

Meta-analysis of randomised controlled trials of fluoxetine v. placebo and tricyclic antidepressants in the short-term treatment of major depression

P. BECH, P. CIALDELLA, M. C. HAUGH, M. A. BIRKETT, A. HOURS, J. P. BOISSEL and G. D. TOLLEFSON

Background Previous meta-analyses of fluoxetine as an antidepressant have many methodological problems, including diagnosis of major depression, validity of outcome measures and lack of intention-to-treat analyses.

Aims To provide an estimate of the effect of fluoxetine compared with placebo and tricyclic antidepressants (TCAs), and to investigate reasons for early discontinuation from acute treatment.

Method Randomised trials were analysed using both intention-to-treat, efficacy and end-point.

Results Fluoxetine was superior to placebo but effect size was low. In trials comparing fluoxetine v. TCA, the results for all trials and for the USA trials showed a trend in favour of fluoxetine. Those for the non-USA trials showed a trend in favour of TCA. When combined, the results showed that significantly fewer patients on fluoxetine discontinued treatment because of adverse events.

Conclusion Fluoxetine is superior to placebo, irrespective of the analytical approach used, whereas the results obtained v. TCAs depend on the approach used. Hence, the results should be interpreted in this light.

Declaration of interest P. Bech is Head of a World Health Organization Collaborating Centre for psychometrics. J. P. Boissel, P. Cialdella, M. C. Haugh, and A. Hours were financed by APRET, a non-profit research organisation, for this project. M. A. Birkett and G. D. Tollefson are employed by Eli Lilly and Company.

Previously published meta-analyses of selective serotonin reuptake inhibitors (SSRIs) v. tricyclic antidepressants (TCAs) or placebo (Anderson & Tomenson, 1994; Greenberg *et al.*, 1994; Anderson, 1998) were based on published data only and did not analyse data for all randomised patients (the intention-to-treat approach) since these were not available in the published reports. The only previous meta-analysis with this approach was reported by Bech & Cialdella (1992). In the present analysis we used the Eli Lilly and Company (Lilly) fluoxetine database and included patients from published and unpublished randomised clinical short-term trials of fluoxetine. A protocol described our objectives, inclusion and exclusion criteria for trials, and the analyses to be performed. We used different analytical approaches for completers and non-completers. Our objectives were to obtain quantitative estimates of the fluoxetine treatment effect compared with: (a) placebo; (b) TCAs; and (c) to analyse the reasons for early discontinuation from treatment.

MATERIAL AND METHOD

Types and sources of data

In keeping with our original protocol all randomised clinical trials completed and analysed up to the end of December 1992 (the Lilly fluoxetine database) that satisfied the selection criteria were included. After this date, no pertinent trials comparing fluoxetine with placebo or TCAs were added to the database. We analysed the trials performed in the USA (USA trials) separately from those performed elsewhere (Canada and Europe; non-USA trials) because the psychiatric methods and clinical trial procedures were sufficiently different in the USA compared with elsewhere, and this could be a source of heterogeneity between the trials (Ansseau, 1992).

In our protocol for this meta-analysis we defined the criteria for selecting trials, before we accessed the trials database: (a) identical or very similar clinical inclusion criteria for patients (major depression as defined by DSM-III (American Psychiatric Association, 1980)); (b) use of the Hamilton Depression Rating Scale (HDRS-17; Hamilton, 1967; and the first 17 items from trials that used more than 17); and (c) a double-blind follow-up phase of at least six weeks. For the non-USA trials, we analysed only trials of fluoxetine v. TCAs since the three non-USA placebo-controlled trials (116 patients) in the database did not satisfy our inclusion criteria or included very few patients. The same inclusion criteria were used for non-USA trials, except that trials with a five-week, double-blind follow-up period were also included since their exclusion would have led to only a handful of trials with a small number of patients being included. The database contained only one USA trial with a five-week double-blind follow-up period, but this was not included.

Trials without a control treatment (e.g. dose-ranging trials) and those with a control treatment other than placebo or a TCA were excluded. In addition, trials in which all control patients received fixed doses ≤ 75 mg/day of a TCA were eliminated, as were those in which treated patients received less than 10 mg/day of fluoxetine. Within a trial, all patients were pooled according to the treatment received, irrespective of the dose received, this being equivalent to comparing a single fluoxetine-treated group with a single TCA-treated group and a single placebo-treated group.

The first evidence-based diagnostic system in psychiatry is the DSM-III. New-generation antidepressants are indicated for major depression as defined using this diagnostic system in most countries, and this is the reason we decided to use DSM-III major depression as the only diagnostic inclusion criterion.

The database contained information for 69 trials, including 6633 patients; of these, 21 trials were USA trials and 48 had been performed elsewhere (non-USA trials). Of the 21 USA trials, five including 400 patients were excluded for the following reasons: three because of the diagnostic system used (Research Diagnostic Criteria; RDC; Spitzer *et al.*, 1978); one because the double-blind follow-up was only for five weeks; and one because the TCA dose was too low. In addition 96 patients

randomised to receive a fixed dose of 5 mg/day were excluded, as per our protocol. Of the 48 non-USA trials, 34 trials including 2047 patients were excluded for the following reasons: 23 because the DSM-III was not used (RDC; Feighner diagnostic criteria; ICD-9; World Health Organization, 1978); five because the control treatment was not a TCA (maprotiline, a monoamine reuptake inhibitor, was considered to be similar to TCAs although it is not tetracyclic, but mianserin was not); two because they were open-label uncontrolled trials; two because only one and two patients, respectively, had been recruited, one because a fixed dose of a TCA was used (clomipramine 75 mg); and one because there were only sparse data available for the 11 patients included. In total, 30 trials (16 USA and 14 non-USA) and 4120 patients (3447 USA and 673 non-USA) were included (62% of the total database) in accordance with the criteria defined in our protocol.

ANALYSIS GROUPS AND METHODS

The analyses of continuous data were performed by M.B. and confirmed by M.H. using the Cochrane Collaboration software (Review Manager, 1997). The analyses of the binary outcomes were performed by P.C. and M.H. using a specific software package (EasyMA; Cucherat *et al*, 1997).

The trials were analysed in groups defined by where they were performed (USA and non-USA trials) and type of control treatment (placebo or TCA). Three types of analyses were performed for each outcome: (a) all randomised patients, classifying prematurely discontinued patients (before Day 42 in USA trials and Day 35 in non-USA trials) as failures (intention-to-treat); (b) all randomised patients who completed at least four weeks of therapy using "a last-observation-carried-forward" technique (efficacy analysis); and (c) all randomised patients with at least one post-baseline visit (end-point analysis) using "a last-observation-carried-forward" technique.

Outcomes

Frank *et al* (1991) suggested using the term remission, rather than recovery, when defining response to drug therapy in the short-term treatment of depression. Partial remission after 4–6 weeks of treatment can be defined as at least a 50% reduction

compared with the baseline value for the HDRS-17 score, which corresponds to very much or much improved on the Clinical Global Impression Scale (CGI) (Guy, 1976). The CGI was used in all the USA trials, but only in a few of the non-USA trials. The primary outcome for USA and non-USA trials was defined as a binary variable on the HDRS-17; partial remission, that is at least 50% reduction compared with the baseline score on the HDRS-17 instrument. The secondary outcome in the USA trials was also a binary variable, defined as a much improved or very much improved on the CGI scale. Another secondary, but quantitative, outcome was the mean change in HDRS-17 scores from baseline to end-point. In this part of the analysis an HDRS subscale, the depression factor (including the six items of depressed mood, guilt, work and interests, retardation, psychic anxiety and general somatic), was also used (HDRS-6; Bech, 1989; O'Sullivan *et al*, 1997).

The reasons for early treatment discontinuation were analysed as binary variables (adverse event, lack of efficacy or any reason).

Meta-analytical methods

Log odds ratio analysis for binary data

We used the logarithm of the odds ratio method, which is based on a multiplicative model, that is the success rate (partial remission) in the treatment group is assumed to be a multiplicative function of that in the control group (Boissel *et al*, 1989). Due to the large number of statistical tests performed the level of statistical significance was set at a robust value $P=0.01$ or less. A test for heterogeneity was also performed, and because this is an insensitive test, the level of statistical significance was set at a value of $P=0.10$ or less. When heterogeneity was detected we analysed the data using a random effects model, which gives more conservative results, but can deal with a certain amount of heterogeneity.

An odds ratio equal to one indicates that there is no difference between the two treatment groups. A value greater than one indicates that more patients in the fluoxetine group were classified as being in partial remission, and therefore that fluoxetine was better; a value of less than one indicates that more patients in the control group were classified as being in partial remission, and therefore that control treatment (placebo or TCA) was better.

However, in the analyses of early treatment discontinuations an odds ratio of less than one indicates fewer discontinuations in the fluoxetine group, and that fluoxetine was better. Conversely, a log odds ratio of greater than one indicates that there were fewer discontinuations in the control (placebo or TCA) group, and therefore that the control treatment was better.

Effect size for the meta-analysis of quantitative data

Effect size analysis was introduced by Glass (1976) as a means of combining data from several independent clinical trials. In our analysis the effect size was defined as the mean change of HDRS from baseline to end-point of the two groups under investigation divided by the standard deviation of the change score (Cohen, 1977). The 95% confidence intervals (95% CIs) were calculated according to Hedges & Olkin (1985). Data for all randomised patients with at least one post-baseline visit (end-point analysis), using a 'last-observation-carried-forward' technique, were included in these analyses. The method of calculation used is in accordance with that described by Whitehead & Whitehead (1991), using either a fixed or random effects model as deemed appropriate. As for the meta-analysis of binary data, a test of heterogeneity (Cochran's Q-test; Laird & DerSimonian, 1986) and a test of significance of the effect size were performed.

RESULTS

General results

The list of trials showing some details of their characteristics, for example, number of investigators, number of patients, dose of medication, are given in Tables 1a and 1b (further details and references of published trials can be obtained from the authors upon request). One trial (non-USA-10) was excluded after the initial analyses which showed that the trial was responsible for a statistically significant heterogeneity, and inspection of the results suggested that they were unlike the others (i.e., partial HDRS-17 response rates were 86.7% and 6.7% for the fluoxetine-treated and TCA-treated groups, respectively). For the USA-trial analysis, data were analysed from 16 single- and multicentre, randomised, double-blind trials involving 3543 patients (see Table 1a). Of the 3447 patients, 1914 had received fluoxetine, 847 had

Table 1a Description of the 16 United States (USA) clinical trials of fluoxetine v. placebo or tricyclic antidepressant (TCA) in major depression (DSM-III)

Variables	USA 1	USA 2	USA 3	USA 4	USA 5	USA 6	USA 7	USA 8	USA 9	USA 10	USA 11	USA 12	USA 13	USA 14	USA 15	USA 16
No. of investigators	1	10	10	30	6	4	3	2	3	1	1	3	1	2	7	1
No. of patients																
Total	40	746	363	671	89	118	109	159	130	64	61	58	30	88	728	89
Fluoxetine	21	639	285	335	46	56	55	79	65	32	31	28	15	46	247	30
TCA	-	-	-	-	-	62	54	80	65	32	30	30	15	42	246	30
Placebo	19	107	78	336	43	-	-	-	-	-	-	-	-	-	235	29
Status (in-patient/out-patient)	In/Out	Out	Out	Out	Out	In/Out	Out	Out	Out	Out	Out	In/Out	Out	Out	Out	Out
Fluoxetine	20-60	20, 40 or 60	5, 20 or 40	20 fixed	20 fixed	20-80	20-80	20-80	20-60	20-60	20-60	20-60	20-60	20-60	20-80	20-80
TCA	-	-	-	-	-	IP	D	D	AT	D	D	DMI	NT	DMI	IP	IP
Dosage range (mg)	-	-	-	-	-	75-300	75-300	75-200	50-200	50-150	75-200	50-300	50-150	50-300	75-300	75-300
Published (P)/unpublished (U)	P	P	P	P	P	P	U	P	P	U	U	U	U	U	P	P
Completers (n, %)																
Fluoxetine	15, 71.4	356, 55.7	180, 63.2	263, 78.5	33, 71.7	23, 41.1	38, 69.1	41, 51.9	45, 69.2	23, 71.9	14, 45.2	23, 82.1	13, 86.7	41, 89.1	136, 55.1	19, 63.3
TCA	-	-	-	-	-	24, 38.7	38, 70.4	31, 38.7	30, 46.1	20, 62.5	6, 20.0	22, 73.3	9, 60.0	25, 59.5	124, 50.4	14, 46.7
Placebo	15, 78.9	64, 59.8	42, 53.8	271, 80.6	33, 76.7	-	-	-	-	-	-	-	-	-	92, 39.1	9, 31.0

AT, amitriptyline; IP, imipramine; D, doxepin; DMI, desipramine; NT, nortryptiline.

Table 1b Description of the 14 non-USA trials (Europe and elsewhere) of fluoxetine v. tricyclic antidepressant (TCA) in major depression (DSM-III)

Variables	Non-USA	Non-USA	Non-USA	Non-USA	Non-USA	Non-USA	Non-USA	Non-USA	Non-USA	Non-USA	Non-USA	Non-USA	Non-USA	Non-USA
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
No. of investigators	4	2	5	3	1	1	14	1	1	1	1	1	1	> 1
No. of patients														
Total	63	59	63	58	30	76	42	75	27	30	26	32	30	62
Fluoxetine	29	29	33	28	15	38	21	37	13	15	11	15	15	30
TCA	34	30	30	30	15	38	21	38	14	15	15	17	15	32
Status (in-patient/out-patient)	Out	Out	In/Out	In/Out	In	In/Out	In/Out	Not-specified	Out	In	In	In/Out	In/Out	In
Fluoxetine dosage range (mg)	40-80	20	40-80	20-40	20-60	20-60	40-80	20-80	20-60	20-60	20-60	20	20	20
TCA	M	IP	DT	DT	CI	D	AT	AT	IP	IP	AT	IP	IP	CI
Dosage range (mg)	50-150	75-150	50-225	100-200	50-175	50-200	100-250	75-300	50-175	50-175	50-175	50-175	50-175	50-200
Published (P)/unpublished (U)	U	P	P	U	P	U	U	U	U	U	U	U	P	P
Completers (n, %)														
Fluoxetine	28, 96.5	22, 75.9	18, 54.5	16, 57.1	14, 93.3	25, 65.8	20, 95.2	24, 64.9	11, 84.6	15, 100.0	9, 81.8	15, 100.0	15, 100.0	18, 60.0
TCA	30, 88.2	24, 80.0	21, 70.0	23, 76.7	14, 93.3	26, 68.4	16, 76.2	28, 73.7	12, 85.7	15, 100.0	12, 80.0	10, 58.8	12, 80.0	19, 59.4

AT, amitriptyline; IP, imipramine; D, doxepin; DT, dothiepin; CI, clomipramine; M, maprotiline.

received placebo and 686 had received TCAs (either amitriptyline, desipramine, doxepin, imipramine, or nortriptyline). A total of 96 patients were excluded, per our protocol, because they had been randomised to receive a fixed dose of 5 mg/day of fluoxetine. Therefore, data for a total of 3447 patients were included in the meta-analyses (1914 in the fluoxetine-treated group, 847 in the placebo-treated group and 686 in the TCA-treated group).

For the non-USA trials, data were analysed from 13 single- and multi-centre, randomised, double-blind trials in which fluoxetine was compared with a TCA in 643 patients (i.e. without non-USA trial 10; see Table 1b). Of these 643 patients, 314 had received fluoxetine and 329 had received TCAs (either amitriptyline, clomipramine, dothiepin, doxepin, imipramine or maprotiline). There were no statistically significant differences between the treatment groups in the percentage of men included in the trials (approximately 40% overall), the mean age (approximately 45 years), or the baseline HDRS-17 score (total mean score approximately 22). Only one of the USA trials *v.* placebo included both in- and out-patients, the others included only out-patients.

The dose ranges for the individual trials are shown in Tables 1a and 1b. Only two of the USA trials *v.* TCA included both in- and out-patients (both started with only in-patients and the protocols were amended during the trials); the other trials included only out-patients. Three non-USA trials *v.* TCA included only in-patients, three included only out-patients, six included both, and this was not specified for the remaining trial. The percentage of patients completing the trial was generally higher in the non-USA trials (Table 1b) than in the USA trials (Table 1a).

Meta-analysis of binary data for treatment effects

HDRS-17

Table 2 shows the results obtained with HDRS-17, using both the percentage of responders and odds ratio analysis. The efficacy analysis had the highest response rates in the comparisons. The overall difference for fluoxetine *v.* placebo was 21.4% in the efficacy analysis but only 13.6% in the intention-to-treat analysis. In all the analyses fluoxetine showed a statistically significant benefit compared with placebo. In the USA trials of fluoxetine *v.* TCA no

Table 2 Meta-analysis of patients classified as responders on the Hamilton Depression Rating Scale (HDRS-17). An odds ratio equal to 1 means no difference between treatment groups. A value greater than 1 favours fluoxetine while a value less than 1 favours control treatment

Trials	Percentage of responders on HDRS-17			Log odds ratio analysis	
	Fluoxetine	Placebo	Difference	Odds ratio	95% CI
USA trials v. placebo					
Intention to treat	37.8	24.2	13.6	2.07*	(1.69-2.53)
Efficacy	56.0	34.6	21.4	2.44*	(1.97-3.04)
End-point	44.5	27.7	16.8	2.22*	(1.83-2.70)
USA trials v. TCA					
Intention-to-treat	39.5	35.6	3.9	1.18	(0.94-1.48)
Efficacy	58.2	60.6	-2.4	0.93	(0.71-1.22)
End-point	45.5	43.0	2.5	1.10	(0.88-1.37)
Non-USA trials v. TCA					
Intention to treat	39.5	44.1	-4.6	0.81	(0.58-1.13)
Efficacy	56.3	65.9	-9.6	0.62	(0.42-0.92)
End-point	46.2	52.9	-6.7	0.73	(0.53-1.02)
All studies v. TCA					
Intention to treat	39.5	38.3	1.2	1.05	(0.87-1.26)
Efficacy	57.5	62.5	-5.0	0.82	(0.65-1.02)
End-point	45.7	46.3	-0.6	0.97	(0.81-1.16)

*P ≤ 0.01; TCA, tricyclic antidepressant.

statistically significant differences were observed. In the non-USA trials no statistically significant differences were observed.

CGI

Table 3 shows the results for the CGI outcome, both the remission rates (percentage of 'very much improved' and 'much improved') and the odds ratio analyses. The analyses for the fluoxetine *v.* placebo trials gave results that were similar to those obtained for HDRS-17 outcome, that is, all differences were statistically significant.

Meta-analysis of quantitative data (effect size)

When the results for all seven trials assessing fluoxetine *v.* placebo are pooled an effect size of -0.30 in favour of fluoxetine was obtained, with a 95% CI of -0.39 to -0.21 (see Fig. 1). For the HDRS-6 outcome an effect size of -0.37 was observed (95% CI: -0.46 to -0.28). Figure 2 shows the results for the trials *v.* TCAs. The pooled effect size for the HDRS-17 outcome in the USA trials was 0.00 with a 95% CI of -0.18 to 0.10. The pooled

Table 3 Meta-analysis of patients classified as 'very much' or 'much' improved on the Clinical Global Impression Scale (CGI)

Trials	Percentage of responders on CGI			Log odds ratio analysis	
	Fluoxetine	Placebo	Difference	Odds ratio	95% CI
USA trials v. placebo					
Intention to treat	43.9	29.6	14.3	1.89*	(1.57-2.29)
Efficacy	65.6	41.3	24.3	2.40*	(1.94-2.98)
End-point	52.7	32.9	19.8	2.20*	(1.83-2.66)
USA trials v. TCA					
Intention to treat	45.9	38.3	7.6	1.38	(1.10-1.72)
Efficacy	67.3	66.4	0.9	1.06	(0.79-1.41)
End-point	53.3	49.1	4.2	1.18	(0.95-1.47)

*P ≤ 0.01; TCA, tricyclic antidepressant.

effect size for the HDRS-6 outcome showed a non-significant trend in favour of fluoxetine, (-0.10 ; 95% CI -0.21 to 0.01). A trend in favour of TCAs was observed for the non-USA trials *v.* TCAs, with a pooled effect size for the HDRS-17 outcome of 0.17 (95% CI 0.01 to 0.34). There was a stronger trend in favour of TCAs for the HDRS-6 outcome, with a pooled effect size of 0.18 (95% CI 0.01 to 0.34). When the results from all the trials comparing fluoxetine *v.* TCAs were pooled the effect size for the HDRS-17 outcome showed a non-significant trend in favour of TCAs (0.05 ; 95% CI -0.04 to 0.14). The pooled effect size for the HDRS-6 outcome also showed a non-significant trend in favour of fluoxetine (-0.02 ; 95% CI -0.11 to 0.07).

Meta-analysis of early treatment discontinuation data (binary)

The results of the analyses of the reasons for discontinuations in the trials *v.* placebo were as predicted, that is significantly more discontinuations in the fluoxetine-treated group due to an adverse event, and significantly more discontinuations in the placebo-treated group due to lack of efficacy, with a non-significant trend for discontinuation for any reason favouring fluoxetine (see Table 4). Using the fixed effects model the test for homogeneity was significant indicating heterogeneity among the trials for the three outcomes, and visual inspection of the graphical results (not shown) suggested this was due to two trials (USA-trial-15 and USA-trial-16). We therefore decided to use a random effects model, which gave more conservative results, but removed the heterogeneity.

The analysis of the reasons for discontinuation in the USA trials of fluoxetine *v.* TCA showed that, while receiving fluoxetine, significantly fewer patients discontinued

their treatment because of an adverse event, and significantly fewer patients discontinued for any reason. No significant difference was seen with respect to discontinuations due to lack of efficacy. The results from a similar analysis for the non-USA trials *v.* TCA did not indicate any significant differences between the two groups, however, the width of the confidence intervals suggest a potential lack of power to detect clinically significant differences. When the USA and non-USA trials were combined the results showed that significantly fewer patients on fluoxetine discontinued treatment due to adverse events or for any reason.

DISCUSSION

DSM-III major depression

In our protocol for this meta-analysis it was our intention to compare USA trials with those performed elsewhere (non-USA trials). An *a priori* condition for such a comparison required that the diagnostic system should be evidence-based and accepted by the health care regulators in both the USA and elsewhere. More non-USA trials than USA trials were excluded because the diagnosis of depression in Europe was made using a classification system other than the DSM-III criteria. We had not anticipated that this would lead to 23 non-USA trials (including around 1400 patients) being excluded. However, since the official indication for the use of SSRIs such as fluoxetine in patients with depression worldwide, including Europe, is major depression, we felt that it was not justified to change the original inclusion criteria in our protocol.

Antidepressive responsiveness to fluoxetine in major depression

A 50% reduction in the baseline HDRS-17 score was the primary outcome in our

study. In the intention-to-treat analysis, both for HDRS-17 and for CGI, fluoxetine showed an advantage of approximately 15% over placebo. This is a similar result to that found in one of the first overviews comparing TCAs with placebo (Smith *et al.*, 1969) as well as that reported in the Medical Research Council trial (Medical Research Council, 1965). The odds ratio analysis confirmed that fluoxetine was significantly superior to placebo, although no difference was seen between the USA trials and non-USA trials.

Improved safety acceptance of fluoxetine

In this meta-analysis the results for discontinuation due to adverse reactions were evaluated by the intention-to-treat analysis. Compared with placebo we observed that significantly more patients ceased treatment with fluoxetine due to adverse events while significantly more patients dropped out on placebo due to lack of efficiency. This is reflected in the relatively lower differences in antidepressive improvement in the intention-to-treat analysis. However, compared with patients in the TCA groups, in the USA trials, and to a lesser extent the non-USA trials, we observed that significantly fewer trials in the fluoxetine group stopped treatment due to adverse events. This seems to explain that the intention-to-treat analysis for the USA trials favoured fluoxetine while that for the non-USA trials did not. However, when combined, significantly fewer patients on fluoxetine compared with those on TCAs discontinued treatment due to adverse events.

These results are in agreement with results of the meta-analyses published by Andersen & Tomenson (1995) and Hotopf *et al.* (1997). In the latter meta-analysis, Hotopf *et al.* analysed the 'old' TCAs (e.g.

Table 4 Meta-analysis of reasons for early treatment discontinuation

Analysis subgroup	Reason for early treatment discontinuation					
	Adverse event		Lack of efficacy		Any reason	
	Odds ratio	95% CI	Odds ratio	95% CI	Odds ratio	95% CI
USA trials <i>v.</i> placebo	1.96*	(1.42–2.72)	0.38*	(0.29–0.50)	0.84	(0.69–1.02)
USA trials <i>v.</i> TCA	0.47*	(0.36–0.61)	1.27	(0.88–1.82)	0.64*	(0.52–0.80)
Non-USA trials <i>v.</i> TCA	1.25	(0.64–2.46)	1.06	(0.54–2.10)	1.16	(0.78–1.70)
Combined trials <i>v.</i> TCA	0.53*	(0.42–0.67)	1.22	(0.89–1.62)	0.75*	(0.62–0.90)

* $P \leq 0.01$; TCA, tricyclic antidepressant.

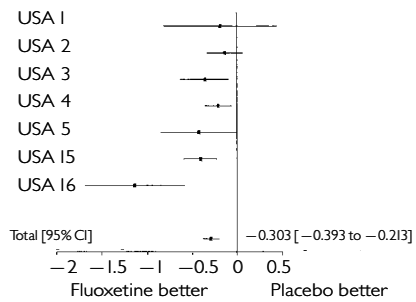


Fig. 1 Effect size analysis for the actual change from baseline to end-point on the 17 item Hamilton Depression Rating Scale in seven USA trials v. placebo (all randomised patients analysed, using a last observation carried forward technique). The horizontal lines represent the result for each trial and the global estimate is given at the bottom of a graph.

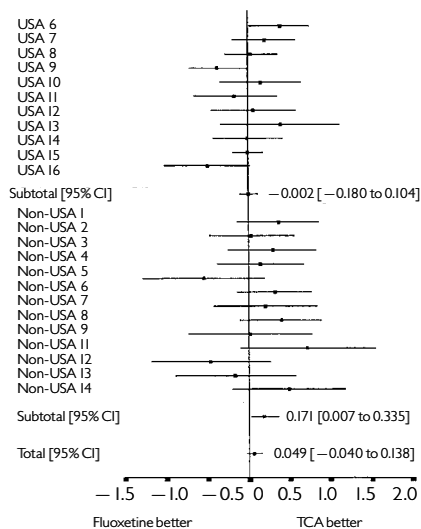


Fig. 2 Effect size analysis for the actual change from baseline to end-point on the 17 item Hamilton Depression Rating Scale scale in 11 USA trials v. tricyclic antidepressant (TCA) and 13 non-USA trials v. TCA (all randomised patients analysed, using a last observation carried forward technique). The horizontal lines represent the result for each trial and the global estimate is given at the bottom of the graph.

imipramine and amitriptyline) separately from the ‘newer’ TCAs (e.g. dothiepin, nortriptyline, clomipramine and doxepin). They found that the lower rate of discontinuation in patients on SSRIs was observed in the comparison with the old TCAs. This may explain our finding concerning the intention-to-treat analysis in the USA v.

CLINICAL IMPLICATIONS

- The different statistical analyses converged in showing that fluoxetine is significantly superior to placebo and equal to tricyclic antidepressants (TCAs).
- In clinical terms fluoxetine had an 15–20% improvement advantage to placebo which was maintained in the core symptoms of depression on the Hamilton Rating Scale.
- The discontinuation rate of fluoxetine due to adverse drug events was significantly lower than with TCAs.

LIMITATIONS

- The meta-analysis criterion of DSM–III major depression excluded a rather high proportion of the European trials.
- The trials were insufficient for evaluating the relationship between dose of fluoxetine and clinical response.
- The USA trials have used older reference TCAs whereas the European trials used newer TCAs.

P. BECH, MD, Frederiksborg General Hospital, Hilleroed, Denmark; P. CIALDELLA, MD, M. C. HAUGH, PhD, A. HOURS, MD, J. P. BOISSEL, MD, Service de Pharmacologie Clinique, Lyon, France; M. A. BIRKETT, BSc, G. D. TOLLEFSON, PhD, Psychopharmacology Clinical Research, Eli Lilly and Co, Indianapolis, Indiana, USA

Correspondence: Dr P. Bech, Psychiatric Research Unit, Frederiksborg General Hospital, Dyrehavevej 48, DK–3400 Hilleroed, Denmark

(First received 22 June 1998, final revision 18 October 1999, accepted 18 October 1999)

non-USA trials, as the old TCAs were used in 65% of the USA trials compared with 47% of the non-USA trials.

Comparison with other meta-analyses with fluoxetine

In previous meta-analyses the effect size was mainly used, for example, Song *et al* (1993), Greenberg *et al* (1994) or Anderson & Tomenson (1994). Our results for fluoxetine v. placebo are in agreement with Greenberg *et al* (1994), although our effect size of -0.30 for the HDRS–17 remission outcome is low. In the Greenberg *et al* (1994) analysis we have detected some publication bias (i.e. unpublished trials not included) and double publication (i.e. data included from two publications of the same trial). When using the core symptoms of depression, the HDRS–6 outcome, we showed

an effect size of -0.37 , indicating that fluoxetine has an effect on the specific symptoms for major depression. This is in agreement with the results from our previous meta-analyses on citalopram and fluvoxamine (Bech, 1989; Bech & Cialdella, 1992). Our results for fluoxetine v. TCAs are in agreement with those published by Anderson & Tomenson (1994), that is, there is no difference in the antidepressive effect. This was confirmed by the HDRS–6 outcome results.

In conclusion, we have shown that results from meta-analyses can differ depending on how patients who withdraw from treatment early are counted in the analyses. Generally, the approach used is intention-to-treat, whereby patients who withdraw from treatment early are considered failures in the trial group to which they were allocated, and it may be important in the future to consider using other approaches (efficacy

and end-point) in meta-analyses, to determine if there is a difference. Thus, in our analyses, we have confirmed the superiority of fluoxetine over placebo for the short-term treatment of major depression, and although we were unable to show a difference in efficacy with TCAs, fewer patients on fluoxetine withdrew due to adverse effects.

REFERENCES

- American Psychiatric Association (1980)** *Diagnostic and Statistical Manual of Mental Disorders* (3rd edn) (DSM-III). Washington, DC: APA.
- Anderson, I. M. (1998)** SSRIs versus tricyclic antidepressants in depressed in-patients: A meta-analysis of efficacy and tolerability. *Depression and Anxiety* **7** (suppl. 1), 11–17.
- & **Tomenson, B. M. (1994)** The efficacy of selective serotonin re-uptake inhibitors in depression: a meta-analysis of studies against tricyclic antidepressants. *Journal of Psychopharmacology*, **8**, 238–249.
- & — (1995) Treatment discontinuation with selective serotonin reuptake inhibitors compared with tricyclic antidepressants: A meta-analysis. *British Medical Journal*, **310**, 1433–1438.
- Ansseau, M. (1992)** The Atlantic gap: clinical trials in Europe and the United States. *Biological Psychiatry*, **31**, 109–111.
- Bech, P. (1989)** Clinical effects of selective serotonin reuptake inhibitors. In *Clinical Pharmacology in Psychiatry* (Psychopharmacology. Series 7) (eds S. G. Dahl & L. F. Gram), pp. 81–93. Berlin & Heidelberg: Springer-Verlag.
- & **Cialdella, P. (1992)** Citalopram in depression: meta-analysis of intended and unintended effects. *International Clinical Psychopharmacology*, **6** (suppl. 5), 45–54.
- Boissel, J. R., Blanchard, J., Panak, E., et al (1989)** Considerations for the meta-analysis of randomized clinical trials. Summary of a panel discussion. *Controlled Clinical Trials*, **10**, 254–281.
- Cohen, J. (1977)** *Statistical Power Analysis for the Behavioral Sciences*. Orlando, FL: Academic Press Inc.
- Cucherat, M., Boissel, J. P., Leizorovicz, A., et al (1997)** EasyMA: a program for the meta-analysis of clinical trials. *Computer Methods and Programs in Biomedicine*, **53**, 187–190.
- Frank, E., Prien, R. F., Jarrett, R. B., et al (1991)** Conceptualisation and rationale for consensus definitions of terms in major depressive disorders. *Archives of General Psychiatry*, **48**, 851–855.
- Glass, G. V. (1976)** Primary, secondary and meta-analysis of research. *Review of Educational Research*, **5**, 3–9.
- Greenberg, R. P., Bornstein, R. F., Zborowski, M. J., et al (1994)** A meta-analysis of fluoxetine outcome in the treatment of depression. *Journal of Nervous and Mental Disease*, **182**, 547–551.
- Guy, W. (1976)** *ECDEU Assessments Manual for Psychopharmacology*. Rockville, MD: National Institute of Mental Health.
- Hamilton, M. (1967)** Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology*, **6**, 278–296.
- Hedges, L. V. & Olkin, I. (1985)** *Statistical Methods for Meta-Analysis*. New York: Academic Press.
- Hotopf, M., Hardy, R. & Lewis, G. (1997)** Discontinuation rates of SSRIs and tricyclic antidepressants: a meta-analysis and investigation of heterogeneity. *British Journal of Psychiatry*, **170**, 120–127.
- Laird, N. & DerSimonian, R. (1986)** Meta-analysis in clinical trials. *Controlled Clinical Trials*, **7**, 177–188.
- Medical Research Council (1965)** Clinical trial treatment of depressive illness. *British Medical Journal*, **1**, 881–886.
- O'Sullivan, R. L., Fava, M., Agustín, C., et al (1997)** Sensitivity of the six-item Hamilton Depression Rating Scale. *Acta Psychiatrica Scandinavica*, **95**, 379–384.
- Review Manager (1997)** Computer Programme Version 3.0.1. Oxford: Update Software.
- Smith, A., Traganza, E. & Harrison, G. (1969)** Studies on the effectiveness of antidepressant drugs. *Psychopharmacology Bulletin*, suppl., 1–53.
- Song, F., Freemantle, N., Sheldon, T. A., et al (1993)** Selective serotonin reuptake inhibitors: a meta-analysis of efficacy and acceptability. *British Medical Journal*, **306**, 683–687.
- Spitzer, R. L., Endicott, J. & Robins, E. (1978)** Research Diagnostic Criteria: Rationale and research reliability. *Archives of General Psychiatry*, **35**, 773–785.
- Whitehead, A. & Whitehead, J. (1991)** A general parametric approach to the meta-analysis of randomized clinical trials. *Statistics in Medicine*, **10**, 1665–1677.
- World Health Organization (1978)** *Mental Disorders: Glossary and Guide to their Classification in Accordance with the Ninth Revision of the International Classification of Diseases (ICD-9)*. Geneva: WHO.