the following 15 dimensions: breathing, mental functioning, communicating, seeing, moving, working, perceived health, hearing, eating, eliminating, sleeping, distress, pain, social participation and depression. Each dimension is divided into 4-5 levels. The measure produces a 15-dimensional profile and a total single 15D score over all the dimensions. The 15D score between 0 and 1 for state H (a profile of levels from the 15 dimensions) is calculated as follows:

$$v_H = \sum_j l_j(x_j)[w_j(x_j)],$$

where $l_j(x_j)$ = the average importance weight that the population attaches to various levels of dimension j (j=1, 2,...,m) relative to those of other dimensions, and
$w_j(x_j)$ = the average relative value the population places on various levels of dimension j.

The average importance weights and level values have been elicited from a representative sample of Finnish population by using a three-stage valuation procedure. The 15D score of a state thus reflects its goodness or badness relative to the maximum score of 1 (full quality of life), and the minimum score of 0 (being dead) as assessed by the general public.

This study enables us to use several alternative outcome measures:

A)   The proportion of patients remaining on the drug (c.f. Jönsson and Bebbington 1994).

B)   The proportion of patients remaining on the drug and having a better quality of life (higher 15D score) at week 6 than at baseline.

C)   The proportion of patients having better quality of life (higher 15D score) at week 6 than at baseline, including those no longer on drugs.

D)   The average quality of life gain between week 6 and baseline (15D score at week 6 minus 15D score at baseline).

E)   The average time-weighted quality of life gain. This is the primary outcome measure of this study. The measure was calculated as follows:

E = $\Sigma\{[7(S_2-S_0) + 14(S_6-S_2)]_i + [3.5(S_1-S_0) + 3.5(S_2-S_1) + 14(S_6-S_2)]_j + [7(S_2-S_0) + 7(S_4-S_2) + 7(S_6-S_4)]_k\}/n$,

where S_0, S_1, S_2, S_4, S_6 = 15D score at baseline and at the end of weeks 1, 2, 4 and 6
        i = 1, 2, ..., m = the patients who remained on the drug to the end of trial,
        j = 1, 2, ..., p = the patients who dropped out between baseline and week 2,
        k = 1, 2, ..., q = the patients who dropped out between week 2 and 6,
        n = m + p + q = sample size,
        multipliers 3.5, 7, 14 = average duration of the gain in days.

This measure takes into account the changes in quality of life during the study period and the probability and time of dropout (e.g. p/n = probability of dropping out between baseline and week 2). It was assumed that quality of life changed linearly and dropouts occurred midway between measurements. In all outcome measures involving 15D, the 'intent-to-treat' approach was adopted and a 15D score was assigned to patients with missing values as follows:

If the patient dropped out due to adverse effects or poor response to treatment, the baseline score was assigned to all remaining missing measurements. If the patient did not drop out, but the week 2 score was missing, the average of scores at baseline and week 6 was assigned, and if the week 6 score was missing, the week 2 score was assumed to apply. If the patient discontinued treatment, because he felt the medication was no longer needed, the latest score was assigned to the remaining missing measurements. This principle was applied also to those with an unknown reason for dropout, since the changes in their HDRS and 15D scores suggested that also they most likely belonged to the 'no need' group.

The costs of the whole treatment process associated with the drug regimens were estimated on the basis of a cost study linked to the trial. From the 27 doctors 18 agreed to participate in the cost study involving a structured interview of the patient at baseline and at the end of treatment at week 6. In the interview the depression-related use of health care and non-health care (social services and informal help) resources as well as loss of resources (productive time lost) was recorded in detail. The number of patients in the cost study was 70 in the M-regimen and 71 in the F-regimen.

The use of health care and non-health care resources represents the direct cost of the treatment process. In addition, the treatment process requires varying amounts of patients' time (travel, waiting and treatment time associated with outpatient visits, hospital stay) that could otherwise be devoted to alternative uses (work or leisure). The treatment process may also require time from the patient's family or friends (e.g. helping in domestic duties, accompanying during visits). The value of that time in the best alternative use is the sacrifice, that the patients (and their informal carers and helpers) have to make in order to get and undergo treatment. Moreover, the illness may result in a general inability to work or engage in household production (loss of productive resources). These items of opportunity cost represent the indirect cost of the illness and its treatment process. In this study only the costs incurred during the six week treatment period are reported.

The resource use was valued at the 1991 prices. The working time of various categories of health care and social service personnel was valued at their average wage rate. Working time lost was valued at the average wage rate in Finland and household production lost at the wage rate of the home helper. Care and help provided by the family, relatives etc. was valued at the average wage rate, if the carer/helper had lost working time, otherwise at the wage rate of the home helper. To all wage rates 40% was added to account for social security contributions. The visits to various outpatient facilities were valued at the average cost per visit in each kind of facility. The hospital

days at various hospital levels/types were valued at the average daily cost at that level/type of hospital in Finland. The drugs used were valued at their retail price.

Differences between the groups in outcome and costs were tested by using t-test, chi-square test and 95% confidence intervals (CI). Since the 15D scores underlying outcome measures D and E are limited to 0-1 it is possible that their distribution is not normal, especially if the scores converge close to either limit. Before applying the t-test to measures D and E, the distributions were tested for normality by using a normal probability plot.

Results relating to outcome are based both on the total samples and the groups that completed the cost study. Cost results are based on the groups that completed the cost study. The short-term cost-effectiveness/utility of the regimens are examined in the light of cost-effectiveness/utility ratios (cost incurred/outcome obtained). Ratios based on outcome measures A, B and C can be regarded as cost-effectiveness ratios. Ratios based on measures D and E are cost-utility ratios, since they are based on a quantitative estimates of the quality-of-life gain.

In addition, attention will be paid to marginal cost-effectiveness/utility. They provide an estimate of the extra cost for the additional effectiveness or utility of the more effective regimen. Cost-effectiveness/utility results are based both on the total samples and the groups that completed the cost study.

Results

At baseline, the treatments groups were well matched in terms of psychiatric measures, quality of life and background variables apart from the distribution by gender: there were more women in the M-group (Chi^2=4.2, p=0.04)(Lönnqvist et al. 1994b). The average baseline 15D score was 0.719 (SD=0.118) in the F-group and 0.719 (SD=0.098) in the M-group.

The average baseline 15D score of patients, who participated in the cost study (0.712, SD=0.109) did not differ from that of the non-participants (0.734, SD=0.111) nor did that of the patients, who completed the cost study (0.7256, SD=0.101) differ from the average score of those, who did not participate or complete the cost study (0.712, SD=0.116). Among the cost study participants, the average baseline 15D score was 0.712 (SD=0.117) in the F-group and 0.712 (SD=0.097) in the M-group. The corresponding scores among those, who completed to cost study, were 0.734 (SD=0.109) and 0.718 (SD=0.092), respectively.

Outcome in the treatment groups

The upper part of table 1 shows the outcome of treatment in terms of various measures in the total samples. There was no significant difference between the regimens in the probability of remaining on the drug. Of the 22 patients (20.6%) who dropped out in the F-group, 16 did so due to adverse effect, 2 to poor response to treatment, 1 to unnecessary medication and 3 for an unknown reason. For the 18 patients (17.6%) in the M-group the reasons were: 10 for adverse effect, 6 for poor response to treatment and 2 for an unknown reason.

When the probability of remaining on the drug was considered in conjunction with the probability of being better-off in quality of life, there appeared to be a statistically significant difference between the groups in favour of moclobemide. When the probability of being better-off in quality of life was considered alone without paying attention to whether the patients had remained on the drug or not, the difference did not reach the limit of statistical significance.

The average quality-of-life gain on a 0-1 scale was about 0.06 points in the F-groups and about 0.08 points in the M-group, but the difference was not statistically significant. The average time-weighted quality-of-life gain, that is, the extra number of healthy days during the treatment period of 42 days, was about 0.5 days in the F-group and about 0.8 days in the M-group, but the difference did not quite reach the significance level of 5%. The distributions of the 15d scores before and after treatment, or of the gain measures did not show any clear deviation from normal for either drug in a normal probability plot.
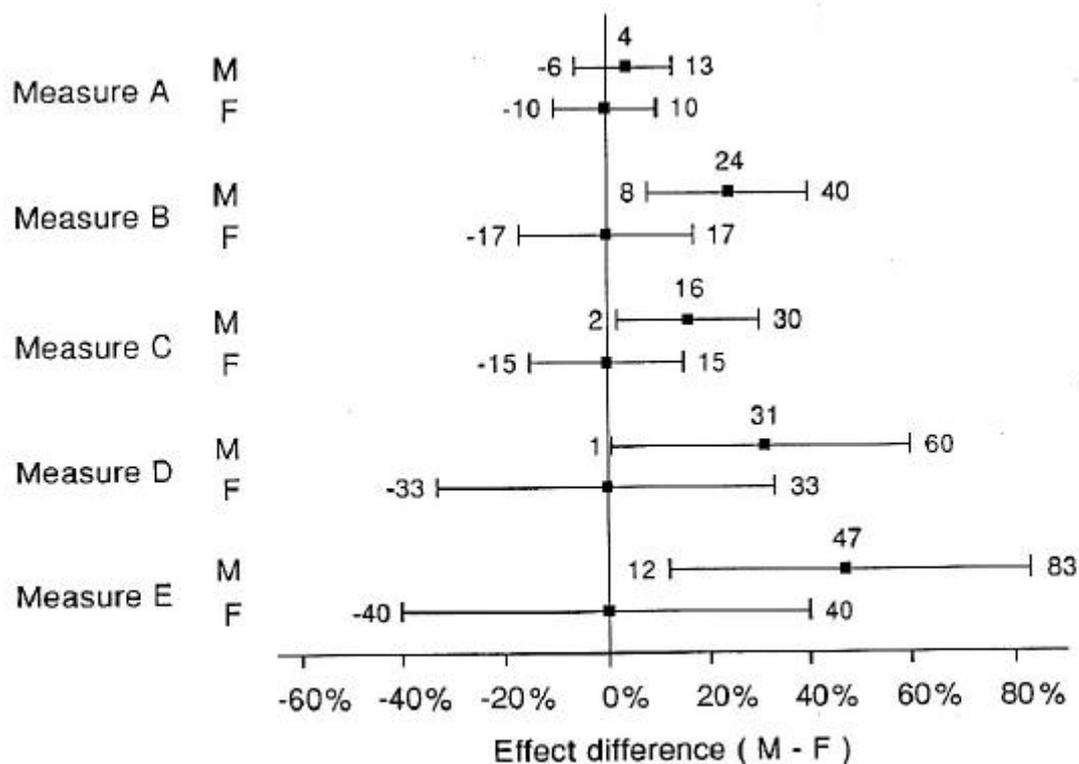
Table 1

The outcome of treatment in terms of various outcome measures
in the total samples and in the groups that completed the cost study

Total samples	Fluoxetine N=107 (95% CI)	Moclobemide n= 102 (95% CI)	Statistical significance
Outcome measure			
A) % of patients remaining on the drug	79.4 (71.2-87.2)	82.4 (74.8-89.9)	Chi^2/p=0.59
B) % of patients remaining on the drug and having a better quality of life at week 6 than at baseline	57.0 (47.5-66.5)	70.6 (61.6-79.6)	Chi^2/p=0.04
C) % of patients having better quality of life at week 6 than at baseline	62.6 (53.3-71.9)	72.6 (63.7-81.4)	Chi^2/p=0.12
D) The average quality of life gain between week 6 and baseline	0.059 (0.039-0.079)	0.078 (0.060-0.095)	t/p=0.17
E) The average time-weighted quality of life gain	0.537 (0.323-0.751)	0.790 (0.601-0.980)	t/p=0.08
Groups that completed the cost study	n=53	n=55	
C) % of patients having better quality of life at week 6 than at baseline	67.9 (54.9-80.9)	85.5 (75.8-95.1)	Chi^2/p=0.03
D) The average quality of life gain between week 6 and baseline	0.080 (0.0480-0.1117)	0.092 (0.0667-0.1180)	t/p=0.54
E) The average time-weighted quality of life gain	0.732 (0.386-1.079)	0.969 (0.683-1.255)	t/p=0.29

In figure 1 the results in terms of various outcome measures are presented graphically showing for each measure the point estimate of effect difference and its 95% confidence interval (CI) as suggested by Braitman (1991). For each measure, the point estimate of F-regimen is on the axis through zero (with its 95% CI across the axis), since its point estimate is 'lower' for each measure. Above it is the corresponding information for M-regimen in comparison with the point estimate of F-regimen. For example, on measure E, the M-regimen shows a 47% improvement compared with the F-regimen [100*(.790-.537)/.537], with a CI of 12% to 83%. This would suggest that moclobemide would be clearly superior to fluoxetine. However, there is a high variability around the point estimate of fluoxetine as well. Therefore, the confidence intervals to some extent overlap so there is a small chance the conclusion does not hold.

Figure 1

The effect difference between the drug regimens (F=fluoxetine, M=moclobemide)

Outcome among patients, who completed the cost study

The lower part of table 1 shows the outcome of treatment in terms of various measures in the treatment groups that completed the cost study. Apart from measure C, the differences between the groups are not statistically significant. All point estimates are higher than in the corresponding total samples, since these groups do not include those, who dropped out with lower 15D scores.

The costs

Table 2 shows the breakdown of direct and indirect costs presented in 1000 Finnish marks (1 AUD=3.96 FIM, 1 USD=5.55 FIM, 1 GBP=8.26 FIM). Within direct costs, there is a significant difference between the regimens in the average cost of study drugs and other psychotrophic drugs, but these items represent only 1.7-1.8% of total costs and 2.5-2.9% of direct costs. The direct costs are higher in the F-group, mainly due to higher inpatient costs, but the difference between the regimens is not significant. In fact, inpatient care is by far the greatest single cost item representing 69-78% of direct costs.

The indirect costs represent 25-43% of total costs. There is a significant difference between the regimens in indirect costs, the costs being higher in the M-group. In total costs there is no significant difference between the regimens.

Figure 2, similar to figure 1, shows graphically for direct, indirect and total costs the point estimate of cost difference and its 95% confidence interval. The figure shows the considerable differences between the groups in direct and indirect costs, but the wide confidence intervals make it to some extent uncertain, whether these differences really exist.

Table 2

The breakdown of direct and indirect costs by items (1000 FIM)

Cost item	Fluoxetine n=53 (95% CI)	%	Moclobemide n=55 (95% CI)	%	Statistical significance
Outpatient care					
a) Visits by the patient to outpatient services	3.29 (2.87-3.71)	14.3	3.37 (2.99-3.73)	15.5	t/p-0.78
b) Home visits	0.08 (-0.00-0.17)	0.4	0.06 (-0.01-0.13)	0.3	t/p=0.66
Inpatient care	13.41 (7.91-18.90)	58.3	8.49 (4.60-12.38)	39.2	t/p=0.15
Study drugs	0.39 (0.36-0.42)	1.7	0.34 (0.33-0.35)	1.6	t/p=0.001
Other psych. drugs	0.03 (0.02-0.05)	0.1	0.02 (0.01-0.03)	0.1	t/p=0.05
Direct costs	17.20 (11.78-22.61)	74.8	12.28 (8.448-16.10)	56.7	t/p=0.14
Work time and household production lost	4.70 (2.99-6.41)	20.5	7.82 (5.63-10.00)	36.1	t/p=0.03
Care/help by family etc	1.08 (0.62-1.54)	4.7	1.56 (0.91-2.21)	7.2	t/p=0.23
Indirect costs	5.78 (4.07-7.49)	25.2	9.38 (7.17-11.59)	43.3	t/p=0.01
Direct and indirect = total costs	22.98 (16.49-29.46)	100.0	21.66 (16.95-26.35)	100.0	t/p=0.74

Figure 2

The cost difference between the drug regimens (F=fluoxetine, M=moclobemide) in the light of direct, indirect and total costs (the point estimates and the 95% confidence intervals)

Cost-effectiveness/utility
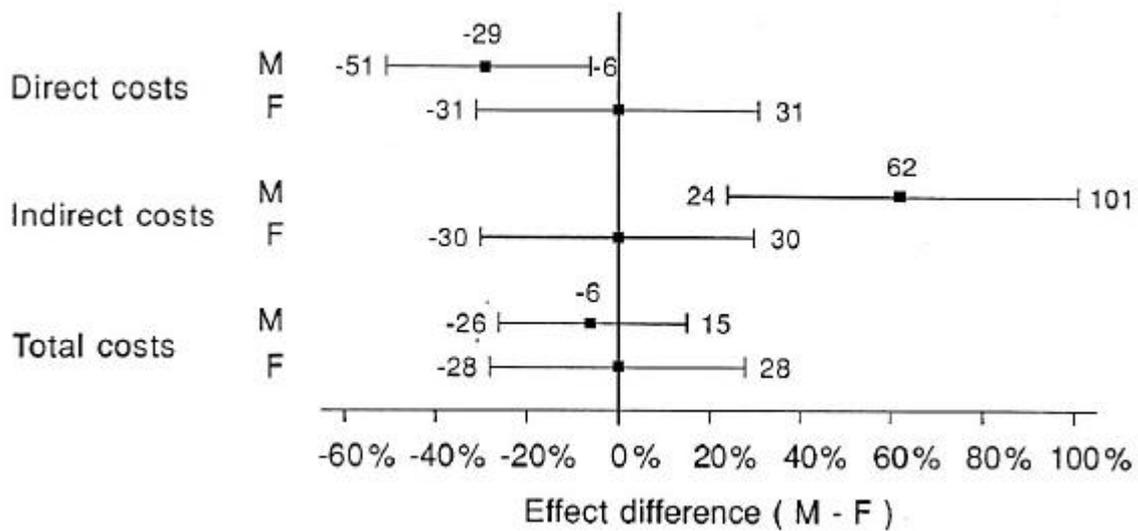
Cost-effectiveness/utility

Table 3 shows cost-effectiveness/utility ratios with various outcome measures, when **direct costs** are used as a measure of cost. At the upper part of the table, the result are from the groups that completed the cost study. At the lower part, the results are from the total samples with the assumption that the average costs would have been the same in the total samples as in the groups that completed the cost study.

All the ratios are in favour of the M-regimen: depending on the outcome measure, the direct cost per unit of outcome is 31-52% less in the M-regimen. The conclusion remains regardless of whether ratios from the cost study or total samples are looked at.

Table 3

Short-term cost-effectiveness and cost/utility ratios with
direct costs and various outcome measures in the treatment groups

Groups that completed the cost study Cost outcome (Cost in 1000 FIM)	Fluoxetine n=53	Moclobemide n=55	Difference in favour of moclobemide
C) Cost/patient having better quality of life at week 6 than at baseline	25.3	14.4	-43%
D) Cost/average quality of life gain between week 6 and baseline	215.2	132.9	-38%
E) Cost/average time-weighted quality of life gain	23.5	12.7	-46%
Total samples	n=107	n=102	
A) Cost/patient remaining on the drug	21.7	14.9	-31%
B) Cost/patient remaining on the drug and having a better quality of life at week 6 than at baseline	30.2	17.4	-42%
C) Cost/patient having better quality of life at week 6 than at baseline	27.5	16.9	-38%
D) Cost/average quality of life gain between week 6 and baseline	290.0	158.1	-45%
E) Cost/average time-weighted quality of life gain	32.1	15.5	-52%

The cost-utility ratios based on outcome measures D and E are the most interesting ones. Cost per average time-weighted quality-of-life gain (measure E) indicates the cost of producing one extra quality-adjusted day, that is, one extra day in full quality of life. The cost is 12.7-15.5 thousands FIM for moclobemide and 23.5-32.0 thousands FIM for fluoxetine. The incremental cost-utility ratios ($Cost_F - Cost_M$)/($U_F - U_M$), based either on the cost study or total samples, suggest that the M-regimen would dominate, that is, the M-regimen appears to produce a better outcome at a lower cost (positive marginal effectiveness/utility and negative marginal cost).

Cost per average quality-of-life gain (measure D) can not be directly interpreted in the same way, since the outcome measure does not contain the time dimension, that is, the duration of the gain. By multiplying the gain by 21 (42 days/2 = 21, assuming a linear change in the quality of life between baseline and week 6), we get an estimate of 6.3-7.5 thousands FIM per quality-adjusted extra day for moclobemide and 10.2-13.8 thousands FIM for fluoxetine. This means that extrapolating linearly between endpoint measurements yields a time-weighted quality-of-life gain estimate double and cost/extra quality-adjusted day estimate half of what is obtained when the

actual development of quality of life during the treatment period is taken into account (measure E). The M-regimen appears to dominate (as is the case regardless of what outcome measure is used).

The cost-effectiveness/utility ratios based on the total costs (not presented here) also suggest the M-regimen would still dominate, although less clearly: it appears to produce a better outcome at the same total cost (positive marginal effectiveness/utility with no difference in total costs).

Sensitivity analysis

The problem with cost-effectiveness/utility ratios is that statistical test for their differences raise a number of complexities, especially if the numerator and denominator cannot be assumed to be independent as cost and effectiveness/utility hardly can, since they are associated through a production function (Drummond and O'Brien 1993). However, to get an idea about the relationship between the ratios and the degree of uncertainty surrounding our conclusion based upon them, figure 3 was produced. The figure shows the difference in point estimate ratios for outcome measures C, D and E with the interval defined by the 'worst' ratio (the combination of the high cost and poor outcome estimate of the 95% confidence intervals) and 'best' ratio (low cost, good outcome).

The figure shows that in the light of the cost-utility ratios, especially of the one based on outcome measure E, it is quite likely that the M-regimen is superior to the F-regimen. For the opposite to be true it would be required that the worst results both in terms of outcome and costs would have to be true for moclobemide and the best results for fluoxetine. Such a combination would be highly unlikely to occur.

Figure 3
The difference in cost-effectiveness/utility ratios between the drug regimens (F=fluoxetine, M=moclobemide)
(the point estimates and the 'best' and 'worst' combinations of the
95% confidence intervals of outcome and direct costs)
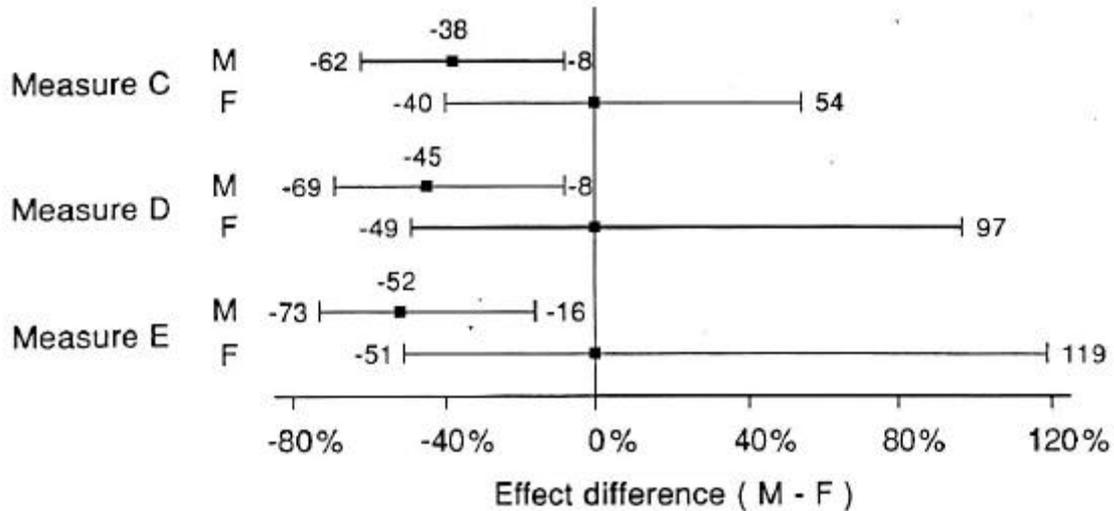
The cost of inpatient care was by far the greatest single cost item and source of variance of direct costs. Although the large difference in average inpatient costs between the regimens was not statistically significant (due to high variance) nor was the difference in proportion of inpatients at baseline, it was still suspected that even a small difference in the latter might make a considerable difference in the former. To check the sensitivity of direct cost estimates in this respect a regression model run for direct costs with 15D score, patient status at baseline (inpatient or outpatient), age and regimen as explanatory variables (R^2=.541, F=30.4, p=0.0000). The patient status turned out to be by far the most important explanatory variable. The average indirect costs predicted by the model with having the proportion of inpatients at the average across the samples were as follows: 13.72 (SD=3.59) thousands FIM for moclobemide and 15.69 (SD=4.26) thousands for fluoxetine (t/p=0.01). Thus an equal proportion of inpatients reduces the difference in direct costs, but does not remove it.

Discussion

So far we are aware this study is unique at least in two respects. It is the first fully empirical economic evaluation of drug regimens in the field of psychiatry. It is also first in that field to link a cost study to a RCT.

The measurement of quality of life was 'genuine' in the sense that it was based on patients' subjective experience, that is, the patients themselves to assessed and recorded their quality of life on a self-administered 15D questionnaire. A very small number of missing responses indicates that they were able to do so. They were also able to do it quite reliably as the correlations ranging from .55 to .73 between the 15D scores at weeks 0, 2 and 6 indicate.

The patients were also surprisingly capable of reporting the use of resources in the cost study interviews. Although the quality of data in some cases necessitated some 'detective work' to amend the quality, at the end the data are reliable enough to give an adequate picture of the relative costs of the regimens.

Five outcome measure were used. One was the proportion of patients remaining on the drug (A), which showed no difference between the regimens. This measure used by Jönsson and Bebbington (1994) assumes that if the patient did not drop out due to adverse effects or poor response to treatment, his treatment was successful. This measure may give too optimistic a picture of success: even if the patient remains on the drug, he may not be better-off or may even be worse-off than without the drug.

This was clearly shown by our second outcome measure, the proportion of patients remaining on the drug and having better quality of life at the end of treatment that at the onset (B). On this measure the M-regimen was significantly better. Yet in that regimen 14% of those, who remained on the drug were worse-off in terms of quality of life at the end of treatment than at the beginning. In the F-group the percentage was 28.

Neither of these measures considers, what happens to the treatment 'failures', i.e., dropouts. For various reasons trials pay very little attention, if at all, to that question. Yet in assessing the performance of treatments, what is the fate and state of those who dropped out should be of great importance, especially if the rate of failure is high. Since the underlying trial of this study did not address that aspect of outcome explicitly, we had to resort to assumptions presented earlier. If the patient dropped out due to adverse effects or poor response to treatment, it is likely that his quality of life on the drug was no better or was worse than without the drug. Therefore, assigning his baseline score to all remaining missing measurements is a reasonable, indeed a cautious, assumption of his post-dropout state. It is also a reasonable assumption to assign the latest score to the remaining missing measurements, if the reason for dropout was 'no need' or unknown, since the development of the HDRS and 15D scores of the latter suggested that also they most likely belonged to the 'no need' group.

With these assumptions it was possible to form outcome measures C, D and E. Measure C shows the proportion of patients having better quality of life at the end of treatment than at the onset regardless of whether they had remained on the drug or not. There was no significant difference between the regimens in this proportion.

Measure C does not take into account the possibility that there might be a quantitative difference between the regimens in how much better-off the patients on average are in terms of quality of life at the end of treatment than at the beginning. Measure D does that, but it showed no significant difference between the regimens.

To put the average quality-of-life gain of 0.06-0.08 in six weeks (measure D) into perspective, it may be illuminating to compare it to that experienced by some other patients groups in the light of 15D. For example, at three months after coronary bypass grafting (n=56), the average gain was 0.1 over the average 15D score of 0.73 before the operation (Sintonen and Pekurinen 1993). On the other hand, during three weeks of rehabilitation, asthma patients with musculosceletal symptoms (n=102) gained on average 0.07 from the average pre-rehabilitation score of 0.84 (Karhu and Punakivi 1993). Thus, the gain of these patient groups was comparable to that achieved by the depression patients.

Measure D is still deficient in the sense that it does not allow for the possibility that there may be differences between the regimens in the speed of the effect emerging, the development of quality of life during the study period (it may go up and down following a different pattern) and the probability and time of dropout. Our principal outcome measure E allows for them, but at the conventional 5% level, the difference between the regimens was not statistically significant. However, if one is willing to accept an 8% rather than 5% risk of the difference having arisen by chance, one could regard the M-regimen as superior. With slightly larger samples, the 5% level might be achievable.

This brings us to the problem of sample size, which is even more acute with the cost data, where the final sample sizes were about half of the total samples. Without prior knowledge of the high variance eventually observed in most cost items, the sample sizes proved to be too small for the cost differences to become statistically significant. For example, given the sample sizes, we ended up with a 14% risk of the difference in direct costs having arisen by chance. However, with sample sizes of 102 and 107 (the total samples) the 5% level might have been attained (provided of course that a difference really exists). Since the subsamples of the cost study were not different from the total samples at baseline, this generalisation to the total samples may be legitimate to do. However, a larger cost study would be needed to give more precision to the estimate of the difference. For a more thorough discussion on the problems with sample size in economic evaluations, a recent article by Drummond and O'Brien (1993) is worth consulting.

Assuming that the observed differences in quality-of-life gain and direct costs really exist, are they also 'clinically' or quantitatively significant or important? If we take the arbitrary limit of 20%

difference, customary in clinical studies (Burnard et al. 1990), all outcome measures apart from A indicate a 'quantitatively' significant difference. As to measure D, which showed a difference of about 0.02 points between the regimens, this question can be considered more substantially. This difference may just be at the lower end of being 'quantitatively' significant in the sense that the patients can feel a difference overall. This is inferred from a survey, where 1231 patients assessed their status on 15D about half a year apart. In the second measurement they also reported separately, whether they now felt better or worse than or much the same as in the first measurement. For those who felt better or worse, the 15D score had changed on average +0.037 and -0.034, respectively, whereas for those who felt much the same the score had remained on average virtually unchanged (Sintonen and Pekurinen 1989).

A difference of 4900 FIM (or 2000 FIM as predicted by the regression model) in direct costs over a treatment period of six weeks is quantitatively significant, especially, when a large amount of such treatment episodes are provided. For example, with 100000 episodes the overall cost difference is 200-490 million, which should be quantitatively important to the health care budget of any country.

Sensitivity analysis does not have a major role in this study, since the measurement of outcome and costs was stochastic, that is, the items have an observed mean and variance. Uncertainty associated with the estimates can then be handled by the testing of statistical significance. The measurement of costs was stochastic as far as resource use was concerned, but to reduce the already high variance, deterministic unit costs/prices, that is, averages over the appropriate settings in Finland, were used. These estimates are so reliable that subjecting them to sensitivity analysis is not necessary.

Therefore, sensitivity analysis was applied only to test, whether the proportion of inpatients at baseline in the regimens makes a difference in direct costs with the inpatient costs being by far the greatest item. The result did not affect our basic conclusion. In addition, since statistical testing of ratios raise complexities, the relationship between the ratios and the degree of uncertainty surrounding our conclusion based upon them was examined with the ranges of the 'best' and 'worst' ratios. The conclusion was that the probability of the fluoxetine regimen turning out to be better in terms of cost-utility is small.

Indirect costs were estimated and are presented in this study, although the main body of results are based on direct costs only for three reasons. First, the cost-effectiveness/utility results would not have changed much even if indirect costs would have been included. Second, their exclusion is recommended (e.g. Commonwealth ...1992) on the grounds that it is doubtful whether there is an economic loss to society through the patient's lack of productive capacity and absence from work (for short-term absence, production will be made up by other workers or by the sick person on the return to work; for long-term absence, production will be made up by a replacement worker, otherwise unemployed). Third, it has been argued that cost-utility analyses can be used legitimately only to examine, what additional utility can be bought by allocating health care resources to different programs within a health care budget. Therefore, only direct costs should be considered (Birch and Gafni 1991, Gerard and Mooney 1993).

The results suggest that the moclobemide regimen may be slightly better in terms of cost-effectiveness/utility than the fluoxetine one. However, in considering the conclusion, some points should be borne in mind.

First, the conclusion is surrounded by a slightly higher degree of uncertainty than is conventionally applied. There is a clear case for a larger study, giving more precision to the cost estimates in particular.

Second, the results refer to short-term cost-effectiveness/utility only. On the basis of these ratios is not possible and legitimate to estimate, what e.g. the cost/QALY would be. The study period may be too short for the full effects of treatment to emerge (Lönnqvist et al. 1994a). It is not known, how long the effect of treatment will on average last. For example, it would make a huge difference in the cost/QALY, if the patients relapse on average in one or 10 year's time. Depression is a long-term illness so a longer-term analysis would be required, where maintenance treatment and the probability of relapses and suicide would have to be considered to be able to say something definitive about the relative cost-effectiveness/utility or cost/QALY of the regimens. Yet choices between drugs have to be made even in the absence of such a long-term evidence. In these choices short-term analyses can be useful by bringing up important factors to be considered in a more systematic way than previously and thus encourage a rational diffusion and use of medicines (Drummond 1992).

Third, one has to recall the difference between **efficacy** and **effectiveness** of the drug. The results of clinical trials tend to indicate the efficacy (benefits under ideal conditions of use), rather than effectiveness (benefit under average conditions of use). The patients are chosen to trials with certain criteria. The trials are usually carried out by more experienced, trained and dedicated staff than is the case in the health care system at large. These factors with better compliance due to more frequent check-ups achieve that better outcomes are to be expected in clinical trials than in the routine treatment.

Strictly taken the outcome results of this study reflect efficacy rather than effectiveness. However, the difference between them may not be noteworthy, since the treatment protocol of the trial did not deviate much from that followed under average conditions. This also means that our estimates of direct costs may be rather close to those to be expected under average conditions of care. Even if cost and outcome estimates would deviate to some extent from those in average conditions, it is still unlikely that the comparative results would be much affected, since for example the difference in the frequency of drug administration between the regimens is not so essential as to cause much difference in patient compliance (Enlund 1982).

REFERENCES

Birch S, Gafni A. Cost effectiveness/utility analyses: do current decision rules lead us to where we want to be? CHEPA Working Paper 91-6, McMaster University, Hamilton, 1991.

Braitman LE. (1991) Confidence intervals assess both clinical significance and statistical significance. Annals of Internal Medicine 114, 515-517.

Burnard B, Kerman WN, Feinstein AR. (1990) Indexes and boundaries for 'quantitative' siginificance' in statistical decisions. Journal of Clinical Epidemiology 43, 1273-1284.

Commonwealth Department of Health, Housing and Community Services (1992) Guidelines for the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee. Australian Government Publishing Service, Canberra.

Drummond M. (1992) Cost-effectiveness guidelines for reimbursement of pharmaceuticals: is economic evaluation ready for its enhanced status? Health Economics 1, 85-92.

Drummond M, O'Brien B. (1993) Clinical importance, statistical significance and the assessment of economic and quality-of-life outcomes. Health Economics 2, 205-212.

Enlund H. (1982) Drugs use and patient compliance in hypertension care in Eastern Finland. Kuopion Korkeakoulun julkaisuja (Kansanterveystiede, Kuopio).

Fadden GB, Kuipers L, Bebbington PE. (1987) Caring and its burdens: a study of the relatives of depressed patients. British Journal of Psychiatry 151, 660-667.

Gerard K, Mooney G. (1993) QALY league tables: handle with care. Health Economics 2, 59-64.

Hamilton M. (1967) Development of a rating scale for primary depressive illness. British Journal of Social and Clinical Psychology 6, 278-296.

Henry JA. (1993) Debits and credits in the management of depression. In Eccleston D. (ed) The economic evaluation of antidepressant drug therapy. British Journal of Psychiatry 163, Suppl. 20, 33-39.

Johnson J, Weissman MM, Klerman GL. (1992) Service utilisation and social morbidity associated with depresssive symptoms in the community. Journal of the American Medical Association 267, 1478-1483.

Jönsson B, Bebbington PE. (1994) What price depression? The cost of depression and cost-effectiveness of pharmacological treatment. British Journal of Psychiatry 164, 665-673.

Karhu I, Punakivi L. (1993) Ilmastonvaikutus TULE-oireisten astmaatikkojen kuntoutuksessa. Miina Sillanpään säätiön julkaisuja B:9, Helsinki.

Katon W, von Korff M, Lin E et al. (1992) Adequacy and duration of antidepressant treatment in primary care. Medical Care 30, 67-76.

Kind P, Sorensen J. (1993) The cost of depression. International Clinical Psychopharmacology 7, 191-195.

Lönnqvist J, Sihvo S, Syvälahti E et al. (1994a) Continuation treatment of depressive disorders with moclobemide and fluoxetine (under editorial consideration).

Lönnqvist J, Sintonen H, Syvälahti E et al. (1994b) Antidepressant efficacy and quality of life in depression: a double-blind study with moclobemide and fluoxetine. Acta Psychiatrica Scandinavica (forthcoming).

Maynard A. (1993) Cost management: The economist's viewpoint. In Eccleston D. (ed) The economic evaluation of antidepressant drug therapy. British Journal of Psychiatry 163, Suppl. 20, 7-13.

O'Donnell O, Maynard A, Wright K. (1988) The economic evaluation of mental health care: A review. Discussion paper 51, Centre for Health Economics, University of York.

Sintonen H, Pekurinen M. (1993) A fifteen-dimensional measure of health-related quality of life (15D) and its applications. In Walker SR, Rosser RM (eds) Quality of life assessment. Key issues in the 1990s. Kluwer Academic Publishers, Dordrecht, 185-195.

Sintonen H, Pekurinen M. (1989) A generic 15 dimensional measure of health-related quality of life (15D). Journal of Social Medicine 26, 85-96.

Stewart AL, Hays RD, Ware JE. (1988) The MOS short-form general health survey. Reliability and validity in a patient population. Medical Care 26,724-735.

Stoudmire A, Frank R, Hedemark N et al. (1986) The economic burden of depression. General Hospital Psychiatry 8, 387-394.

Wells KB, Stewart A, Hays RD et al. (1989) The functioning and well-being of depressed patients. Results from the Health Outcome Study. Journal of the American Medical Association 262, 914-919.

West R. (1992) Depression. London: Office of Health Economics.