The Limits of Biological Treatments for Psychological Distress

Comparisons with Psychotherapy and Placebo
The Limits of Biological Treatments for Psychological Distress

Comparisons with Psychotherapy and Placebo
The Limits of Biological Treatments for Psychological Distress

Comparisons with Psychotherapy and Placebo

Edited by
Seymour Fisher
Roger P. Greenberg
State University of New York
Health Science Center, Syracuse
To my family: Drs. Rhoda, Jerid, and Eve Fisher—
and Dr. Mark Whitmore
With further regards to Dr. Hippocrates who recognized
the temptation to be too aggressively therapeutic.
S.F.

To my wife Vicki and my son Michael who—
without benefit of statistics—keep me
aware of what is meaningful in life.
R.P.G.
CONTENTS

PREFACE xi

1 EXAMINING ANTIDEPRESSANT EFFECTIVENESS: FINDINGS, AMBIGUITIES, AND SOME VEXING PUZZLES
Roger P. Greenberg and Seymour Fisher 1

2 MEASUREMENT ISSUES IN THE EVALUATION OF PSYCHOPHARMACOLOGICAL THERAPY
Edward J. Murray 39

3 PHARMACOTHERAPY OF THE ANXIETY DISORDERS
Ronald S. Lipman 69

4 PSYCHOTHERAPY VERSUS MEDICATION FOR SCHIZOPHRENIA: EMPIRICAL COMPARISONS
Bertram P. Karon 105

5 ATTENTION DEFICIT DISORDER: THE EMPEROR'S CLOTHES, ANIMAL “PHARM,” AND OTHER FICTION
Diane McGuinness 151

vii
CONTENTS

6  THE CLINICAL IMPACT OF THE SIDE EFFECTS OF PSYCHOTROPIC DRUGS
    Mantosh J. Dewan and Marvin Koss 189

7  PERSONALITY FACTORS IN THE MEDIATION OF DRUG RESPONSE
    Sidney E. Cleveland 235

8  ANALYSIS OF STATISTICAL PROCEDURES AND DESIGNS COMMONLY USED IN DRUG RESEARCH STUDIES
    Silas Halperin 263

9  THE EFFICACY OF ELECTROCONVULSIVE THERAPY IN THE TREATMENT OF MAJOR DEPRESSIVE DISORDER
    Harold A. Sackeim 275

10 A SECOND OPINION: RETHINKING THE CLAIMS OF BIOLOGICAL PSYCHIATRY
    Seymour Fisher and Roger P. Greenberg 309

    AUTHOR INDEX 337

    SUBJECT INDEX 361
Sidney E. Cleveland • Baylor College of Medicine, Houston, TX 77024
Montosh J. Dewan • Department of Psychiatry, State University of New York Health Science Center, and Veterans Administration Medical Center, Syracuse, NY 13210
Seymour Fisher • Department of Psychiatry, State University of New York Health Science Center, Syracuse, NY 13210
Roger P. Greenberg • Department of Psychiatry, State University of New York Health Science Center, Syracuse, NY 13210
Silas Halperin • Department of Psychology, Syracuse University, Syracuse, NY 13210
Bertram P. Karon • Department of Psychology, Michigan State University, East Lansing, MI 48824
Marvin Koss • Department of Psychiatry, State University of New York Health Science Center, and Veterans Administration Medical Center, Syracuse, NY 13210
Ronald S. Lipman • Department of Research and Evaluation, Friends Hospital, Philadelphia, PA 19124
Diane McGuinness • University of South Florida, Fort Myers, FL 33913
Edward J. Murray • Department of Psychology, University of Miami, Miami, FL 33158
Harold A. Sackeim • Department of Psychiatry, Columbia University, New York, NY 10032
This book is devoted to finding out how effective the somatic therapies widely used in treating “mental disorders” truly are. Vast quantities of effort and money are being channeled into drug therapies of various kinds. It has been estimated that 10% to 12% of the adult population in the United States uses a psychoactive drug at least once during a calendar year (Klerman, 1986). We feel there are quasi-mythic images circulating concerning the power of “biological psychiatry” that need to be examined. Issues pertaining to therapeutic efficacy are always politicized. The resident therapeutic experts on the scene are often biased because they have already invested their reputations in claims advertised to large numbers of patients. Our intention is to examine as objectively and fairly as possible the existing scientific data bearing on the value of the somatic treatment modes dominant in 20th century psychiatry. The need for such evaluation is pointed up by Valenstein’s (1986) book Great and Desperate Cures, which depicts the eventual calamitous results for psychiatry of launching inadequate scientific appraisals of such treatments as lobotomy and insulin shock therapy. He suggests that the same forces that fostered these treatments, that in retrospect were grossly irrational, “are still active today.”

In an earlier enterprise we (Fisher & Greenberg, 1985) undertook to evaluate the scientific credibility of many of Freud’s primary psychodynamic notions and also the efficacy of the psychoanalytic therapy he devised. This proved to be a complicated enterprise, but we eventually came up with

\footnote{However, one major category, viz., bipolar disturbance, was not examined.}
definitive conclusions concerning which areas of Freud's work were or were not scientifically sound. The present volume represents an analogous probe into the somatic armamentarium for "mental disorders." The time is propitious for a broad survey of the somatic therapies being offered to people who are psychologically in a state of disequilibrium. Which of these therapies are actually most dependable? How much better are they than nonsomatic approaches? What are the physiological costs and threats associated with them? Are there any exaggerations or distortions in the therapeutic claims being made? Where do we really stand at this point in time with reference to the proven advantages of the somatic therapies?

Biological treatment approaches for psychological problems have several built-in seductive appeals for both practitioners and patients that could color objective evaluation. Obviously, there is a need to do something when people feel disabled by their emotions and behaviors. Faced with a cry for help from an individual displaying puzzling or seemingly irrational behavior, clinicians may feel more secure knowing they can point to a relatively easily applied procedure, such as pill taking, to comply with the demand that something be done. The biological perspective has an evident attraction for patients too, in that it suggests that they are not accountable for the creation of their symptoms (because symptoms are due to chemical imbalances) and that they are not responsible for playing an active role in the solution to their discomforts. Dressed in the cloak of hard science, biological treatment approaches also radiate an aura of precision and specificity that is not as readily associated with the seemingly softer behavioral sciences. It is easy to lose sight of the fact that although the treatments can be delivered with exact measurement of chemical composition and dosage level, the results are being gauged in terms of feelings and judgments about behavior. The medicalized treatments with their images of hard tech, hard science control, quiet the unpleasant idea of people's actions being governed by irrational, disordered feelings, thoughts, and impulses. We raise these ideas not to discount the possible benefits of somatic approaches, but to point out how the pressing demand for simple, comfortable, blame-free treatments could skew evaluations of their effectiveness.

Our strategy in evaluating the therapies had been deliberately provocative. We wanted to take nothing for granted and to feel free to raise questions no matter how naive they might appear to be. Our assumption is that there is good reason to approach any established structure of therapeutic practices with skepticism and the expectation that the "Emperor's clothes" may not be what they appear to be. The long-term practitioners of therapeutic modes are inclined to become complacent and to encrust themselves with exaggerated claims. We assume that probing for weakness and looking for cracks in treatment rationales will eventually benefit consumers as well as practitioners of such treatments. Our selection of contributors to this volume was guided
by our intent to be fresh and questioning. We chose persons whom we considered to be challenging and at least a bit anti-authoritarian. But as a first priority, we chose persons who had a proven track record of scholarly capability and who were expertly acquainted with the fundamental issues involved in evaluating treatment outcomes. Incidentally, we wanted contributors who were not too identified with or invested in treatment roles. Although several were engaged in clinical practice, they also had major work investments outside the clinical realm.

We called upon our contributors to burrow into the research pertinent to most of the heralded somatic approaches current in the treatment of psychologically disturbed individuals. A number of the major drug treatments are critiqued, as is electric shock therapy. In the drug area, we (Greenberg and Fisher) embark upon a broad appraisal of the scientific literature concerned with the power of antidepressant agents. We also take an excursion into the realm of placebos and highlight the puzzling dilemma of whether to use active or inactive placebos in drug research trials. Edward Murray focuses microscopically upon the outcome measures customarily employed to evaluate the effects of antidepressants and points up the advantages and disadvantages of a number of them. Ronald Lipman examines the evidence bearing on the efficacy of the antianxiety agents, and Bertram Karon compares the outcomes deriving from treatment of schizophrenics with drugs as contrasted with psychotherapy. Diane McGuinness probes the vast literature dealing with the use of stimulant drugs to cope with the so-called hyperactive or Attention Deficit child. Mantosh Dewan and Marvin Koss present us with a wide-ranging survey of the physiological side effects of many of the major therapeutic drugs. At still another level, Sidney Cleveland pulls together the literature concerned with the psychological mediators of drug effects. Further, Silas Halperin takes a critical look at the statistical procedures widely used in studies of antianxiety and antidepressant agents. Harold Sackeim completes the overall project by reviewing what is known about the therapeutic power of electric convulsive therapy. In a final overview we survey the major conclusions of the various chapters and ponder a number of urgent questions and issues that emerge.

As is seen, the evaluation of treatment modes is a tricky matter. The scientific literature bearing on treatment outcomes is not only of great volume, but also often surprisingly vague, contradictory, and susceptible to multiple interpretations. It does not take much bias either for or against a drug to slant one's interpretations of the available pertinent data. It is all too easy to dismiss one study because it has some "defect" (no matter how minor) or to accept another because it "so clearly" demonstrates a positive or negative effect. If one is to be honest, one must from the outset admit that bias is rampant. It is probably impossible to interpret the current mass of treatment outcome findings without being pulled by hidden agendas. Let the reader and also the
consumer of somatic treatments beware. At this stage of the game, claims and conclusions should be treated as tentative. Even when one is able to discern that a particular treatment is “significantly” more effective than a placebo control, the nagging question remains as to how large the treatment advantage must be to represent a viable choice for a real life person in distress. Considering the financial costs and not inconsiderable and uncomfortable side effects of most somatic treatments, is a 20% advantage over placebo sufficient? Or does the advantage have to be 30% or 40%? How consistently must the scientific literature demonstrate a given magnitude of advantage before it has real life implications? What if 50% of the studies indicate a 40% advantage, but 50% show only a 20% advantage or less? What if more recent studies indicate less of an advantage than did earlier ones?

This leads into the question of what consumers should be told about the therapeutic potential of any somatic treatment offered to them. Obviously, they cannot be apprised of the full complexity of the accumulated research literature. Should they be expected to take on faith that “This will help you”? Should there be at least a rough quantitative statement of the probable advantage of the treatment as compared to doing nothing? Or as compared to alternative forms of treatment? There is little consensus about such questions. Many think the consumer should be willing to put his faith in the “expert clinician.” Many worry that giving too much information to the consumer will destroy potentially important placebo effects. Obviously, such views fall outside of the purview of science and represent political and ethical schemas.

Beyond the confusion already mentioned, we also ask the reader to wonder a bit about the scientific security of present methods for testing therapeutic effectiveness. We think there is evidence of a false sense of assurance about the safeguards provided by the double blind or other variant paradigms. We are not the first to discern cracks in apparently blind designs. However, past criticisms have simple been superficially acknowledged and then substantially ignored. The problem may be more serious than we have been willing to admit. We had better shake off our inertia and attend to a matter that has potentially serious unsettling implications for what we think we know about the therapeutic power of various substances and procedures.

REFERENCES


EXAMINING ANTIDEPRESSANT EFFECTIVENESS: FINDINGS, AMBIGUITIES, AND SOME VEXING PUZZLES

Roger P. Greenberg
Seymour Fisher
State University of New York
Health Science Center

Since Kuhn (1958) detected an apparent antidepressant effect for imipramine, many studies have explored this phenomenon. Despite some past skeptical opinions (Jenner, 1977; Porter, 1970; Wechsler, Grosser, & Greenblatt, 1965), it is now widely accepted that the therapeutic effectiveness of the antidepressants, especially the tricyclics, has been indubitably demonstrated. However, although this chapter does not question that antidepressants are therapeutic, it does question the magnitude of the effect and some of its underlying causes. As we surveyed the literature dealing with antidepressants, we detected some inconsistencies and methodological gaps in current claims and clinical practices. The history of treatment modes for psychological disturbance bristles with examples (e.g., insulin therapy, lobotomy) that were widely accepted and ultimately proved to fall short of early claims. This is, of course, also true in other areas of treatment (Benson & McCallie, 1979). A cautious attitude about treatment claims is dictated by what has gone before. Among other things, this chapter focuses on the way in which the double-blind design has typically been used to evaluate the therapeutic value of the antidepressants and suggests that it has not been applied satisfactorily. More specifically, our intent is the following:

1. To appraise the effectiveness of the antidepressants across multiple reviews and studies. It should be noted, however, that the appraisal restricts itself to the effectiveness for relieving depression and does not deal with other symptom categories to which antidepressants have been applied.

2. To examine the stability of findings involving the antidepressants, with
a special focus on whether reported levels of effectiveness have changed in more recent appraisals.

3. To compare the effectiveness of antidepressants to the outcomes for psychotherapies specifically designed to treat depression.

4. To probe the objectivity of the typical double-blind design employed to evaluate antidepressants, especially in relation to the issue of using placebos that are inactive.

5. To offer suggestions of possible ways of balancing deficiencies in current approaches to measuring the effectiveness of antidepressants and other therapeutic agents.

PAST REVIEWS OF ANTIDEPRESSANT DRUG EFFICACY

Since the introduction of antidepressant drugs, many studies have attempted to assess their effectiveness, and a variety of reviewers have tried their hand at summarizing the reports that have appeared in the literature. The reviewers have focused on the antidepressant effects of tricyclic compounds and monoamine oxidase inhibitors (MAOIs). Obviously there is overlap in the studies summarized by different authors. However, there is occasional disagreement among reviewers in categorizing the same investigation as showing evidence for or against drug use. Results have generally been compressed into either box scores comparing the number of studies showing drugs to be superior to placebos versus those showing no difference in outcome or compilations of the percentage of patients significantly improving on drugs as opposed to placebos. A search of the literature revealed 15 such reviews. Six assessed the drug effects as relatively positive and superior to placebos and 7 suggested more modest, cautious, or equivocal conclusions about drug effects. Two reviews took a somewhat different tack and attempted to measure the degree to which groups treated with antidepressant drugs have exceeded non-drug control groups. Note that these 15 reviews were largely written in the 1960s.

Perhaps a presentation of the positive reviews should begin with the work of Cole (1964) who, in providing an early impressionistic discussion of drug treatments, noted that two thirds of 15 placebo-controlled studies of depressed inpatients showed imipramine to be superior. Three placebo-controlled studies of outpatients also declared imipramine to be the more effective treatment. Similarly, Davis (1965) detailed a box score account of 47 antidepressant drug studies. Most of these reports were placebo controlled and double-blind. The drug was declared superior to placebos in 68% of the studies. At a later date, Davis, Klerman, and Schildkraut (1968) tabulated box scores for 52 double-blind placebo-controlled studies of tricyclics and 28 similarly controlled inves-
tigations of MAOIs. Drugs were declared superior to placebos in 79% of the tricyclic studies and in 54% of the MAOI studies.

Klerman and Cole (1965) found 23 studies comparing imipramine to placebo. Eighteen of these studies (78%) indicated that the drug outcome was better. However, the authors of 3 of these 18 positive papers felt that their investigations as a whole did not present really convincing evidence of imipramine's general superiority to the controls. Overall, the authors of the studies felt that the drug was convincingly superior 15 times in the 23 trials (65%). Incidentally, even this figure may be an overestimation of the number of trials in which imipramine exceeded placebo. Thus, a commentary (Beck, 1967) on this review indicated that many studies were counted by Klerman and Cole as showing positive results even though the superiority of the drug was slight and short of statistical significance on a number of indices.

Klerman and Cole also combined the data across all the studies, permitting an overall judgment as to whether the patients had or had not improved. They concluded that 65% of the 550 imipramine-treated patients improved, whereas 31% of the 459 controls evidenced similar improvement. They went on to suggest that this improvement difference between drugs and placebo might not be as large as the percentages indicated because some of the improved patients were not improved enough to satisfy either themselves or their physicians. Therefore, imipramine was deemed "not entirely satisfactory treatment in many of the depressed patients to whom it was given" (p. 282).

Klein and Davis (1969) reviewed 65 studies comparing tricyclics with placebo and indicated that 50 showed the drug to be superior (77%). They found no striking overall outcome differences among the tricyclics. Although there were differences in diagnostic groupings and types of measurement among studies, it was felt that combining samples might give an approximate summary statement of the efficacy of antidepressants. Therefore, the percentages of improvement reported by the various studies were pooled. The number of patients rated at least moderately improved was combined within the drug and placebo groups respectively and compared with the number found to be slightly improved, unchanged, or worse. In this way the reviewers derived a figure of 70% as the rate of improvement on imipramine and 39% as the placebo improvement rate. Overall it was concluded that imipramine was superior to a placebo, although "not overwhelmingly so," with a treatment superiority of about 30% more improvement for the active drug than for the placebo. Similarly, in fewer studies, amitriptyline was found to be more beneficial than placebo (62% vs. 24%).

In a more recent, widely cited review, Morris and Beck (1974) looked at 146 double-blind studies on antidepressant drugs utilized in the United States in 1972. Two thirds of the 93 studies of tricyclics indicated that the drug was superior to a placebo, whereas one third of the reports found no difference. The box score outcome for the MAOIs revealed somewhat less success, with
an overall superiority over placebo in only 33% of the papers. When MAOI results were limited to only FDA approved drugs, the success rate rose to 61% in favor of the drugs, but Morris and Beck cautioned that this average value could be misleading in view of the considerable variability in efficacy among the MAOIs.

Although the tone of the six reviews cited to this point indicates a superiority of the antidepressant drugs over placebos, seven other reviews indicate more cautious or equivocal conclusions. Brady, in a 1963 review of placebo-controlled studies of imipramine, found that 52% (13 of 25 studies) of these comparisons yielded no significant difference in outcome (cited in Beck, 1967). Atkinson and Ditman (1965) examined 16 reports where the patients were given similar depressive diagnoses and in which an MAOI (tranylcypromine) was used as the sole treatment. Ten of the 16 studies employed a pre-post design without control groups. Study outcomes were divided into three groups: favorable, equivocal, or unfavorable. Of the comparisons, 68% were deemed equivocal or unfavorable.

Friedman, Granick, Cohen, and Cowitz (1966) located 21 reports appearing between 1957 and 1964 of double-blind controlled studies comparing imipramine with placebo on hospitalized endogenous or psychotic depressed patients. Eleven of the 21 studies claimed that imipramine was superior to a placebo. However, Friedman et al. stated that on close inspection even the 11 studies reporting positive effects for the drug with psychotic depressives were inconclusive. Only 2 of the 11 studies claimed the results were statistically significant and patients in the “positive” studies turned out to be a mixed sample of endogenous and nonendogenous depressives. In short, Friedman et al. could not find a single double-blind controlled study that demonstrated clearly favorable results with a sample of hospitalized psychotic depressives. Overall, Friedman and his colleagues concluded that the matter of treating endogenous or psychotic depression with imipramine is “unsettled.” Their investigation of this question in an additional carefully done study presented in the same paper indicated that if all outcome indices were considered, there was equivalent relief from depression when hospitalized psychotic depressives were treated with either imipramine or an active placebo.

McNair (1974), in studying how frequently different self-rating scales indicate antidepressant effectiveness, analyzed the results of 72 research publications reporting on 75 largely antidepressant drug trials (almost all double-blind) published between 1955 and 1972. The studies produced an average of six comparisons per trial by utilizing a variety of measures with various subscores and sometimes having more than one evaluation period. Significant treatment effects were found in only 21% of the 451 comparisons. Although this represents a greater number of significant effects than would be expected by chance, the results are modest. Another overview suggesting conservative conclusions was presented by Rogers and Clay (1975), who examined 30
trials comparing imipramine to placebo. They used the global ratings of improvement in each of the studies and submitted each to Fisher's Exact Test. Although the authors attempted to report results in a sympathetic manner, they found drugs superior to placebo in only 10 of 30 comparisons (33%).

Wechsler et al. (1965), in a review of 103 publications that appeared over a 5-year period, raised questions about the actual magnitude of the antidepressant drug effect. They found the size of the effect to be related to the type of research design. Treatments were found to be more effective when placebo controls were not employed. The association between improvement rates and research design is highlighted by viewing the work on imipramine, the treatment studied most frequently. Imipramine was found to be at least 65% effective in only 1 of the 9 studies comparing it to a placebo, as opposed to 7 out of 9 studies without a control group and 11 out of 17 reports comparing it to another active treatment.

A. Smith, Traganza, and Harrison (1969) presented a “comprehensive overview of antidepressant literature published in the English language.” More than 2,000 articles were screened resulting in a distillation of 490 trials of the efficacy of one or more drugs used in the treatment of depression. The authors discovered, as had Wechsler et al. (1965), that the sophistication of the research design was a dominant factor in determining the level of reported improvement. The more stringently controlled the study, the lower the reported drug improvement rate. Improvement rates for drugs were significantly lowered by either the presence of a control group or the use of blind techniques. Interestingly, increasing study sophistication had the opposite effect on placebo response. The more stringent the controls, the more improvement noted for placebo. Overall, the authors suggested that the methodology of antidepressant research is more significant than the drug being studied in determining the outcome of a clinical trial. They concluded that “the differences between the effectiveness of antidepressant drugs and placebo are not impressive” (p. 19). When studies were restricted to those having placebo controls and blind techniques, active medications had only about a 15% improvement advantage over placebo (with respective median improvement rates of 61.1% vs. 46.3%). Skepticism about the conclusions reached by most authors was raised by the finding that only 18% of the studies in the literature up to that time used statistical tests to decide whether the active drugs were effective. Most frequently (in 67% of the studies) authors simply used global improvement judgments (without statistical comparisons) to reach conclusions about efficacy. The literature review did not indicate what percentage of the time a statistically significant difference was found between drugs and placebos in studies that tested for a difference.

M. L. Smith, Glass, and Miller (1980), utilizing an effect size statistic, determined the degree to which drug treatment groups exceeded control groups on outcome measures drawn from a representative sample of published
studies. Their analysis of 75 antidepressant drug effect sizes indicated that the average antidepressant drug effect size is .40. This means that the average person treated with antidepressants is at the 66th percentile of the placebo control group or, conversely, that the average person treated with placebos will do better than 34% of those treated with antidepressants. The typical patient's standing on outcome variables was bettered by 16 percentile points because of taking drugs rather than a placebo. The fact that this is a relatively modest drug effect is underlined by the finding that of the three major types of psychotropic medications (antipsychotic, antidepressant, and antianxiety), antidepressants produced the smallest effects.

Another meta-analysis, reported by a group from Australia and New Zealand (Quality Assurance Project 1983), also analyzed the results for controlled trials of drug treatments for depression. Results were separately assessed for neurotic (69 studies) and endogenous depressions (54 studies) treated with tricyclics (or other approaches). The effect size (ES) for the treatment of neurotic depression with tricyclics (ES = .52) was slightly higher than the effect size obtained by Smith et al. (1980). A larger effect size was obtained for tricyclic treatment of endogenous depressions (ES = .79). However, the data indicated that the largest effect size for treating neurotic depression occurred with psychotherapy (ES = 1.00) and the largest effect size for dealing with endogenous depressions (where no comparisons with psychotherapy were made) occurred in studies of Electroconvulsive Therapy (ECT) (ES = .93). There were differences in criteria for including studies, handling data, and measuring outcome between the Australian and American meta-analyses, which may account for differential results. Oddly, the Australian/New Zealand analysis found no relationship between research design characteristics and effect size, whereas the M. L. Smith et al. (1980) project discovered—consistent with other researchers—that effect size was related to such design characteristics as degree of blindness in assessing outcome and randomness of assignment to treatment and control groups.

COMMENTARY ON THE REVIEWS

After looking at modern textbooks in psychiatry, a reading of the antidepressant drug literature may come as something of a surprise. Although textbooks frequently proclaim that drugs are a significant solution to the problem of depression, research findings help to place this conclusion in perspective. Even the most positive reviews indicate that 30% to 40% of the studies show no difference in response to drugs and placebos (M. L. Smith et al., 1980). In terms of the percentage of improvement, supporters of medication use suggest that about one third of patients do not improve with antidepressant treatment, one third improve with placebos, and an additional one third show
a response to medication that they would not have achieved with placebos. Thus, with the most positive outlook, about two thirds of the cases—placebo responders and those who do not respond to anything—would do as well or better with placebo treatment as they would do if treated by an active medication. Furthermore, as we have already noted, review evidence suggests that the average depressed person taking placebos will attain a better outcome than 34% of those individuals taking an active medication (M. L. Smith et al., 1980).

The strength of conclusions about antidepressant efficacy is of course limited by the strength of the data from which those conclusions are drawn, and there are a number of problems common to the bulk of studies reviewed by the investigators we have cited. Klein, Gittleman, Quitkin, and Rifkin (1980) have suggested that some of these problems may have decreased the size of the difference between drug-treated and placebo-treated groups. For example, the lack of clear consistent inclusion criteria for diagnosing depression may have created a heterogenous population in many studies and thereby obscured the unique benefits of drugs for certain types of depression. Yet over the years, reviewers of empirical work have been in disagreement as to whether drugs work best with any particular type of depressive group, such as neurotic or psychotic depressives (Beck, 1967; Bielski & Friedel, 1976; Friedman et al., 1966; Raskin & Crook, 1975), and some recent studies utilizing objective, reliable diagnostic criteria have not found any relationship between an endogenous–nonendogenous factor and the outcome of treatment (Blackburn, Bishop, Glen, Whalley, & Christie, 1981; Hersen, Bellack, Himmelhoch, & Thase, 1984; Rush, Beck, Kovacs, & Hollon, 1977).

Clearly, there have been major difficulties in trying to define homogenous diagnostic subgroups of depressed patients. This problem is highlighted by the work of Katschnig and his colleagues (see overview by Katschnig, Nutzinger, & Schanda, 1986) who found vagueness, looseness in terminology, and little overlap among nine different diagnostic systems purporting to distinguish between endogenous and neurotic depression. Furthermore, in their research on 176 depressed inpatients, none of the endogenous/nonendogenous distinctions made by any of the definitions showed an association with preceding life stress, illness course, or a number of outcome measures over a 2- to 3-year period following discharge.

On the other hand, Bielski and Friedel (1976), in reviewing studies predicting response to tricyclic antidepressants, did find a few reports citing evidence for a relationship between positive response to tricyclic medications and individual symptoms that have been associated with an endogenous diagnosis (e.g., anorexia, weight loss, and middle or late insomnia). Yet they note, with some surprise, that the conclusion regarding a relationship between drug response and endogenous symptoms is based on relatively few controlled studies.
Prusoff, Weissman, Klerman, and Rounsaville (1980), in speaking of the clinical folklore surrounding the subtype of depression labeled “endogenous,” underlined the difficulties in dealing with the concept because of a lack of consensus and imprecision about definition. In their work they used the Research Diagnostic Criteria to define depression subtypes. For most subtypes of depression they found no difference in outcome for four different types of treatment (e.g., interpersonal psychotherapy, drugs, a combination of interpersonal psychotherapy and drugs, or supportive psychotherapy on patient demand). Situational depressions (where there was an indication of a precipitating event) seemed to respond best to either drugs or psychotherapy (depending on the outcome measure) with the combination of both adding nothing to either treatment alone. In contrast, endogenous depressions responded better to a combination of psychotherapy and drugs than to drugs alone. In this study psychotherapy alone proved to be ineffective as a treatment for endogenous depression. Interestingly, the diagnoses used for this work were based on several independent symptoms. It was not known whether some symptoms were more predictive of response than others. The criteria used for diagnosis also allowed patients to be placed in more than one diagnostic category at a time. Unlike the widely accepted clinical lore, it was possible in this scheme for a patient's depression to be labeled both situational and endogenous. Of the 26 patients (out of 81) in this investigation who were diagnosed endogenous, 7 were also labeled as having situational depression.

It has been speculated, too, that some of the early studies did not use a high enough medication dose to attain maximum drug–placebo differences (Klein et al., 1980). Although a few researchers have supported this idea (Quitkin, Rabkin, Ross, & McGrath, 1984), there is room for debate on this issue because others, such as Wechsler and his colleagues (1965), found no relationship between the percentage of improvement reported in a large number of studies and either dosage level or the length of treatment. Similarly, the large-scale Australian and New Zealand meta-analysis of antidepressant outcome studies found no relationship between effect size (a measure of outcome magnitude) and dosage level (Quality Assurance Project, 1983). Despite much emphasis by some observers concerning the importance of using sufficiently high dosages to attain therapeutic effects (see overview by Quitkin, 1985), the truth is that little relevant empirical work comparing dosage levels has been done. We have found only two studies explicitly concerned with this issue. One (Watt, Crammer, & Elkes, 1972) did demonstrate a clear therapeutic superiority for a higher dose (150 vs. 300 mg of desmethylimipramine), although only a modest 50% of the patients improved on the higher dose and the raters of improvement were not blind to dosage level. Another (Simpson, Lee, Cuculic, & Kellner, 1976) actually indicated only a borderline advantage for a 300 mg dose of imipramine as compared to a 150 mg dose.
Research on dosage levels may be an inefficient means for gathering evidence on antidepressant efficacy because dosage and blood plasma levels of antidepressants do not appear to be highly related. However, the picture does not become totally clear if one attempts to associate plasma levels (rather than dosage levels) with clinical outcome. Although several studies have shown an association between plasma levels and outcome for some drugs, the magnitude of the relationship and its consistency have raised some concerns. Complicating the attempts to find evidence for a direct relationship between plasma level and outcome are wide interindividual biochemical and pharmacokinetic variations (e.g., Moller et al., 1985). As Simpson, Edmond, and White (1983) note after reviewing this area, “efforts to relate plasma levels to therapeutic outcome have, in general, been disappointing” (p. 27), and the relationships, although “extensively studied since 1962, remain controversial” (p. 29).

Research on the association of blood level concentrations of drugs and outcome has shown a moderate relationship in some clinical groups and no consistent relationship in others (APA Task Force, 1985; Glassman, Perel, Shostak, Kantor, & Fleiss, 1977; Reisby et al., 1977). One report indicated that plasma level was correlated with treatment response in only one third of the patients (Blackwell, 1982). Work with different drugs has also yielded different conclusions. Although there have not been many independent studies, the relationship between blood level measurement and clinical response has shown a degree of consistency with some drugs (e.g., imipramine) and none with others (e.g., amitriptyline) (APA Task Force, 1985). The size of the correlation found in studies that do report a significant relationship suggests that an individual’s blood level of medication accounts, at best, for only a minor part of the “drug effect” (Glassman et al., 1977; Reisby et al., 1977).

Friedel (1982), in concluding that only the antidepressants nortriptyline and imipramine have shown some relationship between plasma level and therapeutic response, suggests that the relationship holds only for those patients with an endogenous type of diagnosis. He wrote:

It is important to recognize that the reported relationships between antidepressant plasma levels and therapeutic response have been determined for the most part in patients with endogenous depression or, as defined in DSM-III, major depressive disorder with melancholia. Patients who have other depressive subtypes that are not typically responsive to the tricyclics, e.g., those with psychotic symptoms, atypical depressions, or neurotic and reactive depressions, cannot be expected to demonstrate a correlation between antidepressant plasma levels and therapeutic outcome. (p. 40)

However, even the conclusion that there is a relationship between clinical response and plasma concentration for endogenous depression has not received
consistent support. For example, a carefully done study on 90 inpatients (with 85% endogenous and 15% psychotic depression diagnoses) showed virtually no significant linear or curvilinear relationships between any measure of plasma concentration and any of several measures of clinical response to either imipramine or amitriptyline (Kocsis, Hanin, Bowden, & Brunswick, 1986).

There are some characteristics of drug studies that exaggerate the differential effects of drugs and placebo. Prominent among these is the type of study design. In general, the less blind the study participants and drug administrators are to whether drugs or placebos are being administered, the greater the drug–placebo differences become (M. L. Smith et al., 1980). This is a particularly important finding because, as we show, there is good reason to believe that most antidepressant drug studies are not conducted under truly double-blind conditions (Ainslie, Stiefel, & Jones, 1966). Relatedly, studies that rely on global impressionistic ratings of change are more likely to obtain significant drug–placebo outcome differences than are investigations making use of more objective, structured measures of psychological symptoms or adjustment (McNair, 1974; M. L. Smith et al., 1980). McNair found that global ratings were unique in yielding significant differences between antidepressant drugs and placebos and that no other type of measure showed differences even 20% of the time. He noted that global ratings do not necessarily reflect only changes in depressive symptomatology and that an individual may express any type of perceived change—such as in anxiety or sleep patterns—with such ratings. Similarly, Klerman and Cole (1965) discovered that significant differences between antidepressants and placebos dropped 42% when morbidity scores based on signs or symptoms rather than global ratings were used.

A word is in order at this point about a common practice in antidepressant drug research designs that artificially reduces apparent placebo response rates. It is by now a fairly standard procedure to initiate drug trials with a “washout” phase (single-blind, from 7 to 10 days) during which all prospective candidates for the study are placed on placebo, and those who show significant improvement are eliminated. Those who do not respond are then randomly assigned to the usual double-blind (drug versus placebo) design. One of the major purposes of the “washout” period is to “eliminate from the clinical trial patients whose symptoms remit, improve, or rapidly fluctuate within a short time span” (Rabkin et al., 1986, p. 274). Obviously, though, it also eliminates individuals who may be sensitive to the therapeutic impact of the placebo experience. One study (Rabkin et al., 1987) that systematically probed the positive response rate during washout preliminary to an antidepressant study found that the rate was in the 19%-20% range. Thus, a significant segment of the potential placebo responders was eliminated even before the formal comparison of placebo versus drug response was initiated. This means that the placebo response rates of all studies using washout phases may be substantially understating what the placebo response rate is in the real world population of
depressed persons seeking treatment. One could perhaps reason that if some of the variance of improvement in the actual drug group is due to placebo, the washout phase would ultimately reduce the improvement rate in that group too. That may well be, and there are complex questions that can be debated in this regard. But, in any case, the fact remains that in the context of washout procedures the formal statistics concerning explicit improvement rates in placebo control groups are probably seriously understated (compared to the real world).

Rabkin et al. (1987) investigated the characteristics of depressed patients who improve during the washout phase and found they are typified by “milder illness symptoms” than those who do not so improve. They also reported that the washout responders differ from those patients who were not eliminated from the study and who subsequently responded to placebo during a 6-week period. The washout responders were “more mildly ill,” “more chronic,” and characterized by “fewer illness precipitants.” It is an interesting curiosity that Rabkin et al. (1987) also discovered that the “proportion of placebo washout responders declined in the winter months (p. 9).

The variability in outcome reported among studies and differences in improvement assessments by different raters indicate powerful bias and attitudinal mediators in how drug outcome will be experienced, interpreted, or assessed. A similar conclusion is suggested by some multicenter studies that use the same drug and the same criteria for patient selection, but then find that patients improve significantly more at certain of the centers than they do at others. For example, in a single study, Greenblatt, Grosser, and Wechsler (1964) compared improvement rates obtained by three antidepressant drugs and a placebo used at three different hospitals. The rank order of treatment effectiveness among the hospitals was found to be approximately the same no matter what the treatment was. One hospital consistently produced the best results, whereas another produced the worst. Results with imipramine showed that marked improvement occurred 67% of the time at the most effective hospital and 31% of the time at the least effective center. A more recent example comes from a study comparing imipramine, alprazolam, and placebo in the treatment of depression at five different settings (Feighner, Aden, Fabre, Rickels, & Smith, 1983). Although the pooled data showed the drugs to be more effective than the placebo, an examination of the results from each of the centers demonstrated considerable variability. For instance, after 6 weeks of treatment, every one of the six outcome measures showed imipramine to be equivalent to placebo in two or more of the five centers. Two of the centers found a difference favoring imipramine on only 1 of 12 comparisons. An equivalent 1 out of 12 comparisons favored placebos. Variability in outcome extends to placebos as well, with some investigators finding relatively low rates of response and others finding improvement in as many as 80% of those treated with placebo (Jenner, 1977).
MORE CURRENT FINDINGS: COMPARISONS WITH PSYCHOTHERAPEUTIC TREATMENTS

Recent research on drug effects is less plagued by some of the major questions that have surrounded earlier work. For the most part the criteria for inclusion into a depressive sample and the measures of outcome are more objective and clearly spelled out. Duration of treatment tends to be longer and drugs are usually prescribed at accepted dosage levels. It is interesting, therefore, to raise the question of whether the drug effects achieved in recent studies are superior, inferior, or equivalent to past findings. Since the efficacy of the major tricyclics has been generally accepted (despite the inconsistencies in the literature), most current research focuses on testing the adequacy of newly introduced drugs. However, one area where findings on the older accepted tricyclics have continued to be collected is in the comparison of drugs to the psychotherapeutic procedures developed specifically to deal with major depressions. Studies of cognitive therapy, interpersonal psychotherapy, behavior therapy, and social skills training have used the widely prescribed antidepressant medications as standards against which psychotherapy efficacy can be tested. We decided to look at such studies to see what improvement rates for drugs look like under more careful methodology of modern studies. Newer studies may differ from many of the older reports in the bias of the investigators. One would expect these authors to be somewhat less invested in finding powerful medication effects. They are more likely to utilize measures, such as the Beck Depression Inventory, that are more precise and objective than the global ratings earlier work relied on. The standards for significant improvement are probably also more stringent in that the newer measures often permit a precise comparison of patients' depression scores at the end of treatment with normative depression scores.

Eight trials comparing a specified type of psychotherapy (for depression) to antidepressant drugs have been reported in seven studies. Three of these trials indicated that psychotherapy was equivalent to medication in fostering improvement (Blackburn et al., 1981, Hospital Outpatient Sample; Murphy, Simons, Wetzel, & Lustman, 1984; Weissman, Prusoff, DiMascio, Neu, Goklaney, & Klerman, 1979) and five indicated that psychotherapy was superior to the drug in promoting substantial change (Bellack, Hersen, & Himmelhoch, 1981; Beutler et al., 1987; Blackburn et al., 1981, General Results for this trial were also reported by Blackburn and Bishop (1981).

Two reports of a follow-up study that included additional patients and results after 6 months of maintenance treatment showed that social skills training continued to be superior to the drug on measures of social skills. There were no differences between social skills training and dynamic psychotherapy on the social skills measures. In the expanded sample both types of psychosocial treatments produced outcomes equivalent (rather than superior) to medication on measures of depression (Bellack, Hersen, & Himmelhoch, 1983; Hersen et al., 1984).
1. EXAMINING ANTIDEPRESSANT EFFECTIVENESS

Practice Sample; McLean & Hakstian, 1979; Rush et al., 1977). None of the trials showed the drug to be superior. One of the studies used a relaxation group as a control (McLean & Hakstian, 1979). This group was presumed to be equivalent to a placebo because there was no reason to believe that relaxation by itself would be an effective treatment for depression. Results showed no difference between treatment with relaxation and treatment with drugs.

Eleven trials presented evidence on whether adding drugs to psychotherapy produced better results. In nine, the addition of drugs made no difference in treatment outcome (Beck, Hollon, Young, Bedrosian, & Budenz, 1985; Bellack et al., 1981; Beutler et al., 1987; Blackburn et al., 1981, General Practice Sample; DeRubeis, 1983; Murphy et al., 1984; Roth, Bielski, Jones, Parker, & Osborn, 1982; Rush & Watkins, 1981; Wilson, 1982). The other two trials showed that the combination of drugs and psychotherapy was better than either alone (Blackburn et al., 1981, Hospital Outpatient Sample; DiMascio et al., 1979; Weissman et al., 1979). Weissman and her collaborators, in finding the combination treatment preferable, indicated that drugs and psychotherapy affected different aspects of the clinical picture. Psychotherapy had its main effect on mood, apathy, suicidal ideation, work, and interest, whereas medication mainly influenced sleep and appetite. Differential treatment response was also noted between those classified as endogenous depression and those diagnosed as situational depression (Prusoff et al., 1980). Endogenous depressed patients responded best to a combination of interpersonal psychotherapy and medication, whereas those with situational depression responded well to interpersonal psychotherapy with no added benefits if drugs were included in the treatment.

Conte and her colleagues (Conte, Plutchik, Wild, & Karasu, 1986) provide an additional perspective in a review of studies published between 1974 and 1984 comparing the outcome of combined treatment (psychotherapy and drug therapy) with either psychotherapy or medication administered alone. They used an elaborate statistical procedure that weighted studies according to their design adequacy. Seventeen reports on 11 patient samples were surveyed. Included in the analysis were some early reports on psychotherapies that were not specifically designed for treating depression in a focused way (e.g., marital psychotherapy, dynamic group psychotherapy, and psychotherapy to prevent relapse in patients who had been responsive to previous drug treatment). The authors of the review concluded that the combined treatment was "slightly" more efficacious than either drugs or psychotherapy when applied alone. However, they also note that the results might be interpreted as showing that most often there is no difference between combined treatments and psychotherapy or drugs administered alone. Indeed, the data showed that there was a four times greater likelihood that the combined condition would equal psychotherapy than be superior
to it and that it was twice as likely that the combined treatment would equal drug treatment alone rather than exceed it.

Some may speculate that the constraints of adhering to an experimental design might render drug therapy less effective in study trials than it would be in "real life" where physicians feel free to flexibly change drugs or dosages. There is at present no empirical evidence for such a speculation. In fact, there are indications in the literature that treatment with medications, as typically used in practice, would not produce significantly better results. Teasdale, Fennell, Hibbert, and Amies (1984) compared treatment outcomes for 34 general practice patients with a major depressive disorder (91% meeting the Research Diagnostic Criteria for definite or probable endogenous major depressive disorder). Patients were randomly assigned to continue the treatment they would normally receive (which typically relied on antidepressant medication) or to receive sessions of cognitive psychotherapy in addition to treatment as usual. Following treatment, patients receiving cognitive therapy, in addition to usual medical treatment, showed a significantly superior outcome on blind and independently assessed measures of symptom severity (i.e., the Hamilton Rating Scale for Depression and the Montgomery Asberg Depression Scale). At termination, patients receiving cognitive therapy also rated themselves as less depressed (on the Beck Depression Scale) than those getting just the usual treatment. Post-treatment, 82% of the psychotherapy patients rated themselves as not depressed (on the Beck Scale) compared to 23% in the treatment as usual condition. Improvement for the usual treatment condition rose to 58% post-treatment if patients rated mildly depressed were included. At 3-month follow-up, results for the usual treatment group improved to a level comparable to the psychotherapy patients.

Another interesting point regarding trials comparing drugs and psychotherapy is made by Hollon and DeRubeis (1981). They note that studies most often use placebo-plus-psychotherapy combinations to represent psychotherapy in comparative trials. According to data they derived from several studies, the placebo-plus-psychotherapy combination is not equivalent to psychotherapy alone and therefore potentially misleading in assessing relative outcomes. In particular, they discovered that psychotherapy alone was more likely to exceed comparative treatments than was psychotherapy-plus-placebo. If proven correct, their arguments indicate that reviews of the literature may underestimate the efficacy of psychotherapy when compared to medication.

Perhaps the findings most relevant to the present chapter concern the percentage of patients who improved substantially as a result of drug treatment. These figures provide an approximation of a "cure rate." Of the 7 samples that provided this type of data, 5 showed that the rate of substantial or marked improvement was between 14% and 27% (Beck et al., 1985; Blackburn & Bishop, 1981, General Practice Sample; Hersen et al., 1984; McLean & Hakstian, 1979; Rush et al., 1977). The remaining 2 samples showed substan-
tial improvement rates of 56% (Murphy et al., 1984) and 77% (Blackburn & Bishop, 1981, Hospital Outpatient Sample). The median percentage of patients who substantially improved on antidepressants in these 7 samples was 25%. This finding suggests that antidepressant medications may be significantly less potent in fully alleviating depressions than some reviews have implied.

A statistical meta-analysis of 56 outcome studies comparing the relative effectiveness of treating unipolar depression in adults with drugs or psychotherapy augments the findings cited earlier on specific types of psychotherapy for depression. The meta-analysis (Steinbrueck, Maxwell, & Howard, 1983) synthesized the results of studies on psychotherapy outcome for depression and studies of drug therapy outcome for depression. The psychotherapy and the drug therapy did not have to occur within the same study to be included, although each treatment had to be compared to a control group. Treatment effectiveness was measured by the computation of effect sizes for each study. It was concluded that, on average, psychotherapy outcome was superior to drug therapy outcome (with an average effect size that was approximately twice as large). This conclusion needs to be viewed with some caution being that the drug studies and the psychotherapy studies were not conducted under exactly comparable conditions. Therefore, differences in outcome might be due to differences in study characteristics rather than differences in the type of treatment. Some of the differences may have favored psychotherapy outcome, whereas other differences favored drug treatment. For example, a bias in favor of psychotherapy outcome may have resulted from drug studies more often employing double-blind procedures. In contrast, drug outcome was favored by the treatment duration differences. The average duration of the drug therapy in the studies reviewed was almost twice as long as the psychotherapy duration (7 vs. 4 weeks). It was not possible in the integration of these diverse studies to statistically ensure that the results were due only to differences in the treatments.

Initial findings of the National Institute of Mental Health Treatment of Depression Collaborative Research Program (Elkin et al., 1986) help to round out the current picture. The study, conducted at three different sites, tested the relative efficacy of two forms of psychotherapy (interpersonal psychotherapy and cognitive behavior therapy) and a standard drug treatment (imipramine) combined with clinical management. A pill-placebo control group with clinical management was also included. Treatments were carefully defined and standardized, and patients were objectively diagnosed using research criteria. First reports on the results indicate no significant differences in outcome between either of the psychotherapies and imipramine in reducing depressive symptoms or affecting overall functioning. By 16 weeks, all treatments (including the placebo condition) resulted in a significant reduction in symptoms of depression. The difference in outcome between active treatments
and placebo was more marked for the severely depressed patients. Improvement for all treatment groups was across a broad range of measures without differential effects in specific areas for different treatments. A more detailed presentation of the results, as well as work on significant predictors of treatment response, is promised for the future.

In summary, a growing number of carefully done trials comparing active, focused psychotherapies (such as cognitive or interpersonal therapy) to antidepressant drug treatment suggests that depressed outpatients receiving psychotherapy do at least as well, and sometimes better, than those receiving drugs. Although drugs may help patients with their sleep disturbances, research shows they are often less efficient than psychotherapy in helping patients with depression and apathy (DiMascio et al., 1979) and frequently ineffective in aiding patients in their social adjustment, interpersonal relationships, or work performance (Lyons, Rosen, & Dysken, 1985; Weissman, Klerman, Paykel, Prusoff, & Hanson, 1974). In contrast, psychotherapy with similar depressed outpatients has led to improvements in overall adjustment, interpersonal communication, and work performance while reducing interpersonal friction and anxious rumination (Weissman et al., 1974).

As our review notes, there are a few indications in the literature that drugs seem to work best with a subgroup of depressed patients who exhibit symptoms that have been associated with an endogenous classification (e.g., anorexia, weight loss, and insomnia). Presently, the studies demonstrating this are relatively few and the definitions utilized by different investigators are not always consistent with each other. Furthermore, as we have previously indicated, some investigators have not found endogenicity to predict either a unique response to drugs or a response that cannot be attained with non-drug treatment. There are also scattered suggestions in the literature that depressed patients with delusions, suicidal ideation, hypochondriacal concerns, or hysterical personality features may have a poor response to tricyclic medications (see review by Bielski & Friedel, 1976). In an update of the Bielski and Friedel review (1976) of predictors of antidepressant response, Friedel (1983) concluded that specific symptoms serve as better predictors of drug response than do diagnostic classifications. Diagnostic categories, even when determined by objective operational criteria such as the Feighner Criteria, the Research Diagnostic Criteria, or the DSM III Criteria for major depressive disorders, seem to produce groupings that are still too heterogeneous to allow for high-level prediction of tricyclic response. The update notes that psychomotor retardation now appears to be the strongest predictor of medication response. It remains for future research, which controls for depression severity and provides clear operational definitions for symptoms, diagnostic conditions, and treatments, to clarify whether the hints in the literature will result in specific, reliable personality predictors for good and poor response to antidepressant drugs.
RELAPSE

The literature suggests that after a period of recovery, depressed individuals frequently experience a return of their previously troubling symptoms. The reality of this observation is carefully documented in a review by Belsher and Costello (1988) who, after highlighting the ambiguities and inconsistencies in how relapse and recovery have been defined, show that approximately 50% of depressed patients relapse within 2 years of recovery. Overall, although the cumulative probability of relapse was proven to increase with time (i.e., 30% at 6 months, 40% at 12 months, and 50% at 2 years), individual patients were less likely to relapse the longer they stayed well. The review noted several factors that increased the likelihood of relapse. Included were the following: recent environmental stress, the lack of social support from family, a history of depressive episodes, and persistent neuroendocrine dysfunction following recovery.

Some researchers have proposed that antidepressant medication may be of particular value in staving off relapse in patients afflicted with recurrent unipolar depression who initially respond positively to the medication. For example, Prien and Kupfer (1986) reviewed six studies where such patients were either continued on antidepressants or switched to a placebo after responding positively to the initial trial on medication. On average, 50% of the patients relapsed when switched to a placebo, whereas only 22% of those continuing on the antidepressant relapsed. Prien and Kupfer went on to suggest that relapse was less likely if responsive patients were continued on the active drug until they had been free of symptoms for 16 to 20 weeks. This research is consistent with some other indications that the nonpsychotic, nonbipolar depressions treated with drugs may not have a particularly positive long-term course once treatment stops. Kovacs, Rush, Beck, and Hollon (1981), in analyzing data from both their study and the work done in Boston and New Haven by Weissman and her co-workers (Weissman, Kasl, & Klerman, 1976), found that approximately two thirds of the patients became symptomatic again at some point during the first year after ceasing to take medication (irrespective of whether they had been on the medication for longer than 1 year or for only 3 months). Although rate of relapse for the 1 year following cognitive therapy seemed lower than the rate for those treated with drugs (44% vs. 65%), the relatively small number of patients involved rendered the differences in clinical course not statistically significant.

Hints are beginning to emerge that the protection against relapse afforded by medication may be at least equaled, and perhaps surpassed, by psychotherapy. Blackburn, Eunson, and Bishop (1986) presented follow-up results for patients with unipolar depression who had initially been assigned randomly for treatment with either cognitive psychotherapy or pharmacotherapy, or a combination of both therapies. Following symptom remission, recovered
patients were continued for 6 months on maintenance trials of their respective treatments. Relapse rates (as defined by symptom scores) were shown to vary with the type of maintenance treatment. Within 6 months, 30% of those on medication maintenance had relapsed, compared to 6% for those receiving booster sessions of psychotherapy, and a zero rate for those maintained on the combination of both treatments. Recurrence rates at 2 years (as defined through the use of hospital records and case notes) were 78% medication maintenance, 23% psychotherapy, and 21% for the combination treatment. These data suggest that periodic psychotherapy sessions were more effective in preventing relapse than were maintenance dosages of the medication. However, these findings are based on a relatively small number of subjects and we do not know from these results if the patients initially treated with medication would have been better maintained with psychotherapy than they were with maintenance medication.

Kupfer and Frank (1987) present some initial data suggesting that relapse rates can be held to less than 10%, 4 months after recovery, if a combined treatment consisting of psychotherapy, an educational workshop, and medication is administered during both the acute and the continuation phases of treatment to patients with recurrent (mainly endogenous) depressions. They specifically attribute the low relapse rate they obtained to the addition of psychotherapy to maintenance medication, because other trials with similar patients receiving drug maintenance alone had been much less effective in preventing relapse. As these authors point out, to more definitively confirm the advantages of combined treatments over drug-only approaches in preventing relapse, the results need to be replicated and compared in a randomized trial to results for patients being continued on drug therapy alone and, we would add, to results for patients being continued on psychotherapy alone.

NEW ANTIDEPRESSANT-PLACEBO STUDIES

A review of the literature on newer antidepressants by Kane and Lieberman (1984) created a unique opportunity for viewing drug effects from another vantage point. The review analyzed 49 efficacy studies of the drugs amoxapine, maprotiline, and trazodone. Of particular interest to us, however, was a subset of 20 studies that looked at one of the newer drugs in comparison to one of the standard tricyclic drugs (i.e., imipramine or amitriptyline) and placebos. All of these studies met the additional criteria (as stipulated by Kane and Lieberman) of appearing in English language publications, including only patients with depression, and having a minimum duration of 3 weeks for drug trial. We decided to focus on the results comparing the effectiveness of placebo to one of the standard drugs in each of the studies where such a comparison was possible. Our aim was to see if the work performed under the more sophisticated conditions of
modern studies would show any change from the older literature in the effectiveness of the standard tricyclic drugs (imipramine or amitriptyline) when compared to placebo. These studies, of course, also offered the advantage of comparing the effectiveness of the standard tricyclics to a placebo in a situation where proving the efficacy of the standard drug was not a prime concern. In general, compared to the older literature the new studies used more careful, clear depression criteria for including patients and utilized currently accepted standards for drug dosage levels and treatment duration.

In order to develop an overall impression of outcome from the available studies, we followed Kane and Lieberman’s decision to look at response rates in each of the studies. We examined each study to see if there was a significant difference in response rates between drug and placebo. Like Kane and Lieberman we were interested in “clinically meaningful” differences and we used the measures that they selected for their review. Therefore, the data from each study were included if any of the following three measures was available, listed in order of decreasing priority: the percentage of patients achieving (a) a specified final score on the Hamilton Depression Scale (HAM-D) that could be considered indicative of complete or near complete recovery; (b) a clinical global improvement rating of moderate or marked; or (c) a 50% reduction in scores from baseline to end point on the HAM-D or a comparable scale.

Of the 20 studies, 16 provided data on at least one of the three specified measures. Table 1.1 presents a list of the 16 studies along with information for each study on patient status, type of drug used, drug dosage, duration of drug trial, and significance of the drug-placebo difference. For purposes of categorization, studies that allowed maximum imipramine doses of 200 to 300 mg per day were considered high dosage (11 of 13 studies) and those that permitted amitriptyline doses up to 200 mg per day were classified as high dosage (3 of 3 studies). As can be seen, all but 2 of the studies permitted doses up to the commonly accepted maximum levels. In 2 of the studies, statistical tests for significance were not reported and had to be computed from the data presented. Two other studies did not report the percentage of improvement, but did indicate if there was or was not a statistically significant drug-placebo difference in the percentage of patients improving on one of the specified measures.

An overview of all 16 studies indicates that the majority (62%) show no difference in the percentage of patients benefiting from an active drug as opposed to a placebo. Additional analyses were conducted to see if drug-placebo outcome differences were affected by a number of patient or study variables: in- or outpatient status, duration of drug trial, dosage level, sample size. It might be assumed that hospitalized patients have a more severe form of depression and that this might affect results. Therefore, we separated the studies into those dealing with inpatients and those performed on outpatients. Both the majority of studies with inpatients (57%) and the majority of studies with outpatients (67%) showed no difference in the percentage of patients
TABLE 1.1
Summary of Newer Drug Trials Comparing Standard Tricyclics to Placebo

<table>
<thead>
<tr>
<th>Study</th>
<th>Patient Status</th>
<th>Drug Status</th>
<th>Drug Dosage</th>
<th>Duration of Trial (in weeks)</th>
<th>Significance of Drug-Placebo Difference (Sig. or N.S.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominquez et al., 1981</td>
<td>O</td>
<td>Im</td>
<td>High</td>
<td>6</td>
<td>N.S.</td>
</tr>
<tr>
<td>Escobar et al., 1980</td>
<td>I</td>
<td>Im</td>
<td>High</td>
<td>4</td>
<td>Sig.</td>
</tr>
<tr>
<td>Fabre et al., 1979</td>
<td>I</td>
<td>Im</td>
<td>High</td>
<td>4</td>
<td>Sig.</td>
</tr>
<tr>
<td>Feighner, 1980</td>
<td>I</td>
<td>Im</td>
<td>High</td>
<td>4</td>
<td>N.S.</td>
</tr>
<tr>
<td>Gershon et al., 1981</td>
<td>O</td>
<td>Im</td>
<td>High</td>
<td>4</td>
<td>N.S.</td>
</tr>
<tr>
<td>Goldberg &amp; Finnerty, 1980</td>
<td>O</td>
<td>A</td>
<td>High</td>
<td>6</td>
<td>N.S.</td>
</tr>
<tr>
<td>Goldberg et al., 1981</td>
<td>O</td>
<td>A</td>
<td>High</td>
<td>6</td>
<td>Sig.</td>
</tr>
<tr>
<td>Keiv &amp; Okerson, 1979</td>
<td>O</td>
<td>Im</td>
<td>Low</td>
<td>6</td>
<td>N.S.</td>
</tr>
<tr>
<td>Kellams et al., 1979</td>
<td>I</td>
<td>Im</td>
<td>High</td>
<td>4</td>
<td>N.S.</td>
</tr>
<tr>
<td>Mann et al., 1981</td>
<td>O</td>
<td>Im</td>
<td>High</td>
<td>4</td>
<td>N.S.</td>
</tr>
<tr>
<td>Rickels &amp; Case, 1982</td>
<td>O</td>
<td>A</td>
<td>High</td>
<td>6</td>
<td>Sig.</td>
</tr>
<tr>
<td>Rickels et al., 1981</td>
<td>O</td>
<td>Im</td>
<td>High</td>
<td>6</td>
<td>N.S.</td>
</tr>
<tr>
<td>R.C. Smith, 1975</td>
<td>O</td>
<td>Im</td>
<td>Low</td>
<td>6</td>
<td>N.S.</td>
</tr>
<tr>
<td>Steinbook et al., 1979</td>
<td>I</td>
<td>Im</td>
<td>High</td>
<td>4</td>
<td>N.S.</td>
</tr>
<tr>
<td>Trapp et al., 1979</td>
<td>I</td>
<td>Im</td>
<td>High</td>
<td>4</td>
<td>N.S.</td>
</tr>
<tr>
<td>Van Der Velde, 1981</td>
<td>O</td>
<td>Im</td>
<td>High</td>
<td>4</td>
<td>Sig.</td>
</tr>
</tbody>
</table>

*Significance calculated from data presented

significantly improving on drugs as opposed to placebos. To see if longer lasting clinical trials produced better results than shorter trials, the studies were divided according to treatment duration. Nine of the studies had a treatment duration of 4 weeks and 7 of the studies had a 6-week treatment duration. Treatment duration made no difference in outcome. The majority of both shorter duration studies (56%) and longer duration studies (71%) showed no significant difference in the percentage of patients who responded to drugs or placebos. As previously noted, only 2 of the studies used a "low
dose" of the drug and those studies were segregated from the higher dose studies to see if dosage made a difference. Neither of the low-dose studies revealed a drug-placebo difference. The majority of high-dose studies (57%) also showed no difference. Because a small sample size might diminish the probability of finding a statistically significant difference between drug and placebo treatments, the studies were divided into those that included sample sizes of 15 or more patients in each treatment group and those that did not. Seventy-five percent of the studies with the smaller sample sizes showed no significant drug-placebo difference and 50% of the large sample size studies showed no difference. The difference in outcome due to sample size was not significant. In sum, a series of chi-square analyses examining the number of studies showing significant drug-placebo differences and the number not showing differences indicated that inpatient status, treatment duration, dosage level, and sample size all played no role in determining significance. Only 4 of the 16 studies (all using outpatients) were conducted with a combination of the longer trial duration, higher dose levels and larger sample sizes. Half of these 4 studies showed significant results and half did not. In order to get a sense of overall improvement in these studies, reports of the percentage of patients who improved in each study were combined. On average, 59% of the patients on tricyclics were rated as improved, whereas 36% of the placebo-treated patients improved. The percentage of improvement on either drugs or placebos tended to be lower for inpatient (drugs 49%; placebos 21%) than it was for outpatients (drugs 67%; placebos 47%).

The review, therefore, does not support the idea that the effectiveness of the standard antidepressant drugs is more clearly demonstrated by newer drug trial investigations. Drug-placebo differences in outcome tended to be modest (with a median difference of 21%) and the majority of studies showed no difference in the percentage of patients significantly improved by drugs in contrast to placebos. There is no indication that current studies using more objective inclusion criteria, longer treatment durations, and acceptable medication dosages are achieving better outcomes than older studies did, and the overall drug-placebo outcome difference is frequently smaller than the 30% to 35% figure commonly reported in the older literature. The instability of response to drugs or placebos was readily apparent across the new drug trial studies. For drug treatment the percentage of patients improving ranged from 20% to 80%; whereas for placebo treatment the percentage of patients improving ranged from 0% to 91%. It could, of course, be argued that trials of longer than 6 weeks would show more significant drug effects; however, there are indications that many improved patients show a worsening during weeks 7 to 12 that necessitates an increase in dosage or a change of drug (Prien & Levine, 1984). It might also be argued that studies allowing adequate maximum drug dosages do not actually utilize the maximum dosage and therefore obscure drug effects. For the present, however, we must conclude
that current research shows only modest drug effects. It should be added that
the earlier review of the recent literature dealing with psychotherapeutic
treatment of depression also underlines the relative modesty of antidepressant
drug effects from still another perspective.

DOUBLE-BLIND DESIGN

It goes without saying that unless a drug treatment produces a therapy result
substantially exceeding that obtained from placebo and/or spontaneous recov­
ery it has little value. This is especially true if one considers the not inconsider­
able and sometimes life-threatening negative side effects that some drugs
initiate. This point is widely accepted (Lasagna, 1979) and underlies the fact
that the double-blind placebo design is considered essential for establishing
therapeutic efficacy. There are acknowledged difficulties in applying the
double-blind design. Foremost among these is the fact that the participants
pick up cues that sometimes make it possible to differentiate between the
patients receiving the active drug and those receiving the placebo (Jenner,
1977; Marini, Sheard, Bridges, & Wagner, 1976; Nash, 1962, Rabkin et al.,
1986; Stallone, Mendlewicz, & Fieve, 1975). This is a serious problem because
previous studies have shown that the less controlled the evaluation of a
therapeutic procedure the more an experimenter can bias the outcome. The
previously cited review by Wechsler et al. (1965) of the antidepressant litera­
ture nicely illustrates this point. They looked at more than 100 studies and
reported that whereas 17 of 30 no-control studies reported at least 65%
 improvement, only 5 of 22 placebo-controlled studies showed that much
improvement. Interestingly, too, Wechsler et al. discovered a significant
positive correlation between degree of therapeutic efficacy of the active drug
in each study and the efficacy of the placebo. If the drug effect was large, the
placebo effect was also large. This suggests that despite the use of the double
blind there was a spread of intensity of therapeutic expectation for the active
drug to the placebo. A similar pattern, indicating that a placebo often has
one half the efficacy of the active drug with which it is being compared, has
also been described (Jospe, 1978).

By and large, the early, largely uncontrolled (non-double-blind) studies
initiated by enthusiasts for a drug treatment are those that come up with the
most dramatic therapeutic results. Karlowski et al. (1975) have actually shown
that the breaking of the double blind by patients influences the patients’
ratings of their symptoms. It should be acknowledged that some studies have
not shown such an effect. There are many sources of uncontrolled information
that can undermine a double-blind design. These have been reviewed else­
where. They may variously involve differences between the active drug and
the placebo with reference to the quantity and specific sites of their side
effects, and also with reference to the time frames in which they generate responses (Guy, Gross, & Dennis, 1967; Jospe, 1978).

A serious source of interference with the double-blind are the cues supplied by the body sensations aroused by an agent. Hill, Haertzen, Wolbach, and Miner (1963) discovered that patients learn to discriminate between drug and placebo largely from cues provided by body sensations and symptoms. Examination of studies of the effectiveness of the antidepressants indicates that such agents produce different patterns of body sensations than do inert placebos. A substance like imipramine usually initiates clearly defined body experiences (e.g., dry mouth, tremor, sweating, constipation). Inactive placebos used in studies of antidepressants also apparently initiate some body sensations, but they are fewer, more inconsistent, and less intense as indicated by the fact that they are less often cited by patients as a source of discomfort causing them to drop out of treatment (Klein et al., 1980). Probably in the great majority of studies of the effectiveness of antidepressants involving a comparison with an inactive placebo there have been significant differences in the body experiences of the drug and placebo groups. Such differences could signal to the patients involved whether they were receiving an active or inactive agent and they could, further, supply discriminating cues to all personnel (e.g., nurses) responsible for the patients' day-to-day treatment. In the case of the personnel, one would expect that they would adopt different attitudes toward those they identified as being "on" versus "off" active treatment and consequently communicate contrasting expectations. Porter (1970) and also Rabkin et al. (1986) actually reported that in a double-blind study of imipramine it was possible by means of side effects to identify a significant number of the patients taking the active drug. Those patients receiving an inactive placebo have fewer signals (from self and from others) indicating they are being actively treated and should be improving. By the same token, patients taking an active drug like imipramine receive multiple signals that may well amplify potential placebo effects linked to the therapeutic context. Is it possible that a large proportion of the difference in effectiveness often reported in comparisons of antidepressants with inactive placebos can be explained as a function of body sensation discrepancies? It is conceivable, and fortunately there are research data that shed light on the matter.

ACTIVE PLACEBO STUDIES

Let us begin with an analysis by Thomson (1982). He reviewed all the double-blind placebo-controlled studies of tricyclic antidepressants completed between 1958 and 1972 that he could find. He discovered that 68 of the studies had employed an inert placebo and 7 an active one (atropine) that produced a variety of body sensations. When the outcomes of the studies were
computed, he found that whereas 59% of the designs in which an inert placebo was employed indicated that the tricyclic had a superior therapeutic effect, this was true in only one study (about 14% of the designs) in which the active placebo was utilized. The difference was statistically significant. Using an active placebo in the experimental designs eliminated any therapeutic advantage for the tricyclics. To check on the reliability of his own judgments, Thomson had a second rater evaluate the therapeutic outcomes that were described in the active placebo studies and this rater's judgments were in perfect agreement. Thomson raised the question whether the atropine employed as the active placebo might, because of its anticholinergic effects, have had an antidepressant effect. He wondered, too, whether researchers careful enough to incorporate an atropine placebo might simply have been more rigorous in their experimental designs. In any case, his findings did forcefully document the possibility that the active placebos produced greater therapeutic "amplification" than did the inert ones. It is important to emphasize that the failure to find drug-placebo differences in the studies cited by Thomson was not due to low rates of improvement in patients receiving antidepressants, but rather results largely from the elevated improvement rates in patients receiving the active placebos. Other findings in the literature suggest that an active substance increases placebo potency (Brune et al., 1962; Kast & Loesch, 1961). There are also data suggesting that the side effects associated with active drugs may enhance the drugs' therapeutic efficacy (Dinnerstein & Halm, 1970; Penick & Fisher, 1965).

Searching the literature concerned with the antidepressants, we found four other instances in which active placebos (or reasonable equivalents) were utilized, but which were not available to the Thomson analysis. In these studies differences between the antidepressant and the active placebo were of a low order. We briefly consider each here. Fahy, Imlah, and Harrington (1963) compared the efficacy of imipramine, electroconvulsive therapy, and thiopentone sleep treatments in treating patients with "moderately severe" depressive symptoms. Thiopentone sleep was considered to be a placebo. No significant differences among the three treatment conditions occurred. McLean and Hakstian (1979) tested the relative therapeutic power of amitriptyline, short-term psychotherapy, behavior therapy, and a control condition somewhat analogous to an active placebo. The control condition involved teaching patients how to relax their muscles and getting them to "appreciate the relation between muscle tension and depression" (p. 821). It was in the main focused on changing body experience. The patients who participated were "moderately clinically depressed." No differences were detected between the drug and control groups for any of several outcome measures. Weintraub and Aronson (1963/1964) examined the effects of imipramine and an active placebo (atropine). The majority of patients were classified as "moderately" to "severely" depressed. Although a group of resident physicians rated their
1. EXAMINING ANTIDEPRESSANT EFFECTIVENESS

patients as significantly more improved on imipramine than on placebo, the chief residents who performed the same set of ratings could not discern a significant difference. An MMPI measure of improvement that was included showed a modest significant advantage for imipramine. Friedman (1975) looked at the relative effectiveness of amitriptyline and an active placebo (atropine) in treating depression. This was done in the context of a "marital therapy" condition and a "minimal personal contact" condition. The actual design involved drug-marital therapy, drug-minimal contact, placebo-marital therapy, and placebo-minimal contact. All patients had a primary diagnosis of depression. At the end of 10 weeks of treatment there were only minor differences between the drug and placebo categories, as defined by rating scales. The effects of the drugs and placebos were more alike than different. Almost all of the ratings indicated an absence of significant differences. During a 2-week period subsequent to treatment when no drug or placebo was administered, differences in relapse rates could not be detected. Also, self-ratings by patients gave largely negative results with respect to drug versus placebo differences. The findings just reviewed are of the same tenor as Thomson reported in his analysis of the comparative efficacies of tricyclics and active placebos.

The fact that increasing the activity of a placebo intensifies its efficacy has been documented in other contexts too. For example, Kast and Loesch (1961) treated patients with "functional digestive disorders" with meprobamate and tridihexylylloide. They showed that adding atropine to the treatment and focusing the patients' attention on the "dry mouth" sensations associated with it changed the "improvement rate" significantly up or down, as a function of psychological sets they suggested. The addition of the "dry mouth" experience could be manipulated to alter improvement rates. Interestingly, Lipman, Park, Rickels, and Chase (1966) likewise found a variable impact on improvement rates of anxious patients treated with either chlordiazepoxide hydrochloride or chlordiazepoxide plus atropine, or placebo (inactive), or atropine. More specifically, focusing the patient's attention on "dry mouth" (produced by atropine) in the context of a physician describing "dry mouth" as a positive therapeutic indicator had a relatively more negative effect on improvement than did such focusing when the physician adopted a neutral attitude toward "dry mouth." Overall, the active placebo (atropine) was found under a "neutral set" condition to produce a poorer response than inactive placebo at one of the clinics participating in the study and a "marginally better" response at the two other participating clinics. Baker and Thorpe (1957) reported that a placebo had greater therapeutic efficacy than mepazine because the particularly sweet taste of the placebo was so sensorially impressive to the subjects. Fangman (1963) contrasted morphine with an inactive placebo and with an active placebo (phenobarbital) in the treatment of pain. Phenobarbital is not usually regarded as an analgesic but it does from the perspective of naive subjects
produce side effects similar to those for morphine. Although morphine proved to be significantly more analgesic than the inactive placebo, it was only marginally, or not at all, more analgesic than the active placebo. Rickels, Lipman, and Raab (1966) indirectly highlighted the fact that the efficacy of a placebo is influenced by how obviously it can be identified as non-active. They carried out crossover studies in which patients received a placebo after first being on either “therapeutically active” or “inactive” drugs. There was a 66% improvement rate for the placebo in those who had first received the “inactive” drug; whereas the improvement rate was only 37% in those who had initially received the “active” medication. The difference was statistically significant. It was concluded: “In other words, when offering a patient the possibility for a direct comparison between drugs, he frequently differentiates between active agent and placebo, and the more effective clinically the first agent, the worse is the subsequent placebo response” (p. 549). Apropos of this observation, one must take a conservative and even somewhat skeptical attitude toward reports in the literature (Prien & Kupfer, 1986) describing substantial differences in the relapse rates of depressed patients who have significantly improved as the result of antidepressant treatment and who are either allowed to continue on their antidepressant medication or placed on placebo. Patients who remain on their medication are roughly 35% less likely to relapse than are those shifted to placebo. Obviously, those given the placebo would receive numerous bodily cues that they were now ingesting an inert substance and also the possibility would be afforded of an unusually vivid direct comparison with the active agent.

Overall, one may conclude that the active placebo is more powerful than the inactive placebo and this probably reflects the fact that the active placebo more convincingly arouses body sensations that affirm that a potent agent has been taken into one’s body. Interestingly, the efficacy of a placebo may in some contexts be significantly correlated with the number of side effects it produces (Moertel, Taylor, Roth, & Tyce, 1976; Shapiro, Struening, Barten, & Shapiro, 1975). In the Shapiro et al. study it was demonstrated that both positive and negative effects of a placebo may be enhanced by a large number of “side effects.” Joyce (1959) found evidence that placebo responders may be particularly sensitive to the autonomic changes occurring in their body, and one may surmise that such responders unrealistically magnify the side effects aroused by a placebo. Evidence exists that many of the effects of active drugs may at times be linked simply to the changes in body experience they induce. Karniol, Dalton, and Lader (1978) reported that the ratings by normal subjects of their “well-being” after ingesting lithium chloride did not correlate with plasma or erythrocyte lithium concentrations, but rather with amount of nausea experienced. The peripheral effects of lithium (viz., nausea) and not blood concentration levels were best predictive of subjective drug effects reported. After reviewing the literature on response variability to psychotropic