

Are Subjects in Pharmacological Treatment Trials of Depression Representative of Patients in Routine Clinical Practice?

Mark Zimmerman, M.D.

Jill I. Mattia, Ph.D.

Michael A. Posternak, M.D.

Objective: The methods used to evaluate the efficacy of antidepressants differ from treatment for depression in routine clinical practice. The rigorous inclusion/exclusion criteria used to select subjects for participation in efficacy studies potentially limit the generalizability of these trials' results. It is unknown how much impact these criteria have on the representativeness of subjects in efficacy trials. This study estimated the proportion of depressed patients treated in routine clinical practice who would meet standard inclusion/exclusion criteria for an efficacy trial.

Method: A total of 803 individuals, aged 16–65 years, who were seen at intake at an outpatient practice underwent a thorough diagnostic evaluation, including the administration of semistructured diagnostic interviews; 346 patients had current major depression. Common inclusion/exclusion criteria used in efficacy studies of antidepressants were applied

to the depressed patients to determine how many would have qualified for an efficacy trial.

Results: Approximately one-sixth of the 346 depressed patients would have been excluded from an efficacy trial because they had a bipolar or psychotic subtype of depression. The presence of a comorbid anxiety or substance use disorder, insufficient severity of depressive symptoms, or current suicidal ideation would have excluded 86.0% (N=252) of the remaining 293 outpatients with nonpsychotic unipolar major depressive disorder from an antidepressant efficacy trial.

Conclusions: Subjects treated in antidepressant trials represent a minority of patients treated for major depression in routine clinical practice. These results show that antidepressant efficacy trials tend to evaluate a subset of depressed individuals with a specific clinical profile.

(*Am J Psychiatry* 2002; 159:469–473)

Depression is the most frequently treated specific psychiatric disorder (1), and antidepressants are among the most frequently prescribed medications in all of medical practice (2). Since their introduction nearly 50 years ago, over two dozen antidepressant medications have been approved for use in the United States. In order to obtain that approval, the Food and Drug Administration's Code of Federal Regulations indicates that well-controlled clinical investigations must be performed to distinguish the effect of a drug from other sources of improvement, such as spontaneous remission of symptoms or the placebo effect. Because the principal aim of most antidepressant efficacy trials is to demonstrate a difference between drug and placebo effects, individuals with purportedly high rates of placebo response, e.g., subjects with mild depression, are generally excluded from participation. Similarly, individuals who are believed to have low rates of response to drugs, e.g., subjects with chronic depression or psychiatric comorbidity, are often excluded as well. With regard to patient selection, the Code of Federal Regulations indicates only that the subjects selected must have the medical condition being studied; there is no recommendation for exclusionary factors (21 C.F.R. 314.126).

One limitation, then, of antidepressant efficacy trials is that only a certain subset of depressed individuals is ever studied. For years investigators have raised concerns regarding the generalizability of such studies to patients treated in the real world. One line of study that has attempted to address this concern has been in comparing the demographic, clinical, and treatment response characteristics of trial participants recruited by means of advertisement ("symptomatic volunteers") with those of trial participants referred from clinical settings. These studies have generally found that the two groups are similar, and this fact has been interpreted as supporting the generalizability of efficacy trials (3–7). However, such a comparison addresses generalizability only as a function of referral source; it remains unknown how representative trial participants are of the entire population of depressed patients as a function of eligibility requirements.

To address this question, as part of the Rhode Island Methods to Improve Diagnostic Assessment and Services project, we systematically evaluated the clinical profile of a large group of depressed individuals who came to our outpatient psychiatric practice for treatment. On the basis of each patient's clinical profile, we sought to determine

how many depressed patients from our practice would qualify for participation in a standard antidepressant efficacy study. If the percentage were low, this would highlight an important limitation of the antidepressant literature and would suggest that more emphasis may need to be placed on evaluating the efficacy of antidepressant medications in less "pure" groups of depressed patients.

Method

Patients

A total of 803 patients, aged 18–65 years, were evaluated in the outpatient practice of Rhode Island Hospital's Department of Psychiatry. This private practice treats predominantly individuals with medical insurance (Medicare but not Medicaid) on a fee-for-service basis, and it is distinct from the hospital's outpatient residency training clinic, which services predominantly lower-income, uninsured, and medical assistance patients. A total of 346 patients had a principal diagnosis of unipolar or bipolar depression. The group included 122 men (35.3%) and 224 women (64.7%), who had a mean age of 38.5 years (SD=11.3). At the time of the evaluation, nearly one-half of the subjects were married (43.4%); the remainder were never married (26.9%), divorced (15.9%), separated (9.2%), widowed (1.4%), or living with someone as if in a marital relationship (3.2%). A total of 11.3% of the group did not graduate from high school, 65.6% graduated from high school or received an equivalency degree, and 23.1% graduated from college. The racial composition of the group was 85.3% white, 4.6% black, 2.3% Hispanic, 0.9% Asian, and 7.0% another group or a combination of racial backgrounds.

Assessment

Upon presentation for outpatient treatment, all of the patients were interviewed by a trained diagnostic rater who administered the Structured Clinical Interview for DSM-IV (SCID) (8), supplemented with questions from the Schedule for Affective Disorders and Schizophrenia (SADS) (9), to assess the severity of symptoms during the week before the evaluation. An extracted score on the Hamilton Depression Rating Scale was derived from SADS ratings by following the algorithm developed by Endicott and colleagues (10). Personality disorders were not assessed in the first 91 patients evaluated. The next 501 patients were interviewed with the borderline personality disorder section of the Structured Interview for DSM-IV Personality Disorders (11). The last 203 patients were interviewed with the entire Structured Interview for DSM-IV Personality Disorders. (Data were missing for the remaining eight patients.) On the day on which the SCID and the Structured Interview for DSM-IV Personality Disorders were administered, the results of these diagnostic interviews were presented to the treating psychiatrist, who also interviewed the patient and reviewed the findings of the evaluation. Rhode Island Hospital's institutional review board approved the research protocol, and all patients provided written informed consent. The training of the raters and the reliability of the diagnostic assessments in the Methods to Improve Diagnostic Assessment and Services project have been described in detail elsewhere (12, 13).

Inclusion/Exclusion Criteria

We reviewed the inclusion/exclusion criteria used in treatment efficacy studies of depression published from 1994 through 1998 in five journals (the *Archives of General Psychiatry*, the *American Journal of Psychiatry*, the *Journal of Clinical Psychiatry*, the *Journal of Clinical Psychopharmacology*, and the *Psychopharmacology Bulletin*). We identified 31 studies of outpatients whose subject groups were not limited to particular demographic groups,

such as the elderly (references available upon request from Dr. Zimmerman). All, or almost all, of the study groups excluded depressed patients who had psychotic features, a history of a manic or hypomanic episode, suicide risk, unstable medical illnesses, and scores below a certain threshold on the Hamilton depression scale; most studies excluded subjects who abused drugs or alcohol. In addition, several studies excluded subjects with comorbid conditions, including dysthymic disorder, eating disorders, obsessive-compulsive disorder (OCD), panic disorder, any comorbid anxiety disorder, and borderline personality disorder.

Four studies excluded subjects whose depressive episode was over 12, 18, or 24 months in duration, whereas 11 studies required that the depressive episode be of at least 1 month's duration (in contrast to the DSM-IV definition, which requires a symptom duration of 2 weeks).

Our goal was to estimate the percentage of depressed patients seen in routine clinical practice who would meet exclusion criteria for a clinical efficacy trial. However, no standard set of exclusion criteria has been developed to select depressed subjects for an efficacy study. Therefore, we evaluated how frequently depressed patients met each of the following exclusion criteria:

1. History of DSM-IV manic or hypomanic episodes.
2. Experiencing psychotic features during the current episode of depression.
3. Significant risk of suicide, as determined by the SADS suicide item.
4. DSM-IV diagnosis of alcohol or drug abuse or dependence within the last 6 months.
5. Mild depression, as determined by low baseline extracted score on the Hamilton depression scale.
6. Presence of underlying dysthymic disorder.
7. Illness duration of less than 4 weeks or more than 2 years.
8. Presence of comorbid nondepressive, nonsubstance use, axis I disorders.
9. Presence of borderline personality disorder.

Features 3–9 were then examined in the subset of patients with unipolar nonpsychotic major depressive disorder. After examining each item individually in the patients, we examined the impact of using multiple exclusion criteria. We applied the five criteria that were used in at least three-quarters of the efficacy studies of antidepressants we reviewed: history of manic or hypomanic episodes, current psychotic features, low score on the Hamilton depression scale, drug or alcohol abuse or dependence during the 6 months before the evaluation, and significant suicidal ideation. After application of these criteria, we also examined how many more patients would be excluded if the exclusion criteria included the presence of a comorbid anxiety disorder or any comorbid axis I disorder.

Results

Thirty-one (9.0%) of the 346 depressed patients had bipolar I or II disorder. Of the remaining 315 unipolar depressed patients, 22 (7.0%) had psychotic features. Thus, approximately one-sixth (N=53) of the 346 depressed patients would have been excluded from an efficacy trial because they had a bipolar or psychotic subtype of depression.

We examined the impact of the other exclusion criteria on the remaining 293 outpatients with nonpsychotic unipolar major depressive disorder (some patients met more than one criterion). Their mean extracted Hamilton depression scale score was 19.2 (SD=6.1). A total of 54.3% (N=159) of the patients scored below 20, which is the most commonly used severity threshold for inclusion in an efficacy trial. The sec-

ond most commonly used cutoff score for study participation is a score of 18; 42.0% (N=123) of the unipolar depressed patients scored below this value. A total of 7.8% (N=23) of the 293 outpatients had evidence of a drug or alcohol use disorder during the 6 months before the evaluation.

The presence of suicidal ideation is a frequent reason for exclusion from efficacy trials; however, the intensity and frequency of the suicidal thoughts warranting exclusion have not been clearly defined. Table 1 lists the guidelines for rating the SADS suicidal ideation item and the distribution of patients rated at each level. These ratings refer to suicidality during the week before the evaluation. Twenty-four patients (8.2%) were rated at 4 or higher, indicating the presence of frequent suicidal thoughts and planning. If the decision to exclude patients on the basis of the presence of suicidal ideation were equivalent to a SADS rating of 3 (i.e., suicidal ideation of moderate severity), then nearly 20% of the patients would have been excluded from participation.

The most common comorbid psychiatric condition that excluded patients from an efficacy trial was a diagnosis of anxiety disorder (53.2%, N=156). The specific rates of anxiety disorders were 2.4% (N=7) for panic disorder without agoraphobia, 14.3% (N=42) for panic disorder with agoraphobia, 1.0% (N=3) for agoraphobia without panic, 8.2% (N=24) for OCD, 16.4% (N=48) for generalized anxiety disorder, 30.0% (N=88) for social phobia, and 12.6% (N=37) for posttraumatic stress disorder. A total of 29.0% (N=85) of the depressed patients had a single anxiety disorder, 18.8% (N=55) had two anxiety disorders, and 5.5% (N=16) had three or more anxiety disorders.

The mean duration of the index depressive episode was 196.6 weeks (SD=466.1). Episode duration was 2–3 weeks for 16 (5.5%) of the depressed patients. Although these patients met the DSM-IV criteria for major depressive disorder, they would be excluded from studies requiring a minimum duration of 1 month. Chronic major depressive disorder of more than 24 months' duration was present in 101 (34.5%) of the patients.

Slightly more than 10% (N=35) of the patients had borderline personality disorder, and somewhat less than 10% (N=26) had an underlying dysthymic disorder. A total of 15.0% (N=44) of the patients had a nonsubstance-use, nonanxious axis I disorder. Specific diagnoses were as follows: intermittent explosive disorder (3.1%, N=9), body dysmorphic disorder (2.4%, N=7), attention deficit disorder (2.4%, N=7), undifferentiated somatoform disorder (1.7%, N=5), bulimia nervosa (1.4%, N=4), pathological gambling (1.0%, N=3), pain disorder (1.0%, N=3), hypochondriasis (1.0%, N=3), somatization disorder (0.3%, N=1), trichotillomania (0.3%, N=1), and conversion disorder (0.3%, N=1).

Figure 1 illustrates the impact of the exclusion criteria when applied in sequential order. (Figure 1 does not identify patients who met multiple exclusion criteria.) Exclusion of patients with any of the five features usually used as criteria in efficacy trials (bipolar subtype, psychotic

TABLE 1. Severity of Suicidal Thoughts and Behavior Among 293 Outpatients With Nonpsychotic Unipolar Major Depressive Disorder, According to SADS Suicidal Ideation Item

Severity in Past Week	N	%
0. Not at all	99	33.8
1. Slight, e.g., occasional thoughts of death (without suicidal thoughts); "I would be better off dead" or "I wish I were dead"	55	18.8
2. Mild, e.g., frequent thoughts that he/she would be better off dead or occasional thoughts of suicide but no thoughts about a specific method	81	27.6
3. Moderate, e.g., often thinks of suicide or has thoughts about a specific method	34	11.6
4. Severe, e.g., often thinks of suicide and has thought of, or mentally rehearses, a specific plan or has made a suicide gesture of a communicative rather than potentially medically harmful type	16	5.5
5. Extreme, e.g., has made preparations for a potentially serious suicide attempt	5	1.7
6. Very extreme, e.g., suicide with definite attempt to die or potentially cause medical harm	3	1.0

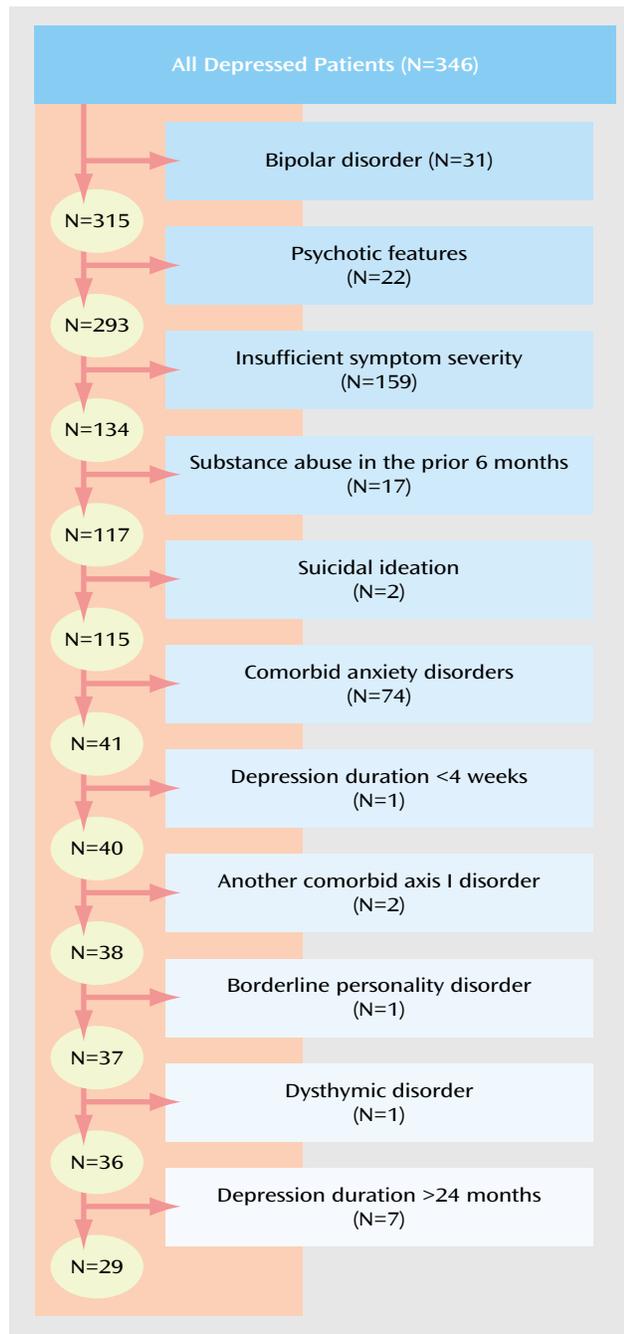
subtype, Hamilton depression scale score of <20, substance abuse or dependence in past 6 months, suicidal ideation) eliminated two-thirds of the patients in our study. If patients with a comorbid anxiety disorder were also excluded, then 86.0% (N=252) of the patients would have been excluded from an efficacy study. After accounting for the effect of these six exclusion criteria, the impact of the remaining criteria (e.g., borderline personality disorder, dysthymic disorder, episode duration greater than 24 months) excluded only a handful of additional patients.

Discussion

Concerns regarding the generalizability of antidepressant efficacy studies have been raised for years (14, 15). The present study sought to determine to what extent participants in antidepressant efficacy trials are representative of patients treated in the "real world." We found that only a small minority of the depressed patients who came into our outpatient psychiatric practice for treatment would qualify for an antidepressant efficacy trial. Depending on the criteria used, as few as 14% of the depressed patients from our practice would have been eligible to participate in such a trial because they met various exclusion criteria. This number could actually be an overestimate, since it does not account for the individuals who would be excluded due to medical comorbidity, prior non-response to somatic therapy, or illicit drug abuse as detected by urine screenings. On the other hand, if the diagnostic evaluations conducted in clinical trials had been performed in a less rigorous manner than what occurs during a semistructured SCID interview, then a broader range of patients might have been included in these trials than the number suggested by the exclusion criteria.

These results point out that antidepressant efficacy trials tend to evaluate only a small subset of depressed individuals with a specific clinical profile. It would be erroneous to assume that because antidepressant medications

FIGURE 1. Impact of Sequential Application of Exclusion Criteria on the Number of 346 Depressed Outpatients Who Would Have Qualified for Participation in an Antidepressant Efficacy Trial



are efficacious in these individuals, they must also be efficacious in all patients diagnosed with major depression. Strictly speaking, the results of standard antidepressant efficacy trials are directly generalizable to only a small fraction of the depressed patients who come in for treatment and are given antidepressant medication.

Should this conclusion affect prescribing habits in any way? Each practicing clinician must decide how to use the results of medical research, given the limitations that in-

variably will be present. As it stands now, there are probably several instances in which clinicians do not prescribe antidepressant medications, even though the full criteria for major depressive disorder are met. For example, when treating patients with very mild symptoms or those experiencing grief, clinicians may recommend not taking somatic therapy. Similarly, antidepressant medications are often avoided for depressed patients who are actively abusing illicit substances. Presumably, somatic therapy is not initiated in these instances because it is not known to offer a better prognosis than the passage of time. Withholding medication is justified on the grounds that it would expose patients to risks and side effects without sufficient evidence to suggest that it might be helpful. Even if antidepressant therapy were initiated, it would probably occur in the context of an informed consent process, in which the lack of sufficient evidence to support such a decision would be openly discussed.

However, it is our sense that the situation is quite different when clinicians treat patients whose clinical profiles would be grounds for exclusion due to a purportedly poor prognosis. For example, there are few data suggesting that antidepressant medications are efficacious for depressed patients with comorbid borderline personality disorder (16). Nevertheless, although the data supporting drug-placebo differences are no more apparent for these patients than for patients with mild depression, antidepressant medications are rarely withheld from such patients. In fact, somatic therapy appears to be prescribed *more* vigorously to these patients than to patients with “pure” depression, on which the antidepressant efficacy literature was founded (17). While it may seem less problematic to withhold an unproven treatment for patients with mild rather than more severe symptoms, the underlying issue is still the same: if antidepressants are ineffective, then their prescription incurs unjustifiable exposure to risks and side effects. In addition, we imagine that few clinicians engage such patients in an informed discussion that reviews the absence of data supporting the use of antidepressant medications. This appears to represent a contradiction in practice habits compared to what occurs with mildly depressed patients and patients experiencing a grief reaction.

The present study, of course, does not address *whether* antidepressant medications are efficacious in the subgroups of depressed patients that are usually excluded from efficacy trials. Rather, we have only attempted to determine how large a group this is and consequently how frequently clinicians must consider this limitation in treating their patients. Our results suggest that this consideration may be relevant in the overwhelming majority of cases.

One potential limitation of our results is that our analyses assumed that all depressed patients were given antidepressant medications. If, in fact, many were not, then our results would not apply to antidepressant trials. To address this limitation, we reviewed the records of a subset of 147 patients with major depressive disorder for whom pre-

scription information was available. We found that more than 93% (N=137) of the depressed patients were treated with antidepressants at the time of the initial evaluation and another 3.4% (N=5) were given antidepressants within 2 months of the initial evaluation.

It should also be noted that preliminary evidence exists that suggests antidepressant medications may in fact be efficacious for some subgroups of depressed patients who are typically excluded from efficacy trials. For example, drug-placebo differences have been found in studies involving subjects with comorbid anxiety (18–20), comorbid medical illnesses (21), mild depression (15, 22–24), and chronic depression (25, 26). Unfortunately, the number of studies evaluating these less “pure” groups of depressed individuals pales in comparison with the plethora of studies that have been performed with traditional subjects.

For now clinicians need to continue to extrapolate from a literature that has focused almost exclusively on a small subset of patients with “pure” depression. It would be impossible, of course, to establish the efficacy of antidepressant medications in every conceivable subgroup of depressed patients. Therefore, some amount of extrapolation will always be necessary. Nevertheless, the present study should serve as a reminder to practicing clinicians of how little we still know regarding the efficacy of antidepressant medications in large groups of depressed patients. It also highlights how far we still have to go in our efforts to bridge the gap between efficacy research, which evaluates treatment under rigorously controlled conditions, and effectiveness research, which evaluates treatment in the “real world.”

Received May 10, 2001; revision received Aug. 14, 2001; accepted Sept. 13, 2001. From the Department of Psychiatry and Human Behavior, Brown University School of Medicine, Rhode Island Hospital, Providence; and Bayside Medical Center. Address reprint requests to Dr. Zimmerman, Bayside Medical Center, 235 Plain St., Providence, RI 02905; mzimmerman@lifefspan.org (e-mail).

Supported in part by NIMH grants MH-48732 and MH-56404.

References

- Pincus HA, Zarin DA, Tanielian TL, Johnson JL, West JC, Pettit AR, Marcus SC, Kessler RC, McIntyre JS: Psychiatric patients and treatments in 1997. *Arch Gen Psychiatry* 1999; 56:441–449
- Top 200 drugs of 1995. *Pharm Times* 1996; 62:27–36
- Brauzer B, Goldstein BJ: Symptomatic volunteers: another patient dimension for clinical trials. *J Clin Pharmacol New Drugs* 1973; 13:89–98
- Breckenridge JS, Zeiss AM, Breckenridge JN: Solicitation of elderly depressives for treatment outcome research: a comparison of referral source. *J Consult Clin Psychol* 1985; 53:552–554
- Miller CA, Hooper CL, Bakish D: A comparison of patients with major depressive disorder recruited through newspaper advertising versus consultation referrals for clinical drug trials. *Psychopharmacol Bull* 1997; 33:69–73
- Rapaport MH, Zisook S, Frevert T, Seymour S, Kelson JR, Judd LL: A comparison of descriptive variables for clinical patients and symptomatic volunteers with depressive disorders. *J Clin Psychopharmacol* 1996; 16:242–246
- Covi L, Lipman RS, McNair DM, Czerlinsky T: Symptomatic volunteers in multicenter drug trials. *Prog Neuropsychopharmacol* 1979; 3:521–533
- First MB, Spitzer RL, Gibbon M, Williams JBW: Structured Clinical Interview for DSM-IV Axis I Disorders. Washington, DC, American Psychiatric Press, 1997
- Endicott J, Spitzer RL: A diagnostic interview: the Schedule for Affective Disorders and Schizophrenia. *Arch Gen Psychiatry* 1978; 35:837–844
- Endicott J, Cohen J, Nee J, Fleiss JL, Serantakos: Hamilton Depression Rating Scale: extracted from regular and change versions of the Schedule for Affective Disorders and Schizophrenia. *Arch Gen Psychiatry* 1981; 38:98–103
- Pfohl B, Blum N, Zimmerman M: Structured Interview for DSM-IV Personality: SIDP-IV. Washington, DC, American Psychiatric Press, 1997
- Zimmerman M, Mattia JI: Psychiatric diagnosis in clinical practice: is comorbidity being missed? *Compr Psychiatry* 1999; 40:182–191
- Zimmerman M, Mattia JI: Differences between clinical and research practices in diagnosing borderline personality disorder. *Am J Psychiatry* 1999; 156:1570–1574
- Angst J, Bech P, Bruinvels H, Engel R, Helmchen H, Hippus H, Lingjaerde O, Racoagni G, Saletu B, Sedvall G, Silverstone JT, Stefanis CN, Stoll K, Woggon B: Consensus conference on the methodology of clinical trials of antidepressants, Zurich, March 1988: report of the consensus committee. *Pharmacopsychiatry* 1989; 22:3–7
- Stewart JW, McGrath PJ, Quitkin FM: Can mildly depressed outpatients with atypical depression benefit from antidepressants? *Am J Psychiatry* 1992; 149:615–619
- Ilardi SS, Craighead WE: Personality pathology and response to somatic treatments for major depression: a critical review. *Depression* 1994/1995; 2:200–217
- Bender DS, Dolan RT, Skodol AE, Sanislow CA, Dyck IR, McGlashan TH, Shea MT, Zanarini MC, Oldham JM, Gunderson JG: Treatment utilization by patients with personality disorders. *Am J Psychiatry* 2001; 158:295–302
- Tollefson GD, Holman SL, Saylor ME, Potvin JH: Fluoxetine, placebo, and tricyclic antidepressants in major depression with and without anxious features. *J Clin Psychiatry* 1994; 55:50–59
- Silverstone PH, Ravindran A: Once-daily venlafaxine extended release (XR) compared with fluoxetine in outpatients with depression and anxiety. *J Clin Psychiatry* 1999; 60:22–28
- Dunbar GC, Cohn JB, Fabre LF, Feighner JP, Fieve RR, Mendels J, Shrivastava R: A comparison of paroxetine, imipramine, and placebo in depressed outpatients. *Br J Psychiatry* 1991; 159:394–398
- Cohen-Cole SA, Kaufman KG: Major depression in physical illness: diagnosis, prevalence, and antidepressant treatment (a ten year review: 1982–1992). *Depression* 1993; 1:181–204
- Paykel ES, Freeling P, Hollyman JA: Are tricyclic antidepressants useful for mild depression? a placebo-controlled trial. *Pharmacopsychiatry* 1988; 21:15–18
- Dunlop SR, Dornseif BE, Wernick JF, Potvin JH: Pattern analysis shows beneficial effect of fluoxetine treatment in mild depression. *Psychopharmacol Bull* 1990; 26:173–180
- Pande AC, Saylor ME: Severity of depression and response to fluoxetine. *Int Clin Psychopharmacol* 1993; 8:243–245
- Kocsis JH, Frances AJ, Voss C, Mann JJ, Mason BJ, Sweeney J: Imipramine treatment for chronic depression. *Arch Gen Psychiatry* 1988; 45:253–257
- Khan A, Dager SR, Cohen S, Avery DH, Scherzo B, Dunner DL: Chronicity of depressive episode in relation to antidepressant-placebo response. *Neuropsychopharmacology* 1991; 4:125–130