

Short-term efficacy of tricyclic antidepressants revisited: a meta-analytic study

Jitschak G. Storum^{a,b,*}, André J.A. Elferink^a, Barbara J. van Zwieten^a, Wim van den Brink^b, Berthold P.R. Gersons^b, Roel van Strik^a, André W. Broekmans^{a,c}

^aMedicines Evaluation Board of the Netherlands, Kalvermarkt 53, P.O. Box 16229, 2500 BE Den Haag, The Netherlands

^bPsychiatric Department of the Academic Medical Centre, University of Amsterdam, Tafelbergweg 25, 1105 BC Amsterdam, The Netherlands

^cDepartment of Pharmaco-epidemiology and Pharmacotherapy, University of Utrecht, Utrecht, The Netherlands

Received 4 December 2000; accepted 15 February 2001

Abstract

The original data from the placebo-arms and the tricyclic-arms of all parallel randomized controlled three-arm studies, which had been conducted in the period 1979–1991 for a drug under development in order to obtain marketing authorization for the indication major depression, were included in a meta-analysis. Thirty-two placebo-controlled studies including 4314 patients were analyzed. The intention to treat analysis resulted in 46% responders (at least 50% improvement on the Hamilton Depression Rating Scale) in the tricyclic antidepressant group and 31% in the placebo-group ($CI_{95\% \text{-difference}}$ 11.5–17.1%). The number needed to treat for responders was 7 ($CI_{95\%}$ 5–8). In 10 out of 32 studies, a statistically significant difference in favor of tricyclic antidepressant compared to placebo was found for responders. The responder rate in the placebo-group varied from 6 to 52%. We conclude that tricyclic antidepressants are efficacious in the short-term treatment of major depression. However, the magnitude of the effect is rather modest. Because 69% of the placebo-controlled studies with a tricyclic antidepressant did not show a statistically significant difference in favor of tricyclic antidepressant and the placebo rate varied considerably from study to study, equivalence studies with tricyclic antidepressant as comparator without a placebo-control are not sufficient for demonstrating efficacy. Therefore in major depression, placebo-controlled studies are still necessary to demonstrate efficacy. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Major depression; Meta-analysis; Placebo; Tricyclic antidepressants

1. Introduction

Kuhn (1958) tested the effect of imipramine in psychotic patients. It was found ineffective in schizophrenia but it made depressed schizophrenics less depressed. Since then, many imipramine-like drugs, the so-called tricyclic antidepressants, have been shown to be effective in depressive patients and tricyclic antidepressants became the gold standard in the treatment of major depression. However, an accurate estimation of the magnitude of the effect could not be determined as many of the studies were conducted

at a time when the development of proper study design was in its infancy (Klein et al., 1980).

In the 1980s and 1990s new antidepressants were developed and evaluated in newly developed study designs. Most of the pivotal efficacy studies conducted with these new compounds were randomized double-blind studies with both a negative control (placebo) and a positive control (tricyclic antidepressant). These studies were part of registration dossiers that pharmaceutical companies worldwide submitted to regulatory authorities to obtain a marketing authorization.

These studies can result in four outcomes. Both investigational drug and TCA are superior to placebo. This is considered a positive study. A second possibility is that TCA is superior to placebo and the investigational drug is not. This is a negative study for the investigational drug. A

*Corresponding author. Tel.: +31-70-356-7400; fax: +31-70-356-7515.

E-mail address: jg.storum@cbg-meb.nl (J.G. Storum).

third possibility is that both compounds do not show a statistically significant difference in favour of placebo. This result is not considered a negative study for the investigational drug but failed study as the gold standard (TCA) is in that case not superior to placebo. The fourth possibility is that the investigational drug is superior to placebo and TCA not. This result often gives much discussion concerning among others, the patients included in the study and/or the dose of TCA used in the study. As pharmaceutical companies like to satisfy regulatory requirements, studies are conducted with the objective of a positive result both for investigational drug and TCA (positive study). All other results are not desirable for pharmaceutical companies as they can lead to a delay in granting a licence. Therefore the efficacy results of these studies are very informative concerning the magnitude of the effect of TCAs.

Recently Kahn et al. (2000) used the public domain data on FDA-reviewed studies to assess mean improvement in investigational antidepressants, active comparators (mostly tricyclic antidepressants) and placebo. Their original aims, to establish rates of therapeutic response, and to evaluate the effect size between placebo and antidepressants expressed as confidence intervals, could not be reached, because results were not available on a week-by-week basis, and not for individual patients.

Here, we report a meta-analysis based on a database that included individual patient data on a week-by-week basis. The aim of this meta-analysis is to provide an accurate estimation of the magnitude of the effect of tricyclic antidepressants and to determine the clinical relevance of tricyclic antidepressants in the short-term treatment of major depression. In addition, this study intends to evaluate whether it will still be necessary in future studies to use placebo controls for demonstration of efficacy of new drugs.

2. Experimental procedures

2.1. Selection of studies

The pharmaceutical companies concerned (AHP, Bristol, Duphar, Lilly, Lundbeck, Organon, Pfizer, Roche, SKB), were asked to make available all data from placebo-arms and tricyclic-arms of all three-arm studies which had been conducted for a drug under development. These studies were part of registration dossiers which had been submitted to the Medicine Evaluation Board (MEB) in the period from the first selective serotonin re-uptake inhibitor to the most recent compound to be granted a license. The completeness was verified with the descriptions of the studies in the assessment reports of the MEB. In these studies the patients included were diagnosed as having major depression according to the diagnostic criteria from the Diagnostic and Statistical Manual III (R) criteria

(American Psychiatric Association, 1980, 1987) and the Hamilton Depression Rating Scale (Hamilton, 1960) was used as outcome measure for efficacy.

2.2. Medicines Evaluation Board

The Medicines Evaluation Board of the Netherlands (MEB) is the regulatory authority of the Netherlands. To obtain a marketing authorization, a company must submit a dossier to the MEB with all clinical trials conducted for a drug under development. On the basis of the assessment report on this dossier, the MEB takes a decision with regard to the marketing authorization application and, in the case of a positive opinion, finalizes the product information texts.

2.3. Hamilton Depression Rating Scale

The Hamilton Depression Rating Scale is a standard measure that is included in most studies of depression (Hamilton, 1960). Besides the original 17-item version there is among others, a 21-item version of this scale. Various cut-off points have become standard for determining levels of severity. For the 17-item version the cut-off score for 'moderately depressed' lies between 18 and 24 points and for 'severely depressed' ≥ 25 points (Endicott et al., 1981). Response is defined as at least 50% improvement on the Hamilton Depression Rating Scale from baseline (Montgomery, 1995).

2.4. Statistical analysis

The primary outcome variable was response. Secondary outcome variable was the mean improvement from baseline on the Hamilton Depression Rating Scale.

Analyses were based on the intention-to-treat population and on the completer population. The intention to treat analysis included all patients who were randomized to receive double-blind treatment, had taken at least one dose, and had at least one efficacy evaluation. Intermediate missing values were interpolated by using the last observation. The last observation was carried forward (LOCF) for patients who had dropped out before the end of the study. The completer population was defined as all patients who were randomized and had completed the double-blind period. For studies longer than 6 weeks the cut-off point was taken at 6 weeks.

The Hamilton Depression Rating Scale 17-item version was used for the meta-analysis. In studies where the Hamilton Depression Rating Scale 21-item version was used this version was converted to the Hamilton Depression Rating Scale₁₇. Only when this was not possible was the 21-item scale used.

Responder rates were analysed in terms of risk differences. Test statistics were based on the mean differences between tricyclic antidepressant and placebo in terms of

improvement of the Hamilton Depression Rating Scale from baseline. The overall estimates and confidence intervals were based on a fixed model, weighted for variance differences between the studies.

3. Results

3.1. Studies

All but one company gave permission to use their study data. This resulted in 32 studies which were included in the meta-analysis (Table 1). The studies were started between

1979 and 1991 and approval for licensing was granted between 1985 and 1997. All studies were conducted in the USA. Out-patients were included in all but one study. In 27 studies imipramine was used as positive control and amitryptiline in five studies. The duration of the double-blind period was 6 weeks in 26 studies (81%). In 13 studies the Hamilton Depression Rating Scale₁₇ was used, in 15 studies the Hamilton Depression Rating Scale₂₁, and in four studies data of both the Hamilton Depression Rating Scale₁₇ and the Hamilton Depression Rating Scale₂₁ were available.

The mean baseline scores in the studies ranged from 22.8 to 26.8 points in the Hamilton Depression Rating

Table 1
Characteristics of the studies included

Study no.	Year of start of study	Patients	Number of patients		Age		F/M ratio ^a	Duration of study Weeks	HDRS score at baseline		Dose at endpoint mg
			TCA	Placebo	Mean	S.D.			Mean	(S.D.) ^b	
1	1979	Out	22	18	41.8	(13.0) ^c	0.6	4	23.8	(4.20)	182 ^d
2	1979	Mixed	47	41	40.4	(11.5)	2.7	4	23.6	(5.37)	196 ^d
3	1980	Mixed	33	17	43.0	(13.6)	4.0	6	28.9	(6.71)	144 ^d
4	1980	Mixed	27	28	46.1	(12.7)	1.8	6	29.0	(6.01)	164 ^d
5	1980	Out	74	80	41.0	(13.1)	1.2	4	26.2	(5.91)	149 ^d
6	1981	Out	241	232	41.1	(12.2)	1.8	6	24.5	(4.55)	162 ^d
7	1982	Out	30	29	40.9	(13.3)	2.0	6	22.9	(3.59)	141 ^d
8	1983	Out	72	143	38.4	(12.3)	1.4	6	25.4	(3.78)	154 ^d
9	1984	In	14	15	42.0	(9.8)	0.8	6	26.8	(5.55)	125 ^d
10 ^e	1984	–	70	73	44.0	(12.0)	–	6	22.8	(4.67)	–
11 ^e	1984	–	144	140	39.2	(11.5)	–	6	23.3	(3.68)	–
12	1985	Out	233	235	40.5	(12.0)	1.0	6	23.4	(3.48)	140 ^f
13 ^e	1986	Out	13	11	45.5	(9.6)	1.0	6	26.0	(4.32)	140 ^f
14 ^e	1986	Out	49	48	41.2	(10.8)	1.4	6	23.4	(3.64)	105 ^f
15 ^e	1986	Out	49	50	38.3	(12.9)	2.5	6	27.1	(4.16)	131 ^f
16	1986	Out	102	95	41.8	(13.4)	1.5	6	27.2	(3.78)	106 ^d
17	1987	Out	25	27	69.5	(6.7)	0.9	12	26.8	(5.75)	124 ^d
18	1987	Out	70	69	40.9	(10.9)	2.0	6	29.4	(4.44)	132 ^d
19	1987	Out	95	93	42.0	(12.0)	1.8	6	28.5	(4.41)	123 ^d
20	1988	Out	76	78	42.4	(12.6)	1.8	6	22.5	(5.29)	128.5 ^d
21	1988	Out	82	82	38.8	(10.3)	2.2	6	22.8	(3.63)	135 ^d
22	1990	Out	41	47	40.1	(10.6)	0.5	6	26.9	(4.19)	80–240 ^g
23	1990	Out	46	43	35.0	(7.8)	3.0	6	26.3	(3.74)	80–240 ^g
24	1990	Out	44	45	38.2	(10.8)	2.3	6	26.2	(3.72)	80–240 ^g
25	1990	Out	47	46	39.1	(11.3)	1.9	6	24.7	(3.50)	80–240 ^g
26	1990	Out	46	50	39.4	(10.2)	1.2	6	27.6	(3.61)	80–240 ^g
27	1991	Out	36	35	42.7	(10.7)	1.0	6	24.0	(3.22)	165 ^h
28	1991	Out	37	35	41.3	(10.9)	0.8	6	24.0	(3.58)	160.5 ^h
29	1991	Out	50	52	43.0	(9.5)	1.4	6	25.9	(3.42)	253.5 ^h
30	1991	Out	55	57	38.7	(10.9)	1.7	6	23.3	(2.56)	189 ^h
31	1991	Out	79	78	38.3	(11.0)	1.6	8	23.7	(3.30)	137.5 ^h
32	1991	Out	83	90	43.1	(12.9)	1.5	8	23.9	(3.19)	159 ^h

^a F/M ratio, female male ratio; –, no data.

^b Hamilton Depression Rating Scale₁₇ (HDRS₁₇) studies 6, 7, 9–13, 27–32; Hamilton Depression Rating Scale₂₁ (HDRS₂₁) studies 1–5, 8, 16–19, 22–26; Hamilton Depression Rating Scale_{17&21} studies 14, 15, 20, 21.

^c Age was unequally distributed over the tricyclic antidepressant-arm (TCA) and placebo arm.

^d Mean daily dose at last observation.

^e In studies 10, 11, 13, 14 amitryptiline was used as comparator, in the other studies imipramine was the comparator.

^f Mean average dose.

^g No sufficient data, planned maintenance dose according to protocol.

^h Mean modal dose at final week of study.

Scale₁₇ studies and from 23.6 to 29.4 points in the Hamilton Depression Rating Scale₂₁ studies, indicating that the severity of the depression was moderate to severe.

Dosage schedules varied greatly with respect to the duration of the titration period and the maximum dose allowed. In some studies a flexible dose was used during the complete duration of the study, whereas in other studies the dosage was fixed during the last part of the study.

3.2. Drop-out

The mean drop-out rates were 34.5% in tricyclic antidepressant groups and 35.1% in the placebo groups (not

significantly different statistically). There was no difference in drop-out rate between tricyclic antidepressant-arms and placebo-arms. Studies which showed statistically significant results did not differ from studies which did not with respect to drop-out or time to drop-out (correlation coefficient 0.96). Time to drop out did not differ between the two treatment groups.

3.3. Responders

The percentage of responders are shown in Fig. 1 (intention to treat population; last observation carried forward). There were 46% responders in the tricyclic

Study	TCA	Placebo
1	50%	6%
2	45%	41%
3	30%	41%
4	48%	43%
5	41%	29%
6	46%	26%
7	40%	21%
8	53%	42%
9	29%	27%
10	44%	29%
11	57%	36%
12	37%	20%
13	38%	45%
14	55%	29%
15	47%	24%
16	48%	40%
17	40%	26%
18	39%	33%
19	40%	26%
20	47%	42%
21	41%	41%
22	46%	26%
23	52%	33%
24	48%	29%
25	64%	52%
26	52%	50%
27	56%	40%
28	41%	51%
29	48%	35%
30	69%	37%
31	48%	29%
32	40%	33%
HDRS ₁₇ studies	46%	30%
HDRS ₂₁ studies	46%	33%
All studies	46%	31%
Analysis	Difference	CI _{95%}
	14.3%	11.45% ↔ 17.14%
Heterogeneity	χ ²	p
	38.5	0.167

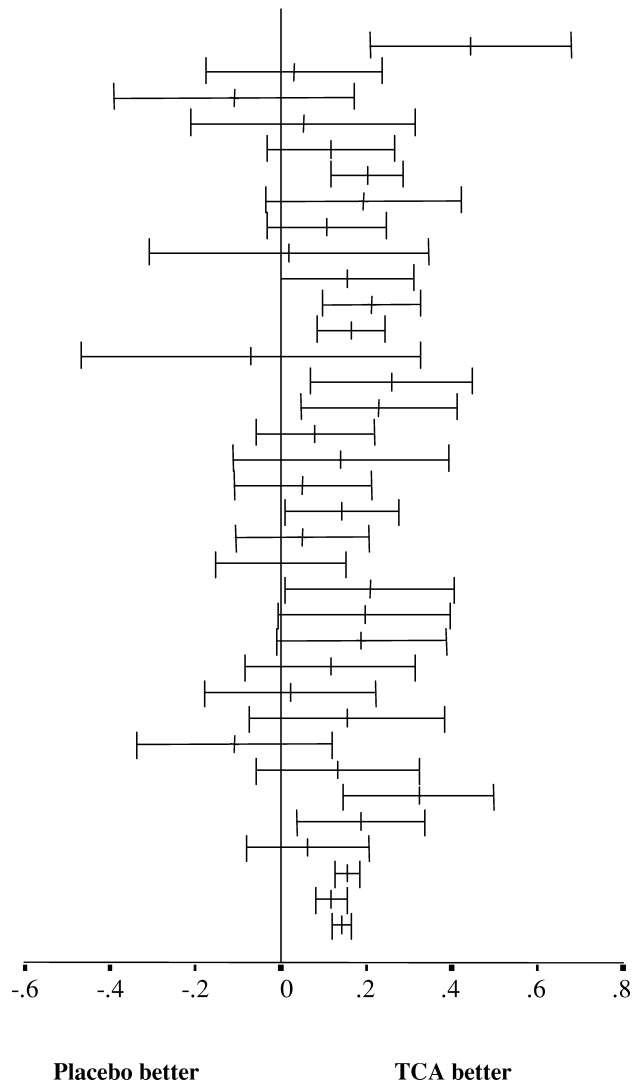


Fig. 1. Percentage of responders at last observation, and difference in proportion of responders between tricyclic antidepressant and placebo by study. A responder was defined as a patient with a 50% improvement from baseline Hamilton Depression Rating Scale. The studies are arranged in order of start-date. The horizontal bars represent the 95% confidence intervals for the difference between tricyclic antidepressant and placebo.

antidepressant group and 31% in the placebo group (CI_{95%-difference} 11.5–17.1%). The number needed to treat was 7 (CI_{95%} 5–8). The number needed to treat is defined as the estimated number of patients who need to be treated with the new treatment rather than standard treatment (e.g. placebo) for one additional patient to benefit (Laupacis et al., 1988; Altman, 1998).

In 10 out of the 32 studies (31%) there was a statistically significant treatment effect in favor of tricyclic antidepressant. In none of the studies there was a statistical difference in favor of placebo. For the completers data set the difference in percentage responders was 16.5% (CI_{95%-difference} 13.9–19.2%).

The responder rate in the placebo-groups varied from 6 to 52%.

3.4. Mean improvement on the Hamilton Depression Rating Scale

The mean improvement from baseline and the differences in mean improvement between the tricyclic antidepressant and placebo arms (intention to treat population; last observation carried forward analysis) are presented in Fig. 2. In 10 studies (31%) a statistically significant treatment effect in favor of tricyclic antidepressant was observed.

Study	Baseline Score		Mean improvement last observation	
	TCA	Placebo	TCA	Placebo
1	24.3	23.2	10.9	4.1
2	23.4	23.8	10.6	9.2
3	29.0	28.6	9.5	9.9
4	29.6	28.4	12.5	10.0
5	26.2	26.3	9.5	6.0
6	24.6	24.4	9.9	6.5
7	22.2	23.6	8.0	3.7
8	25.1	25.6	11.2	9.7
9	27.5	26.1	6.9	3.3
10	23.5	22.1	9.2	6.3
11	23.2	23.4	11.7	8.3
12	23.3	23.5	8.1	5.1
13	26.8	25.0	11.6	9.7
14	23.6	23.2	12.2	6.5
15	27.5	26.8	12.6	7.8
16	27.1	27.4	11.8	10.6
17	26.5	27.0	9.4	7.5
18	29.4	29.3	10.7	10.5
19	28.7	28.4	11.0	8.6
20	22.5	22.4	9.0	8.7
21	23.1	22.4	8.5	8.6
22	27.9	26.0	12.7	8.5
23	26.5	26.0	12.2	8.5
24	25.9	26.4	10.4	6.6
25	25.2	24.3	13.9	10.2
26	27.3	27.9	12.2	11.4
27	23.9	24.1	11.7	8.9
28	24.2	23.8	10.0	10.8
29	25.7	26.1	10.8	8.0
30	23.5	23.1	13.8	8.1
31	23.7	23.8	10.1	8.3
32	24.3	23.5	9.9	7.2
HDRS ₁₇ studies	23.7	23.9	9.99	7.14
HDRS ₂₁ studies	26.7	26.9	11.21	8.74
All studies	24.9	25.0	10.4	7.69

Analysis	Diff	CI _{95%}	Heterogeneity		
			χ ²	df	p
HDRS ₁₇ studies	-2.76	-3.35 ⇔ -2.18	13.8	16	0.47
HDRS ₂₁ studies	-2.43	-3.25 ⇔ -1.60	27.5	14	0.04
All studies	-2.65	-3.13 ⇔ -2.17	41.7	31	0.10

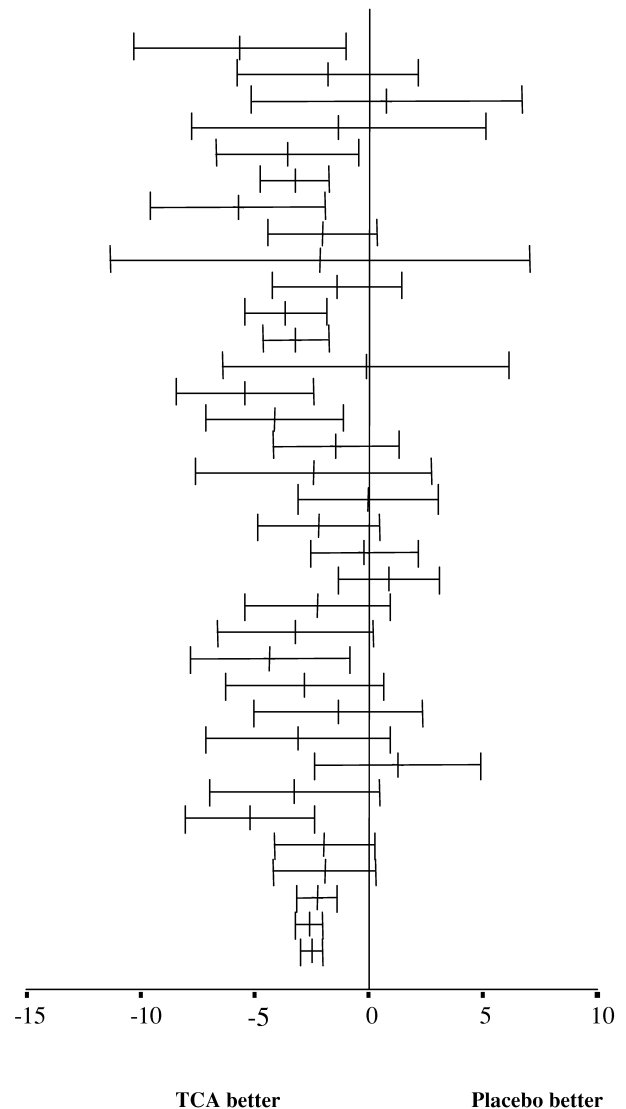


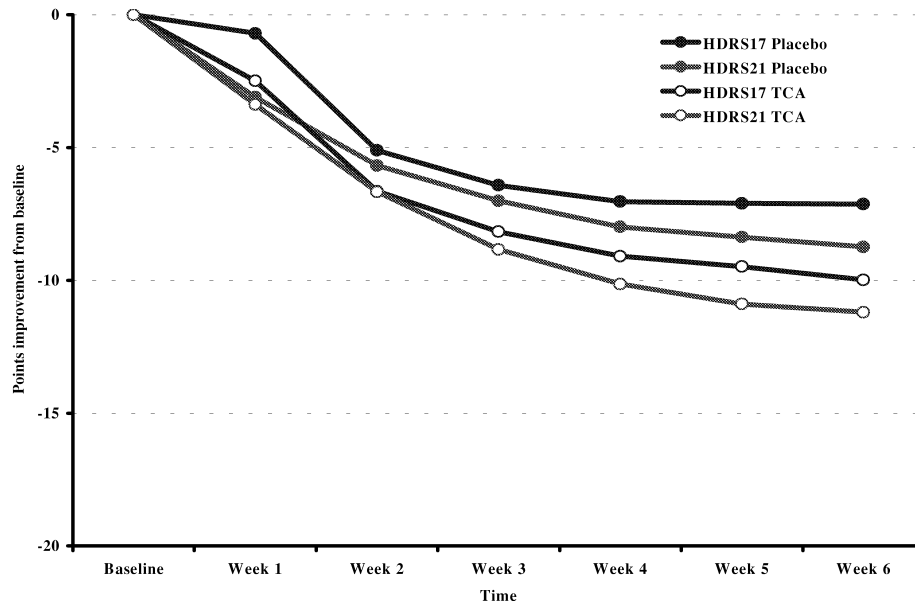
Fig. 2. Baseline Hamilton Depression Rating Scale score, mean improvement from baseline at last observation and estimated difference with 95% confidence intervals. The studies are arranged in order of start-date.

The mean improvement in the placebo-group varied from 13 to 45%.

3.5. Week-by-week improvement

The week-by-week improvement on the Hamilton De-

pression Rating Scale is presented in Fig. 3. The difference in treatment effect between tricyclic antidepressant and placebo increases steadily during the duration of the study. This difference becomes and remains statistically significant from week 2 on. The magnitude of the effect in Hamilton Depression Rating Scale₁₇ studies (intention to



HDRS ₁₇ studies							
TCA	24.1	-2.5	-6.6	-8.2	-9.1	-9.5	-10.0
Placebo	23.8	-0.7	-5.1	-6.4	-7.0	-7.1	-7.1
Difference		-0.3	-1.5	-1.7	-2.0	-2.3	-2.8
CI _{95%}		-0.6 ⇔ -0.1	-2.0 ⇔ -1.1	-2.2 ⇔ -1.2	-2.5 ⇔ -1.5	-2.8 ⇔ -1.7	-3.3 ⇔ -2.2
Heterogeneity							
χ ²		26.7	33.6	27.5	25.2	31.9	27.5
p=		0.05	0.01	0.04	0.07	0.01	0.04
HDRS ₂₁ studies							
TCA	26.7	-3.4	-6.7	-8.8	-10.1	-10.9	-11.2
Placebo	26.4	-3.1	-5.7	-7.0	-8.0	-8.4	-8.7
Difference		-0.3	-1.0	-1.8	-2.2	-2.5	-2.4
CI _{95%}		-0.8 ⇔ 0.1	-1.6 ⇔ -0.4	-2.5 ⇔ -1.2	-2.9 ⇔ -1.4	-3.3 ⇔ -1.7	-3.3 ⇔ -1.6
Heterogeneity							
χ ²		14.9	15.4	22.6	21.0	16.5	13.8
p=		0.38	0.35	0.07	0.10	0.28	0.47

Fig. 3. Mean improvement on the Hamilton Depression Rating Scale score from baseline for tricyclic antidepressant and placebo for studies on the Hamilton Depression Rating Scale₁₇ (HDRS₁₇) and Hamilton Depression Rating Scale₂₁. The data below presents the weighed absolute values per treatment arm, difference between tricyclic antidepressant (TCA) and placebo and the result of the heterogeneity tests per time point. A negative value indicates an improvement in favor of tricyclic antidepressant. The estimates are weighed for variance differences between the studies.

treat population; last observation carried forward analysis) was -2.8 ($CI_{95\%}$ -3.3 to -2.2) and in Hamilton Depression Rating Scale₂₁ studies -2.4 ($CI_{95\%}$ -3.3 to -1.6).

3.6. Dose–effect relationship

In an explorative multiple regression model the impact of doses and other variables on the outcome has been evaluated. In 13 studies (1525 patients) the daily dose at last observation was known for both treatment arms. In these studies imipramine was the active comparator. An independent dose–effect relationship could be shown for the outcome responders, and for mean improvement on the Hamilton Depression Rating Scale in the tricyclic antidepressant group. The probability of becoming a responder increased with about 0.05 per 100 mg increase in tricyclic antidepressant dose. The score on the Hamilton Depression Rating Scale improved 1.1 points for every 100 mg increase in tricyclic antidepressant dose (intention to treat population; last observation carried forward analysis).

4. Discussion

Our meta-analysis shows that tricyclic antidepressants are efficacious in the short-term treatment of major depression. However, seven patients need to be treated to obtain one additional responder above placebo indicating that tricyclic antidepressants are rather weak compounds. This is also shown by the small difference of 2.8 points in improvement on the HDRS between the TCA group and the placebo-group. In the tricyclic antidepressant group the improvement from baseline was 42% and in the placebo-group 31%, indicating that the success rate in the placebo-group was responsible for more than 70% of the effect of the active drug. This is in line with the results of the meta-analysis performed by Kirsch and Sapirstein (1998) and Kahn et al. (2000).

As a dose-relationship was present, one of the explanations of the results could be the rather low dosages of active drug used in the studies. However the slope was modest. Therefore the modest results cannot be explained by sub-therapeutic dosages of TCA.

The studies included in our meta-analysis were part of registration dossiers for granting a license for the indication major depression. These dossiers contain all clinical trials conducted for a drug under development, including all short-term, parallel, randomized studies with both a tricyclic antidepressant and placebo as controls. Therefore, our analysis was based on published as well as unpublished material. Moreover, these dossiers do not contain aggregate and/or separate center studies from multicenter studies as is often found in the literature. Therefore, although one company did not participate in submitting data, the data used in our meta-analysis are sufficiently informative for the purpose of the investigation.

Is it still necessary to conduct placebo-controlled studies for showing efficacy in the short-term treatment of major depression? Rothman and Michels (1994) considered it to be unethical to conduct placebo-controlled studies in major depression, stating that equivalence studies might be sufficient to demonstrate efficacy. However, our results indicate that in the short-term treatment of major depression it is not possible to demonstrate efficacy based on equivalence studies because tricyclic antidepressants are not superior to placebo in 22/32 (69%) of the placebo-controlled studies. As large studies show more often a statistically significant difference in favor of tricyclic antidepressants than small studies, one could argue that the reason for our finding is that many studies were underpowered, suggesting that large equivalence studies could be sufficient to demonstrate efficacy. However, we also found that the size of the success rate in the placebo-group considerably varied from study to study, indicating that a non-inferiority margin in an equivalence study cannot be defined. Therefore an equivalence study is not an option to demonstrate efficacy and placebo-controlled studies are still necessary to demonstrate efficacy in major depression.

Acknowledgements

We thank the following pharmaceutical companies: AHP, Bristol-Myers Squibb, Eli Lilly, Lundbeck, Organon, Pfizer, SmithKline Beecham and Solvay Pharmaceuticals, for providing the data that are used in this meta-analysis. Moreover we thank M. van der Graaff (Nefarma), Prof. Dr. J.P. Vandenbroucke (LUMC), Prof. Dr. F.R. Rosendaal (LUMC), Prof. Dr. H.G.M. Rooymans (LUMC), Dr. A.M. van Hemert (LUMC) and especially Dr. J.E. Hoogendijk (UMCU) for their contribution making this study possible.

References

- Altman, D.G., 1998. Confidence intervals for the number needed to treat. *Br. Med. J.* 317, 1309–1312.
- American Psychiatric Association, 1980. In: 3rd Edition. *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association, Washington, DC.
- American Psychiatric Association, 1987. In: 3rd Edition, revised. *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association, Washington, DC.
- Endicott, J., Cohen, J., Nee, J., Fleiss, J., Sarantakos, S., 1981. Hamilton Depression Rating Scale. Extracted from regular and change versions of the schedule for affective disorders and schizophrenia. *Arch. Gen. Psychiatry* 38, 98–103.
- Hamilton, M., 1960. A rating scale for depression. *J. Neurol. Neurosurg. Psychiatry* 23, 50–55.
- Khan, A., Warner, H.A., Brown, W.A., 2000. Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials. An analysis of the Food and Drug Administration database. *Arch. Gen. Psychiatry* 57, 311–317.
- Kirsch, I., Sapirstein, G., 1998. Listening to prozac but hearing placebo: a

- meta-analysis of antidepressant medication. *Prev. Treat.* (<http://journals.apa.org/prevention/volume1/pre001002a.html>).
- Klein, D.F., Gittelman, R., Quitkin, F., Rifkin, A., 1980. In: 2nd Edition. *Diagnosis and Drug Treatment of Psychiatric Disorders: Adults and Children*. Williams and Wilkins, Baltimore.
- Kuhn, R., 1958. The treatment of depressive states with G-22355 (imipramine hydrochloride). *Am. J. Psychiatry* 115, 459–464.
- Laupacis, A., Sackett, D.L., Roberts, R.S., 1988. An assessment of clinically useful measures of the consequences of treatment. *N. Engl. J. Med.* 318, 1728–1733.
- Montgomery, S.A., 1995. Clinical relevance of response and improvement in psychopharmacology. A statement from the European College of Neuropsychopharmacology, from an ECNP workshop, Jerusalem, October 1995. *Eur. Neuropsychopharmacol.* 5, 511–533.
- Rothman, K.J., Michels, K.B., 1994. The continuing unethical use of placebo controls. *N. Engl. J. Med.* 331, 394–398.