Prevention & Treatment

Prevention & Treatment, Volume 1, Article 0002a, posted June 26, 1998

Listening to Prozac but Hearing Placebo: A Meta-Analysis of Antidepressant Medication

Irving Kirsch, Ph.D.
University of Connecticut, Storrs, CT

Guy Sapirstein, Ph.D.
Westwood Lodge Hospital, Needham, MA

ABSTRACT

Mean effect sizes for changes in depression were calculated for 2,318 patients who had been randomly assigned to either antidepressant medication or placebo in 19 double-blind clinical trials. As a proportion of the drug response, the placebo response was constant across different types of medication (75%), and the correlation between placebo effect and drug effect was .90. These data indicate that virtually all of the variation in drug effect size was due to the placebo characteristics of the studies. The effect size for active medications that are not regarded to be antidepressants was as large as that for those classified as antidepressants, and in both cases, the inactive placebos produced improvement that was 75% of the effect of the active drug. These data raise the possibility that the apparent drug effect (25% of the drug response) is actually an active placebo effect. Examination of pre–post effect sizes among depressed individuals assigned to no-treatment or wait-list control groups suggest that approximately one quarter of the drug response is due to the administration of an active medication, one half is a placebo effect, and the remaining quarter is due to other nonspecific factors.

EDITORS' NOTE

The article that follows is a controversial one. It reaches a controversial conclusion—that much of the therapeutic benefit of antidepressant medications actually derives from placebo responding. The article reaches this conclusion by utilizing a controversial statistical approach—meta-analysis. And it employs meta-analysis controversially—by meta-analyzing studies that are very heterogeneous in subject selection criteria, treatments employed, and statistical methods used. Nonetheless, we have chosen to publish the article. We have done so because a number of the colleagues who originally reviewed the manuscript believed it had considerable merit, even while they recognized the clearly contentious conclusions it reached and the clearly arguable statistical methods it employed.

We are convinced that one of the principal aims of an electronic journal ought to be to bring our readers information on a variety of current topics in prevention and treatment, even though much of it will be subject to heated differences of opinion about worth and ultimate significance. This is to be expected, of course, when one is publishing material at the cutting-edge, in a cutting-edge medium.

We also believe, however, that soliciting expert commentary to accompany particularly controversial articles facilitates the fullest possible airing of the issues most germane to appreciating both the strengths and the weaknesses of target articles. In the same vein, we welcome comments on the article from readers as well, though for obvious

reasons, we cannot promise to publish all of them.

Feel free to submit a comment by emailing admin@apa.org.

Peter Nathan, Associate Editor (Treatment)
Martin E. P. Seligman, Editor

We thank R. B. Lydiard and Smith-Kline Beecham Pharaceuticals for supplying additional data. We thank David Kenny for his assistance with the statistical analyses. We thank Roger P. Greenberg and Daniel E. Moerman for their helpful comments on earlier versions of this paper.

Correspondence concerning this article should be addressed to Irving Kirsch, Department of Psychology, U-20, University of Connecticut, 406 Babbidge Road Storrs, CT 06269-1020.
E-mail: Irvingk@uconnvm.uconn.edu

More placebos have been administered to research participants than any single experimental drug. Thus, one would expect sufficient data to have accumulated for the acquisition of substantial knowledge of the parameters of placebo effects. However, although almost everyone controls for placebo effects, almost no one evaluates them. With this in mind, we set about the task of using meta-analytic procedures for evaluating the magnitude of the placebo response to antidepressant medication.

Meta-analysis provides a means of mathematically combining results from different studies, even when these studies have used different measures to assess the dependent variable. Most often, this is done by using the statistic *d*, which is a standardized difference score. This effect size is generally calculated as the mean of the experimental group minus the mean of the control group, divided by the pooled standard deviation. Less frequently, the mean difference is divided by standard deviation of the control group (Smith, Glass, & Miller, 1980).

Ideally, to calculate the effect size of placebos, we would want to subtract the effects of a no-placebo control group. However, placebos are used as controls against which the effects of physical interventions can be gauged. It is rare for an experimental condition to be included against which the effects of the placebo can be evaluated. To circumvent this problem, we decided to calculate within-cell or pre–post effect sizes, which are the posttreatment mean depression score minus the pretreatment mean depression score, divided by the pooled standard deviation (cf. Smith et al., 1980). By doing this for both placebo groups and medication groups, we can estimate the proportion of the response to antidepressant medication that is duplicated by placebo administration, a response that would be due to such factors as expectancy for improvement and the natural course of the disorder (i.e., spontaneous remission). Later in this article, we also separate expectancy from natural history and provide estimates of each of these effects.

Although our approach is unusual, in most cases it should provide results that are comparable to conventional methods. If there are no significant pretreatment differences between the treatment and control groups, then the subtraction of mean standardized pre–post difference scores should result in a mean effect size that is just about the same as that produced by subtracting mean standardized posttreatment scores. Suppose, for example, we have a study with the data displayed in Table 1. The conventionally calculated effect size would be would be 1.00. The pre–post effect sizes would be 3.00 for the treatment group and 2.00 for the control group. The difference between them is 1.00, which is exactly the same effect calculated from posttreatment scores alone. However, calculating the effect size in this manner also provides us with the information that the effect of the control procedure was 2/3 that of the treatment procedure, information that we do not have when we only consider posttreatment scores. Of course, it is rare for two groups to have identical mean pretreatment scores, and to the extent that those scores are different, our two methods of calculation would provide different results. However, by controlling for baseline differences, our method should provide the more accurate estimate of differential outcome.

Table 1
Hypothetical Means and Standard Deviations for a Treatment Group and a Control Group

	Treatment		Control	
	Pretreatment	**Posttreatment**	**Pretreatment**	**Posttreatment**
M	25.00	10.00	25.00	15.00
SD	5.50	4.50	4.50	5.50

The Effects of Medication and Placebo

Study Characteristics

Studies assessing the efficacy of antidepressant medication were obtained through previous reviews (Davis, Janicak, & Bruninga, 1987; Free & Oei, 1989; Greenberg & Fisher, 1989; Greenberg, Bornstein, Greenberg, & Fisher, 1992; Workman & Short, 1993), supplemented by a computer search of PsycLit and MEDLINE databases from 1974 to 1995 using the search terms *drug-therapy or pharmacotherapy or psychotherapy or placebo and depression or affective disorders. Psychotherapy* was included as a search term for the purpose of obtaining articles that would allow estimation of changes occurring in no-treatment and wait-list control groups, a topic to which we return later in this article. Approximately 1,500 publications were produced by this literature search. These were examined by the second author, and those meeting the following criteria were included in the meta-analysis:

1. The sample was restricted to patients with a primary diagnosis of depression. Studies were excluded if participants were selected because of other criteria (eating disorders, substance abuse, physical disabilities or chronic medical conditions), as were studies in which the description of the patient population was vague (e.g., "neurotic").
2. Sufficient data were reported or obtainable to calculate within-condition effect sizes. This resulted in the exclusion of studies for which neither pre–post statistical tests nor pretreatment means were available.
3. Data were reported for a placebo control group.
4. Participants were assigned to experimental conditions randomly.
5. Participants were between the ages of 18 and 75.

Of the approximately 1,500 studies examined, 20 met the inclusion criteria. Of these, all but one were studies of the acute phase of therapy, with treatment durations ranging from 1 to 20 weeks (M = 4.82). The one exception (Doogan & Caillard, 1992) was a maintenance study, with a duration of treatment of 44 weeks. Because of this difference, Doogan and Caillard's study was excluded from the meta-analysis. Thus, the analysis was conducted on 19 studies containing 2,318 participants, of whom 1,460 received medication and 858 received placebo. Medications studied were amitriptyline, amylobarbitone, fluoxetine, imipramine, paroxetine, isocarboxazid, trazodone, lithium, liothyronine, adinazolam, amoxapine, phenelzine, venlafaxine, maprotiline, tranylcypromine, and bupropion.

The Calculation of Effect Sizes

In most cases, effect sizes (d) were calculated for measures of depression as the mean posttreatment score minus the mean pretreatment score, divided by the pooled standard deviation (SD). Pretreatment SDs were used in place of pooled SDs in calculating effect sizes for four studies in which posttreatment SDs were not reported (Ravaris, Nies, Robinson, et al., 1976; Rickels & Case, 1982; Rickels, Case, Weberlowsky, et al.,

1981; Robinson, Nies, & Ravaris, 1973). The methods described by Smith et al. (1980) were used to estimate effect sizes for two studies in which means and *SD*s were not reported. One of these studies (Goldberg, Rickels, & Finnerty, 1981) reported the *t* value for the pre–post comparisons. The effect size for this study was estimated using the formula:

$$d = t\,(2/n)^{1/2}$$

where *t* is the reported *t* value for the pre–post comparison, and *n* is the number of subjects in the condition. The other study (Kiev & Okerson, 1979) reported only that there was a significant difference between pre- and posttreatment scores. As suggested by Smith et al. (1980), the following formula for estimating the effect size was used:

$$d = 1.96\,(2/n)^{1/2},$$

where 1.96 is used as the most conservative estimation of the *t* value at the .05 significance level used by Kiev and Okerson. These two two effect sizes were also corrected for pre–post correlation by multiplying the estimated effect size by $(1 - r)^{1/2}$, *r* being the estimate of the test–retest correlation (Hunter & Schmidt, 1990). Bailey and Coppen (1976) reported test–retest correlations of .65 for the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) and .50 for the Hamilton Rating Scale for Depression (HRS-D; Hamilton, 1960) . Therefore, in order to arrive at an estimated effect size, corrected for the pre–post correlation, the estimated effect sizes of the HRS-D were multiplied by 0.707 and the effect sizes of the BDI were multiplied by 0.59.

In studies reporting multiple measures of depression, an effect size was calculated for each measure and these were then averaged. In studies reporting the effects of two drugs, a single mean effect size for both was calculated for the primary analysis. In a subsequent analysis, the effect for each drug was examined separately. In both analyses, we calculated mean effect sizes weighted for sample size (*D*; Hunter & Schmidt, 1990).

Effect Sizes

Sample sizes and effect sizes for patients receiving medication or placebo are presented in Table 2. Mean effect sizes, weighted for sample size, were 1.55 *SD*s for the medication response and 1.16 for the placebo response. Because effect sizes are obtained by dividing both treatment means by a constant (i.e., the pooled *SD*), they can be treated mathematically like the scores from which they are derived. [1] In particular, we have shown that, barring pretreatment between-group differences, subtracting the mean pre–post effect size of the control groups from the mean pre–post effect size of the experimental groups is equivalent to calculating an effect size by conventional means. Subtracting mean placebo response rates from mean drug response rates reveals a mean medication effect of 0.39 *SD*s. This indicates that 75% of the response to the medications examined in these studies was a placebo response, and at most, 25% might be a true drug effect. This does not mean that only 25% of patients are likely to respond to the pharmacological properties of the drug. Rather, it means that for a typical patient, 75% of the benefit obtained from the active drug would also have obtained from an inactive placebo.

Table 2

Studies Including Placebo Control Groups

	Drug		Placebo	
Study	***n***	***d***	***n***	***d***
Blashki et al. (1971)	43	1.75	18	1.02
Byerly et al. (1988)	44	2.30	16	1.37
Claghorn et al. (1992)	113	1.91	95	1.49
Davidson & Turnbull (1983)	11	4.77	8	2.28
Elkin et al. (1989)	36	2.35	34	2.01
Goldberg et al. (1981)	179	0.44	93	0.44
Joffe et al. (1993)	34	1.43	16	0.61
Kahn et al. (1991)	66	2.25	80	1.48
Kiev & Okerson (1979)	39	0.44	22	0.42
Lydiard (1989)	30	2.59	15	1.93
Ravaris et al. (1976)	14	1.42	19	0.91
Rickels et al. (1981)	75	1.86	23	1.45
Rickels & Case (1982)	100	1.71	54	1.17
Robinson et al. (1973)	33	1.13	27	0.76
Schweizer et al. (1994)	87	3.13	57	2.13
Stark & Hardison (1985)	370	1.40	169	1.03
van der Velde (1981)	52	0.66	27	0.10
White et al. (1984)	77	1.50	45	1.14
Zung (1983)	57	.88	40	0.95

Inspection of Table 2 reveals considerable variability in drug and placebo response effect sizes. As a first step toward clarifying the reason for this variability, we calculated the correlation between drug response and placebo response, which was found to be exceptionally high, $r = .90$, $p < .001$ (see Figure 1). This indicates that the placebo response was proportionate to the drug response, with remaining variability most likely due to measurement error.

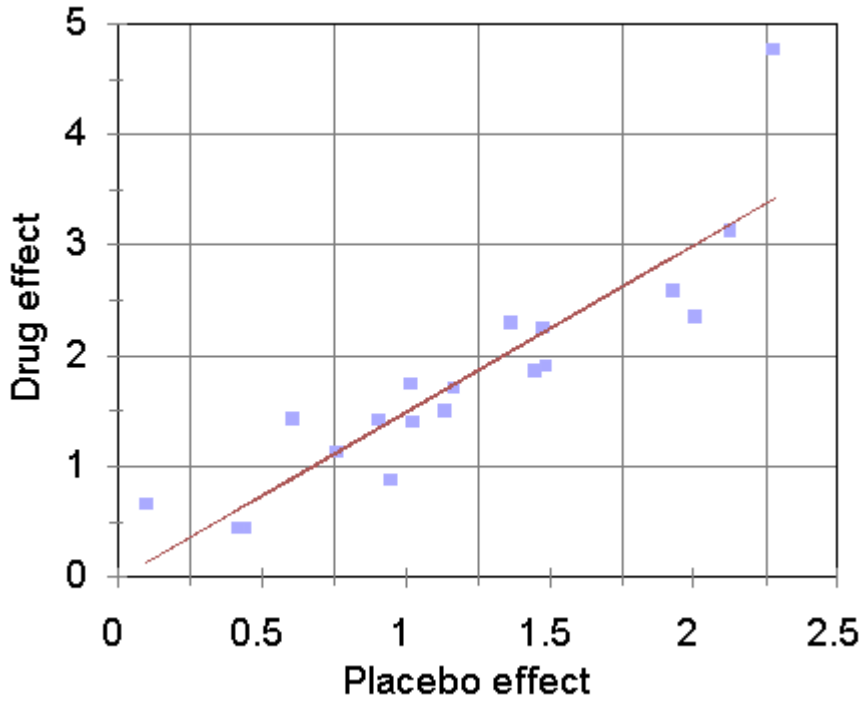

Figure 1. The placebo response as a predictor of the drug response.

Our next question was the source of the common variability. One possibility is that the correlation between placebo and drug response rates are due to between-study differences in sample characteristics (e.g., inpatients vs. outpatients, volunteers vs. referrals, etc.). Our analysis of psychotherapy studies later in this article provides a test of this hypothesis. If the correlation is due to between-study differences in sample characteristics, a similar correlation should be found between the psychotherapy and no-treatment response rates. In fact, the correlation between the psychotherapy response and the no-treatment response was nonsignificant and in the opposite direction. This indicates that common sample characteristics account for little if any of the relation between treatment and control group response rates.

Another possibility is that the close correspondence between placebo and drug response is due to differences in so-called nonspecific variables (e.g., provision of a supportive relationship, color of the medication, patients' expectations for change, biases in clinician's ratings, etc.), which might vary from study to study, but which would be common to recipients of both treatments in a given study. Alternately, the correlation might be associated with differences in the effectiveness of the various medications included in the meta-analysis. This could happen if more effective medications inspired greater expectations of improvement among patients or prescribing physicians (Frank, 1973; Kirsch, 1990). Evans (1974), for example, reported that placebo morphine was substantially more effective than placebo aspirin. Finally, both factors might be operative.

We further investigated this issue by examining the magnitude of drug and placebo responses as a function of type of medication. We subdivided medication into four types: (a) tricyclics and tetracyclics, (b) selective serotonin reuptake inhibitors (SSRI), (c) other antidepressants, and (d) other medications. This last category

consisted of four medications (amylobarbitone, lithium, liothyronine, and adinazolam) that are not considered antidepressants.

Weighted (for sample size) mean effect sizes of the drug response as a function of type of medication are shown in Table 3, along with corresponding effect sizes of the placebo response and the mean effect sizes of placebo responses as a proportion of drug responses. These data reveal relatively little variability in drug response and even less variability in the ratio of placebo response to drug response, as a function of drug type. For each type of medication, the effect size for the active drug response was between 1.43 and 1.69, and the inactive placebo response was between 74% and 76% of the active drug response. These data suggest that the between-drug variability in drug and placebo response was due entirely to differences in the placebo component of the studies.

Table 3
Effect Sizes as a Function of Drug Type

Statistic	Type of drug			
	Antidepressant			Other drugs
	Tri- and tetracyclic	SSRI	Other	
N	1,353	626	683	203
K	13	4	8	3
D—Drug	1.52	1.68	1.43	1.69
D—Placebo	1.15	1.24	1.08	1.29
Placebo/drug	.76	.74	.76	.76

N = number of subjects; *K* = number of studies; *D* = mean weighted effect size; placebo/drug = placebo response as a proportion of active drug response.

Differences between active drug responses and inactive placebo responses are typically interpreted as indications of specific pharmacologic effects for the condition being treated. However, this conclusion is thrown into question by the data derived from active medications that are not considered effective for depression. It is possible that these drugs affect depression indirectly, perhaps by improving sleep or lowering anxiety. But if this were the case and if antidepressants have a specific effect on depression, then the effect of these other medications ought to have been less than the effect of antidepressants, whereas our data indicate that the response to these nonantidepressant drugs is at least as great as that to conventional antidepressants.

A second possibility is that amylobarbitone, lithium, liothyronine, and adinazolam are in fact antidepressants. This conclusion is rendered plausible by the lack of understanding of the mechanism of clinical action of common antidepressants (e.g., tricyclics). If the classification of a drug as an antidepressant is established by its efficacy, rather than by knowledge of the mechanism underlying its effects, then amylobarbitone, lithium, liothyronine, and adinazolam might be considered specifics for depression.

A third possibility is that these medications function as active placebos (i.e., active medications without specific activity for the condition being treated). Greenberg and Fisher (1989) summarized data indicating that the effect of antidepressant medication is smaller when it is compared to an active placebo than when it is compared to an inert placebo (also see Greenberg & Fisher, 1997). By definition, the only difference between active and inactive placebos is the presence of pharmacologically induced side effects. Therefore, differences in responses to active and inert placebos could be due to the presence of those side effects. Data from other studies indicate that most participants in studies of antidepressant medication are able to deduce whether they have been assigned to the drug condition or the placebo condition (Blashki, Mowbray, &

Davies, 1971; Margraf, Ehlers, Roth, Clark, Sheikh, Agras, & Taylor, 1991; Ney, Collins, & Spensor, 1986). This is likely to be associated with their previous experience with antidepressant medication and with differences between drug and placebo in the magnitude of side effects. Experiencing more side effects, patients in active drug conditions conclude that they are in the drug group; experiencing fewer side effects, patients in placebo groups conclude that they are in the placebo condition. This can be expected to produce an enhanced placebo effect in drug conditions and a diminished placebo effect in placebo groups. Thus, the apparent drug effect of antidepressants may in fact be a placebo effect, magnified by differences in experienced side effects and the patient's subsequent recognition of the condition to which he or she has been assigned. Support for this interpretation of data is provided by a meta-analysis of fluoxetine (Prozac), in which a correlation of .85 was reported between the therapeutic effect of the drug and the percentage of patients reporting side effects (Greenberg, Bornstein, Zborowski, Fisher, & Greenberg, 1994).

Natural History Effects

Just as it is important to distinguish between a drug response and a drug effect, so too is it worthwhile to distinguish between a placebo response and a placebo effect (Fisher, Lipman, Uhlenhuth, Rickels, & Park, 1965). A drug response is the change that occurs after administration of the drug. The effect of the drug is that portion of the response that is due to the drug's chemical composition; it is the difference between the drug response and the response to placebo administration. A similar distinction can be made between placebo responses and placebo effects. The placebo response is the change that occurs following administration of a placebo. However, change might also occur without administration of a placebo. It may be due to spontaneous remission, regression toward the mean, life changes, the passage of time, or other factors. The placebo effect is the difference between the placebo response and changes that occur without the administration of a placebo (Kirsch, 1985, 1997).

In the preceding section, we evaluated the placebo response as a proportion of the response to antidepressant medication. The data suggest that at least 75% of the drug response is a placebo response, but it does not tell us the magnitude of the placebo effect. What proportion of the placebo response is due to expectancies generated by placebo administration, and what proportion would have occurred even without placebo administration? That is a much more difficult question to answer. We have not been able to locate any studies in which pre- and posttreatment assessments of depression were reported for both a placebo group and a no-treatment or wait-list control group. For that reason, we turned to psychotherapy outcome studies, in which the inclusion of untreated control groups is much more common.

We acknowledge that the use of data from psychotherapy studies as a comparison with those from drug studies is far less than ideal. Participants in psychotherapy studies are likely to differ from those in drug studies on any number of variables. Furthermore, the assignment of participants to a no-treatment or wait-list control group might also effect the course of their disorder. For example, Frank (1973) has argued that the promise of future treatment is sufficient to trigger a placebo response, and a wait-list control group has been conceptualized as a placebo control group in at least one well-known outcome study (Sloane, Staples, Cristol, Yorkston, & Whipple, 1975). Conversely, one could argue that being assigned to a no-treatment control group might strengthen feelings of hopelessness and thereby increase depression. Despite these problems, the no-treatment and wait-list control data from psychotherapy outcome studies may be the best data currently available for estimating the natural course of untreated depression. Furthermore, the presence of both types of untreated control groups permits evaluation of Frank's (1973) hypothesis about the curative effects of the promise of treatment.

Study Characteristics

Studies assessing changes in depression among participants assigned to wait-list or no-treatment control

groups were obtained from the computer search described earlier, supplemented by an examination of previous reviews (Dobson, 1989; Free, & Oei, 1989; Robinson, Berman, & Neimeyer, 1990). The publications that were produced by this literature search were examined by the second author, and those meeting the following criteria were included in the meta-analysis:

1. The sample was restricted to patients with a primary diagnosis of depression. Studies were excluded if participants were selected because of other criteria (eating disorders, substance abuse, physical disabilities or chronic medical conditions), as were studies in which the description of the patient population was vague (e.g., "neurotic").
2. Sufficient data were reported or obtainable to calculate within-condition effect sizes.
3. Data were reported for a wait-list or no-treatment control group.
4. Participants were assigned to experimental conditions randomly.
5. Participants were between the ages of 18 and 75.

Nineteen studies were found to meet these inclusion criteria, and in all cases, sufficient data had been reported to allow direct calculation of effect sizes as the mean posttreatment score minus the mean pretreatment score, divided by the pooled *SD*. Although they are incidental to the main purposes of this review, we examined effect sizes for psychotherapy as well as those for no-treatment and wait-list control groups.

Effect Sizes

Sample sizes and effect sizes for patients assigned to psychotherapy, wait-list, and no-treatment are presented in Table 4. Mean pre–post effect sizes, weighted for sample size, were 1.60 for the psychotherapy response and 0.37 for wait-list and no-treatment control groups. Participants given the promise of subsequent treatment (i.e., those in wait-list groups) did not improve more than those not promised treatment. Mean effect sizes for these two conditions were 0.36 and 0.39, respectively. The correlation between effect sizes (r = -.29) was not significant.

Table 4

Studies Including Wait-List or No-Treatment Control Groups

Study	Psychotherapy		Control	
	n	***d***	***n***	***d***
Beach & O'Leary (1992)	15	2.37	15	0.97
Beck & Strong (1982)	20	2.87	10	-0.28
Catanese et al. (1979)	99	1.39	21	0.16
Comas-Diaz (1981)	16	1.87	10	-0.12
Conoley & Garber (1985)	38	1.10	19	0.21
Feldman et al. (1982)	38	2.00	10	0.42
Graff et al. (1986)	24	2.03	11	-0.03
Jarvinen & Gold (1981)	46	0.76	18	0.34
Maynard (1993)	16	1.06	14	0.36
Nezu (1986)	23	2.39	9	0.16
Rehm et al. (1981)	42	1.23	15	0.48
Rude (1986)	8	1.75	16	0.74
Schmidt & Miller (1983)	34	1.25	10	0.11
Shaw (1977)	16	2.17	8	0.41
Shipley & Fazio (1973)	11	2.12	11	1.00
Taylor & Marshall (1977)	21	1.94	7	0.27
Tyson & Range (1981)	22	0.67	11	1.45
Wierzbicki & Bartlett (1987)	18	1.17	20	0.21
Wilson et al. (1983)	16	2.17	9	-0.02

Comparison of Participants in the Two Groups of Studies

Comparisons of effect sizes from different sets of studies is common in meta-analysis. Nevertheless, we examined the characteristics of the samples in the two types of studies to assess their comparability. Eighty-six percent of the participants in the psychotherapy studies were women, as were 65% of participants in the drug studies. The age range of participants was 18 to 75 years (M = 30.1) in the psychotherapy studies and 18 to 70 years (M = 40.6) in the drug studies. Duration of treatment ranged from 1 to 20 weeks (M = 4.82) in psychotherapy studies and from 2 to 15 weeks (M = 5.95) in pharmacotherapy studies. The HRS-D was used in 15 drug studies involving 2,016 patients and 5 psychotherapy studies with 191 participants. Analysis of variance weighted by sample size did not reveal any significant differences in pretreatment HRS-D scores between patients in the drug studies (M = 23.93, SD = 5.20) and participants in the psychotherapy studies (M = 21.34, SD = 5.03). The Beck Depression Inventory (BDI) was used in 4 drug studies involving 261 patients and in 17 psychotherapy studies with 677 participants. Analysis of variance weighted by sample size did not reveal any significant differences in pretreatment BDI scores between participants in drug studies (M = 21.58, SD = 8.23) and those in psychotherapy studies (M = 21.63, SD = 6.97). Thus, participants in the two types of studies were comparable in initial levels of depression. These analyses also failed to reveal any pretreatment differences as a function of group assignment (treatment or control) or the interaction between type of study and group assignment.

Estimating the Placebo Effect

Just as drug effects can be estimated as the drug response minus the placebo response, placebo effects can be estimated as the placebo response minus the no-treatment response. Using the effect sizes obtained from the two meta-analyses reported above, this would be 0.79 (1.16 - 0.37). Figure 2 displays the estimated drug, placebo, and no-treatment effect sizes as proportions of the drug response (i.e., 1.55 SDs). These data indicate that approximately one quarter of the drug response is due to the administration of an active medication, one half is a placebo effect, and the remaining quarter is due to other nonspecific factors.

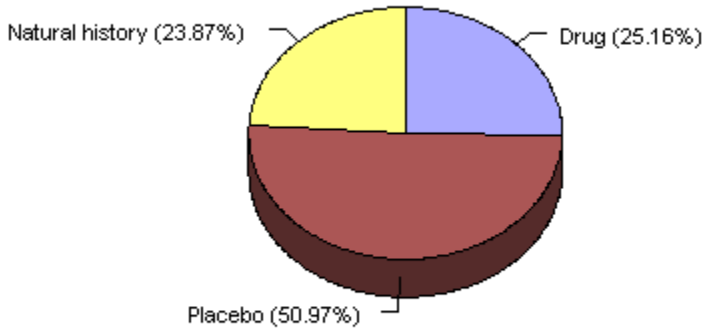

Figure 2. Drug effect, placebo effect, and natural history effect as proportions of the response to antidepressant medication.

Discussion

No-treatment effect sizes and effect sizes for the placebo response were calculated from different sets of

studies. Comparison across different samples is common in meta-analyses. For example, effect sizes derived from studies of psychodynamic therapy are often compared to those derived from studies of behavior therapy (e.g., Andrews & Harvey, 1981; Smith et al., 1980). Nevertheless, comparisons of this sort should be interpreted cautiously. Participants volunteering for different treatments might come from a different populations, and when data for different conditions are drawn from different sets of studies, participants have not been assigned randomly to these conditions. Also, assignment to a no-treatment or wait-list control group is not the same as no intervention at all. Therefore, our estimates of the placebo effect and natural history component of the response to antidepressant medication should be considered tentative. Nevertheless, when direct comparisons are not available, these comparisons provide the best available estimates of comparative effectiveness. Furthermore, in at least some cases, these estimates have been found to yield results that are comparable to those derived from direct comparisons of groups that have been randomly assigned to condition (Kirsch, 1990; Shapiro & Shapiro, 1982).

Unlike our estimate of the effect of natural history as a component of the drug response, our estimate of the placebo response as a proportion of the drug response was derived from studies in which participants from the same population were assigned randomly to drug and placebo conditions. Therefore, the estimate that only 25% of the drug response is due to the administration of an active medication can be considered reliable. Confidence in the reliability of this estimate is enhanced by the exceptionally high correlation between the drug response and the placebo response. This association is high enough to suggest that any remaining variance in drug response is error variance associated with imperfect reliability of measurement. Examining estimates of active drug and inactive placebo responses as a function of drug type further enhances confidence in the reliability of these estimates. Regardless of drug type, the inactive placebo response was approximately 75% of the active drug response.

We used very stringent criteria in selecting studies for inclusion in this meta-analysis, and it is possible that data from a broader range of studies would have produced a different outcome. However, the effect size we have calculated for the medication effect (D = .39) is comparable to those reported in other meta-analyses of antidepressant medication (e.g., Greenberg et al., 1992, 1994; Joffe, Sokolov, & Streiner, 1996; Quality Assurance Project, 1983; Smith et al., 1980; Steinbrueck, Maxwell, & Howard, 1983). Comparison with the Joffe et al. (1996) meta-analysis is particularly instructive, because that study, like ours, included estimates of pre–post effect sizes for both drug and placebo. Although only two studies were included in both of these meta-analyses and somewhat different calculation methods were used, [2] their results were remarkably similar to ours. They reported mean pre–post effect sizes of 1.57 for medication and 1.02 for placebo and a medication versus placebo effect size of .50.

Our results are in agreement with those of other meta-analyses in revealing a substantial placebo effect in antidepressant medication and also a considerable benefit of medication over placebo. They also indicate that the placebo component of the response to medication is considerably greater than the pharmacological effect. However, there are two aspects of the data that have not been examined in other meta-analyses of antidepressant medication. These are (a) the exceptionally high correlation between the placebo response and the drug response and (b) the effect on depression of active drugs that are not antidepressants. Taken together, these two findings suggest the possibility that antidepressants might function as active placebos, in which the side-effects amplify the placebo effect by convincing patients of that they are receiving a potent drug.

In summary, the data reviewed in this meta-analysis lead to a confident estimate that the response to inert placebos is approximately 75% of the response to active antidepressant medication. Whether the remaining 25% of the drug response is a true pharmacologic effect or an enhanced placebo effect cannot yet be determined, because of the relatively small number of studies in which active and inactive placebos have been compared (Fisher & Greenberg, 1993). Definitive estimates of placebo component of antidepressant medication will require four arm studies, in which the effects of active placebos, inactive placebos, active

medication, and natural history (e.g., wait-list controls) are examined. In addition, studies using the balanced placebo design would be of help, as these have been shown to diminish the ability of subjects to discover the condition to which they have been assigned (Kirsch & Rosadino, 1993).

References

Andrews, G., & Harvey, R. (1981). Does psychotherapy benefit neurotic patients? A reanalysis of the Smith, Glass, and Miller data. *Archives of General Psychiatry, 36,* 1203-1208.

Bailey, J., & Coppen, A. (1976). A comparison between the Hamilton Rating Scale and the Beck Depression Inventory in the measurement of depression . *British Journal of Psychiatry, 128,* 486-489.

Beach, S. R. H., & O'Leary, K. D. (1992). Treating depression in the context of marital discord: Outcome and predictors of response of marital therapy versus cognitive therapy. *Behavior Therapy, 23,* 507-528.

Beck, J. T., & Strong, S. R. (1982). Stimulating therapeutic change with interpretations: A comparison of positive and negative connotation. *Journal of Counseling Psychology, 29*(6), 551-559.

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561-571.

Blashki, T. G., Mowbray, R., & Davies, B. (1971). Controlled trial of amytriptyline in general practice. *British Medical Journal, 1,* 133-138.

Byerley, W. F., Reimherr, F. W., Wood, D. R., & Grosser, B. I. (1988). Fluoxetine, a selective serotonine uptake inhibitor for the treatment of outpatients with major depression. *Journal of Clinical Psychopharmacology, 8,* 112-115.

Catanese, R. A., Rosenthal, T. L., & Kelley, J. E. (1979). Strange bedfellows: Reward, punishment, and impersonal distraction strategies in treating dysphoria. *Cognitive Therapy and Research, 3*(3), 299-305.

Claghorn, J. L., Kiev, A., Rickels, K., Smith, W. T., & Dunbar, G. C. (1992). Paroxetine versus placebo: A double-blind comparison in depressed patients. *Journal of Clinical Psychiatry, 53*(12), 434-438.

Comas-Diaz, L. (1981). Effects of cognitive and behavioral group treatment on the depressive symptomatology of Puerto Rican women. *Journal of Consulting and Clinical Psychology, 49*(5), 627-632.

Conoley, C. W., & Garber, R. A. (1985). Effects of reframing and self-control directives on loneliness, depression, and controllability. *Journal of Counseling Psychology, 32*(1), 139-142.

Davidson, J., & Turnbull, C. (1983). Isocarboxazid: Efficacy and tolerance. *Journal of Affective Disorders, 5,* 183-189.

Davis, J. M., Janicak, P. G., & Bruninga, K. (1987). The efficacy of MAO inhibitors in depression: A meta-analysis. *Psychiatric Annals, 17*(12), 825-831.

Dobson, K. S. (1989). A meta-analysis of the efficacy of cognitive therapy for depression. *Journal of Consulting and Clinical Psychology, 57*(3), 414-419.

Doogan, D. P., & Caillard, V. (1992). Sertaline in the prevention of depression. *British Journal of Psychiatry, 160,* 217-222.

Elkin, I., Shea, M. T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., Glass, D. R., Pilkonis, P. A., Leber, W. R., Docherty, J. P., et al. (1989). National Institute of Mental Health, Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Archives of General Psychiatry, 46*(11), 971-982.

Evans, F. J. (1974). The placebo response in pain reduction. In J. J. Bonica (Ed.), *Advances in neurology: Vol. 4. Pain* (pp. 289-296). New York: Raven.

Feldman, D. A., Strong, S. R., & Danser, D. B. (1982). A comparison of paradoxical and nonparadoxical interpretations and directives. *Journal of Counseling Psychology, 29,* 572-579.

Fisher, S., Lipman, R.S., Uhlenhuth, E.H., Rickels, K., and Park, L.C. (1965). Drug effects and initial severity of symptomatology. *Psychopharmacologia, 7,* 57-60.

Fisher, S., & Greenberg, R. P. (1993). How sound is the double-blind design for evaluating psychiatric drugs? *Journal of Nervous and Mental Disease, 181,* 345-350.

Frank, J. D. (1973). *Persuasion and healing* (rev. ed.). Baltimore: Johns Hopkins.

Free, M. L., & Oei, T. P. S. (1989). Biological and psychological processes in the treatment and maintenance of depression. *Clinical Psychology Review, 9,* 653-688.
Goldberg, H. L., Rickels, K., & Finnerty, R. (1981). Treatment of neurotic depression with a new antidepressant. *Journal of Clinical Psychopharmacology, 1*(6), 35S-38S (Supplement).
Graff, R. W., Whitehead, G. I., & LeCompte, M. (1986). Group treatment with divorced women using cognitive–behavioral and supportive–insight methods. *Journal of Counseling Psychology, 33,* 276-281.
Greenberg, R. P., Bornstein, R. F., Greenberg, M. D., & Fisher, S. (1992). A meta-analysis of antidepressant outcome under "blinder" conditions. *Journal of Consulting and Clinical Psychology, 60,* 664-669.
Greenberg, R.P., Bornstein, R.F., Zborowski, M.J., Fisher, S., & Greenberg, M.D. (1994). A meta-analysis of fluoxetine outcome in the treatment of depression. *Journal of Nervous and Mental Disease, 182,* 547-551.
Greenberg, R. P., & Fisher, S. (1989). Examining antidepressant effectiveness: Findings, ambiguities, and some vexing puzzles. In S. Fisher & R. P. Greenberg (Eds.) *The limits of biological treatments for psychological distress.* Hillsdale, NJ: Erlbaum.
Greenberg, R. P., & Fisher, S. (1997). Mood-mending medicines: Probing drug, psychotherapy, and placebo solutions. In S. Fisher & R. P. Greenberg (Eds.), *From placebo to panacea: Putting psychiatric drugs to the test* (pp. 115-172). New York: Wiley.
Hamilton, M. A. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23,* 56-61.
Hedges, L. V., & Olkin, I. (1995). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.
Jarvinen, P. J., & Gold, S. R. (1981). Imagery as an aid in reducing depression. *Journal of Clinical Psychology, 37*(3), 523-529.
Joffe, R. T., Singer, W., Levitt, A. J., & MacDonald, C. (1993). A placebo controlled comparison of lithium and triiodothyronine augmentation of tricyclic antidepressants in unipolar refractory depression. *Archives of General Psychiatry, 50,* 387-393.
Joffe, R., Sokolov, S., & Streiner, D. (1996). Antidepressant treatment of depression: A metaanalysis. *Canadian Journal of Psychiatry, 41,* 613-616.
Khan, A., Dager, S. R., Cohen, S., et al. (1991). Chronicity of depressive episode in relation to antidepressant-placebo response. *Neuropsychopharmacology, 4,* 125-130.
Kiev, A., & Okerson, L. (1979). Comparison of the therapeutic efficacy of amoxapine with that of imipramine: A controlled clinical study in patients with depressive illness. *Clinical Trials Journal, 16*(3), 68-72.
Kirsch, I. (1985). Response expectancy as a determinant of experience and behavior. *American Psychologist, 40,* 1189-1202.
Kirsch, I. (1990). *Changing expectations: A key to effective psychotherapy*. Pacific Grove, CA: Brooks/Cole.
Kirsch, I. (1997). Specifying nonspecifics: Psychological mechanisms of placebo effects. In A. Harrington (Ed.), *The placebo effect: An interdisciplinary exploration* (pp. 166-186). Cambridge, MA: Harvard University Press.
Kirsch, I., & Rosadino, M. J. (1993). Do double-blind studies with informed consent yield externally valid results? An empirical test. *Psychopharmacology, 110,* 437-442.
Lydiard, R. B. et al. (1989). Fluvoxamine, imipramine and placebo in the treatment of depressed outpatients. *Psychopharmacology Bulletin, 25*(1), 63-67.
Margraf, J., Ehlers, A., Roth, W. T., Clark, D. B., Sheikh, J., Agras, W. S., & Taylor, C. B. (1991). How "blind" are double-blind studies? *Journal of Consulting and Clinical Psychology, 59,* 184-187.
Maynard, C. K. (1993). Comparisons of effectiveness of group interventions for depression in women. *Archives of Psychiatric Nursing, 7*(5), 277-283.
Ney, P. G., Collins, C., & Spensor, C. (1986). Double blind: Double talk or are there ways to do better research? *Medical Hypotheses, 21,* 119-126.
Nezu, A. M. (1986). Efficacy of a social problem solving therapy approach for unipolar depression. *Journal of Consulting and Clinical Psychology, 54*(2), 196-202.

Quality Assurance Project. (1983). A treatment outline for depressive disorders. *Australian and New Zealand Journal of Psychiatry, 17,* 129-146.
Ravaris, C. L., Nies, A., Robinson, D. S., et al. (1976). A multiple-dose, controlled study of phenelzine in depression-anxiety states. *Archives of General Psychiatry, 33,* 347-350.
Rehm, L. P., Kornblith, S. J., O'Hara, M. W., et al. (1981). An evaluation of major components in a self control therapy program for depression. *Behavior Modification, 5*(4), 459-489.
Rickels, K., & Case, G. W. (1982). Trazodone in depressed outpatients. *American Journal of Psychiatry, 139,* 803-806.
Rickels, K., Case, G. W., Weberlowsky, J., et al. (1981). Amoxapine and imipramine in the treatment of depressed outpatients: A controlled study. *American Journal of Psychiatry, 138*(1), 20-24.
Robinson, L. A., Berman, J. S., & Neimeyer, R. A. (1990). Psychotherapy for the treatment of depression: A comprehensive review of controlled outcome research. *Psychological Bulletin, 108,* 30-49.
Robinson, D. S., Nies, A., & Ravaris, C. L. (1973). The MAOI phenelzine in the treatment of depressive-anxiety states. *Archives of General Psychiatry, 29,* 407-413.
Rude, S. (1986). Relative benefits of assertion or cognitive self-control treatment for depression as a function of proficiency in each domain. *Journal of Consulting and Clinical Psychology, 54,* 390-394.
Schmidt, M. M., & Miller, W. R. (1983). Amount of therapist contact and outcome in a multidimentional depression treatment program. *Acta Psychiatrica Scandinavica, 67,* 319-332.
Schweizer, E., Feighner, J., Mandos, L. A., & Rickels, K. (1994). Comparison of venlafaxine and imipramine in the acute treatment of major depression in outpatients. *Journal of Clinical Psychiatry, 55*(3), 104-108.
Shapiro, D. A., & Shapiro, D. (1982). Meta-analysis of comparative therapy outcome studies: A replication and refinement. *Psychological Bulletin, 92,* 581-604.
Shaw, B. F. (1977). Comparison of cognitive therapy and behavior therapy in the treatment of depression. *Journal of Consulting and Clinical Psychology, 45,* 543-551.
Shipley, C. R., & Fazio, A. F. (1973). Pilot study of a treatment for psychological depression. *Journal of Abnormal Psychology, 82,* 372-376.
Sloane, R. B., Staples, F. R., Cristol, A. H., Yorkston, N. J., & Whipple, K. (1975). *Psychotherapy versus behavior therapy*. Cambridge, MA: Harvard University Press.
Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore: Johns Hopkins University Press.
Stark, P., & Hardison, C. D. (1985). A review of multicenter controlled studies of fluoxetine vs. imipramine and placebo in outpatients with major depressive disorder. *Journal of Clinical Psychiatry, 46,* 53-58.
Steinbrueck, S.M., Maxwell, S.E., & Howard, G.S. (1983). A meta-analysis of psychotherapy and drug therapy in the treatment of unipolar depression with adults. *Journal of Consulting and Clinical Psychology, 51,* 856-863.
Taylor, F. G., & Marshall, W. L. (1977). Experimental analysis of a cognitive–behavioral therapy for depression. *Cognitive Therapy and Research, 1*(1), 59-72.
Tyson, G. M., & Range, L. M. (1987). Gestalt dialogues as a treatment for depression: Time works just as well. *Journal of Clinical Psychology, 43,* 227-230.
van der Velde, C. D. (1981). Maprotiline versus imipramine and placebo in neurotic depression. *Journal of Clinical Psychiatry, 42,* 138-141.
White, K., Razani, J., Cadow, B., et al. (1984). Tranylcypromine vs. nortriptyline vs. placebo in depressed outpatients: a controlled trial. *Psychopharmacology, 82,* 258-262.
Wierzbicki, M., & Bartlett, T. S. (1987). The efficacy of group and individual cognitive therapy for mild depression. *Cognitive Therapy and Research, 11*(3), 337-342.
Wilson, P. H., Goldin, J. C., & Charboneau-Powis, M. (1983). Comparative efficacy of behavioral and cognitive treatments of depression. *Cognitive Therapy and Research, 7*(2), 111-124.
Workman, E. A., & Short, D. D. (1993). Atypical antidepressants versus imipramine in the treatment of major depression: A meta-analysis. *Journal of Clinical Psychiatry, 54*(1), 5-12.
Zung, W. W. K. (1983). Review of placebo-controlled trials with bupropion. *Journal of Clinical Psychiatry, 44*(5), 104-114.

[1] A reviewer suggested that because effect sizes are essentially *z*-scores in a hypothetically normal distribution, one might use percentile equivalents when examining the proportion of the drug response duplicated by the placebo response. As an example of why this should not be done, consider a treatment that improves intelligence by 1.55 *SD*s (which is approximately at the 6th percentile) and another that improves it by 1.16 *SD*s (which is approximately at the 12th percentile). Our method indicates that the second is 75% as effective as the first. The reviewer's method suggests that it is only 50% as effective. Now let's convert this to actual IQ changes and see what happens. If the IQ estimates were done on conventional scales (*SD* = 15), this would be equivalent to a change of 23.25 points by the first treatment and 17.4 points by the second. Note that the percentage relation is identical whether using *z*-scores or raw scores, because the *z*-score method simply divides both numbers by a constant.

[2] Instead of dividing mean differences by the pooled *SD*s, Joffe et al. (1996) used baseline *SD*s, when these were available, in calculating effect sizes. When baseline *SD*s were not available, which they reported to be the case for most of the studies they included, they used estimates taken from other studies. Also, they used a procedure derived from Hedges and Olkin (1995) to weight for differences in sample size, whereas we used the more straightforward method recommended by Hunter and Schmidt (1990).