Selecting Methodologies for the Evaluation of Differences in Time to Response Between Antidepressants

Stuart A. Montgomery, M.D.; Per Bech, M.D., Ph.D.; Pierre Blier, M.D., Ph.D.; Hans-Jürgen Möller, M.D.; Andrew A. Nierenberg, M.D.; Roger M. Pinder, Ph.D., D.Sc.; Frederic M. Quitkin, M.D.; Paul-Egbert Reimitz, Ph.D., Dip.Stat.; Jerrold F. Rosenbaum, M.D.; A. John Rush, M.D.; H. H. Stassen, Ph.D.; and Michael E. Thase, M.D.

Background: The delay in the therapeutic effect of antidepressants is a considerable impediment to their successful clinical use, and attention has recently been focused on antidepressant drugs that may have a faster onset of action.

Data Synthesis: Several methodologies exist for evaluating differences in time to response between antidepressants including the identification of the timepoint at which statistically or clinically significant differences between treatment groups emerge, pattern analysis, and survival analytical approaches. All have conceptual as well as practical advantages and disadvantages.

Conclusion: The survival analytical approach is generally considered to be the most rigorous and sensitive in detecting differences in the speed of response of antidepressants, but the other methodologies provide useful information.

(J Clin Psychiatry 2002;63:694–699)

Received Oct. 9, 2000; accepted Nov. 29, 2001. *From the Imperial College School of Medicine, London, United Kingdom (Dr. Montgomery); the Psychiatric Research Unit, Frederiksborg General Hospital, Hillerød, Denmark (Dr. Bech); the Department of Psychiatry, University of Florida, Gainsville (Dr. Blier); the Psychiatric Clinic, Ludwig-Maximilians University, Munich, Germany (Dr. Möller); the Depression Clinical and Research Program (Dr. Nierenberg) and the Department of Psychiatry (Dr. Rosenbaum), Massachusetts General Hospital, Boston; Organon, Inc., West Orange, N.J. (Dr. Pinder); the Department of Psychiatry, Columbia University, New York, N.Y. (Dr. Quitkin); NV Organon, Oss, the Netherlands (Dr. Reimitz); the University of Texas, Southwestern Medical Center, Dallas (Dr. Rush); the Department of Psychiatry, University of Zurich, Zurich, Switzerland (Dr. Stassen); and the Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pa. (Dr. Thase).*

The authors acknowledge the financial support of Organon Inc. for the consensus meeting from which this article was drawn.

Financial disclosure appears at the end of this article.

Corresponding author and reprints: Professor Stuart A. Montgomery, P.O. Box 8751, London W13 8WH, United Kingdom (e-mail: stuart@samontgomery.co.uk).

One of the major disadvantages of antidepressants is the apparent delay in onset of their therapeutic effect[1] since consistent significant differences between placebo and active antidepressants are generally seen only after 4 to 6 weeks. Delay in response leaves the patient exposed to risks of suicide and other morbidity associated with depression, even after the decision to treat is taken. Moreover, adverse effects are generally at their worst during this early period, and compliance with treatment can be compromised. In addition, the extended suffering of patients and their families, continuing direct health care costs, and indirect psychosocial costs associated with loss of work and damage to personal relationships are considerable. Although sustained remission is recognized as the ultimate goal of treatment, early improvement, with consequent reduction in the time spent ill, is an important aim. Improvement in the early stages of treatment has been shown to be correlated with subsequent response in a variety of studies[2,3] of antidepressants.

Attention is increasingly focused on the need for antidepressants with rapid onset of effect, and it is therefore important to identify the most appropriate methodology to investigate the relative speed of onset of action. Formal attempts to demonstrate early onset of action have been made with various antidepressants,[4,5] but the methodology has varied considerably. For example, the studies have not been consistent in defining onset of action or determining whether improvement, response, or remission should be addressed. The literature on onset of action of antidepressants is considerable; a review of studies published from 1969 to 1994 identified approximately 240 English-language articles containing information on this topic.[6] A wide range of statistical approaches had been used to evaluate the data.

To examine time to antidepressant response, this article reviews the available methodologies that have been used and the statistical approaches that have been applied, which include the traditional analyses of depression scale scores, pattern analysis, and survival analytical techniques.

METHODOLOGIES

Traditional Analyses of Depression Scale Scores

The standard method for establishing the efficacy of antidepressants has been to compare the mean response to treatment with a putative antidepressant and placebo in a double-blind group comparison, usually over 4 to 8 weeks. The repeated measures of the depression symptoms at intervals during the study are analyzed using multivariate analysis of variance.

The initial reports of early onset of response were largely based on a comparison of the mean severity scores between treatments at different timepoints. Thus, De Paula and Omer[7] in 1980 proposed that onset of effect would be defined by the postbaseline assessment at which a statistically significant difference between 2 treatment groups is seen for the first time. In a comparison with placebo, the method determines the delay in onset of action of the antidepressant and relates this delay to time of response expected from historical controls in order to assess whether onset is "early."

The method is equally useful in comparing 2 active drugs, in which the comparison determines whether there is an advantage in time to response of 1 antidepressant over another. A number of measures might be used for detecting statistically significant differences. These include the mean scores or amelioration on a severity rating scale or a comparison of the number of patients reaching a predetermined response level on the defined rating scale.

A refinement to this approach is to require that a statistically significant difference between treatments be of relevance clinically. The analysis applied by Huitfeldt and Montgomery[8] in 1983 in examining observed differences in response to 2 antidepressants introduces this additional hurdle of clinical relevance. In their analysis, they proposed the demanding criterion of a separation by 3 points on the Hamilton Rating Scale for Depression (HAM-D)[9] or the Montgomery-Asberg Depression Rating Scale (MADRS)[10] to identify probable relevance, and by 4 points as definite.

Pattern Analysis

Although originally intended as a tool to differentiate placebo response from true drug response, pattern analysis has been applied in attempts to identify early onset of action. Quitkin and colleagues originally proposed this method in 1984[11] and subsequently validated it in a replication in 1987.[12] The original 6-week, double-blind study assessed the weekly Clinical Global Impressions-Improvement (CGI-I)[13] score and classified patients as "improved" (CGI-I score of 1 [very much improved] or 2 [much improved]), "unimproved" (CGI-I score of 3 or more), or "missing" (no data available). Patients were excluded from the pattern analysis if baseline or endpoint assessments were missing or if more than one assessment was missing.

At each weekly assessment, patients were assigned a category number: 1 = improved or 0 = unimproved. Missing data were assigned the score of the bracketing weeks if the scores were the same or a 0 if they differed. Thus, a 6-digit number could be generated for each patient that shows the pattern of the patient's clinical status over time in a 6-week study. For example, a patient with the number 011111 responded during week 2 and remained well (according to the criteria of the analysis) throughout the study.

Six types of response patterns were identified: early persistent response, delayed persistent response, early nonpersistent response, delayed nonpersistent response, no response, and response only in the last week. A response was defined as persistent if an improvement was not followed by a relapse on any of the subsequent weeks. Early onset was defined as improved first at the end of week 1 or week 2, and delayed onset was defined as improved first at the end of week 3 or later. The number of patients in each of the first 4 categories can be compared, usually by using the Fisher exact test.

Quitkin and colleagues[11,12] suggested that a true drug response is characterized by a persistent pattern of improvement and that early but nonpersistent response reflected more closely a placebo effect. For evaluation of onset of action, those with an early or early persistent response are therefore of particular importance. An antidepressant with an early response would be characterized by an early and sustained response pattern, for example 011111 or 111111.

Survival Analytical Techniques

Survival analysis has been used extensively to handle situations in which subjects remain in the study for variable amounts of time. Survival analytical techniques are used to analyze longitudinal data to model the time to occurrence of a defined event. Classically, the method was used to measure relative rates of survival under different conditions with the defined endpoint as death, but the method is equally applicable to any dichotomous response variable. In this approach, the analysis focuses on the individual course of recovery, and time is treated as a function of symptom reduction.

Kaplan-Meier method. In the method proposed by Kaplan and Meier[14] in 1958, patients entered into a study are followed up during a treatment (or observation) period and assessed for the endpoint of interest. The defined endpoint may not be observed for all patients; for example, if patients discontinue prematurely, are lost to follow-up, or do not fulfill the response criterion at any time, no further information about how long they will take to respond or whether they will respond is available. Patients for whom the event of interest has not been observed are treated as "censored data" with the assumption that there is no relationship between time to response and censoring time, or, alternatively, these data can be analyzed separately. The

clinical course of patients belonging to specific treatment groups is plotted, and the Kaplan-Meier method summarizes the data set with an estimate of the cumulative distribution of times to the event of interest in the population. This is a nonparametric method that does not require specific assumptions about the underlying distribution of survival times. The Kaplan-Meier estimates of 2 (or more) survivor curves can be compared and differences tested for statistical significance using a nonparametric test such as the log-rank test or the Wilcoxon test.

Stassen modification. Stassen and coworkers[2,3,15] have applied the Kaplan-Meier approach to large data sets from multiple comparative trials of antidepressants in investigating the onset of response to treatment. The event of interest was the onset of improvement defined in each individual case on the basis of significantly reduced psychopathology scores relative to baseline. To address the durability of response, they introduced the concept of sustained response in defining the event, requiring a reduction in baseline score with no subsequent deterioration or variation only within strict limits in the achieved improvement score.

In their analyses of a large imipramine, moclobemide, and placebo database, Stassen and colleagues,[2] utilizing the total score on the HAM-D, defined early improvement as a 20% reduction of the baseline score and a sustained improvement as no more than a 15% increase observed subsequently during the study. This limit of 15% was observed as a "natural" variation in response around baseline scores from spontaneous fluctuation or error variation in the instrument or the observer. The method was also applied to response defined using the conventional 50% reduction in a pivotal scale score and applied systematically to other percentage change scores, both higher and lower.

This method of analysis was used to establish that pindolol augmentation of fluoxetine or paroxetine was associated with a significantly earlier response, where response was defined as 50% reduction on the HAM-D or the MADRS and sustained response as 50% reduction with no more than 15% deterioration for the remainder of the 6-week study.[16,17] Early response with pindolol was observed only in first and second episodes of depression but not in recurrent or chronic depression.[17,18]

Laska and Siegel modification. Laska and Siegel[19] proposed a modification of the classical Kaplan-Meier approach that gives weight to the censored data. This method considers that the assessment of time to onset has to take into account not only distribution of the time to onset for those who will obtain a response but also the proportion of the population who will not. The survival functions based on a specific clinical event criterion can be estimated using the classical Kaplan-Meier or Stassen approaches; the proportion of patients in each group who will not reach the onset criterion can be estimated on the basis of the Kaplan-Meier estimates. The hypothesis that the proportion of patients who will obtain onset is the same for all treatment groups can then be tested. If this hypothesis is not rejected (i.e., there is no difference in the percentage of subjects who ultimately will respond), a rank test such as the log-rank test or the Wilcoxon test may be used to test the second hypothesis of equal conditional survival distributions. If the proportions differ between the groups, then a nonparametric test may be used to test the second hypothesis of equal conditional survival distributions.

Depending on the clinical situation, the time of improvement or response can be defined as the time at which a patient fulfills the criterion for the first time, independent of the outcome on succeeding visits (classical analysis), or the time at which a subject fulfills the criterion for the first time and remains above the threshold or does not fall below it by more than a specific amount (Stassen modification), for some specified period of time thereafter. This may be referred to as sustained improvement or sustained response.

RELATIVE STRENGTHS AND WEAKNESSES OF AVAILABLE TECHNIQUES

The available techniques for the evaluation of onset of therapeutic action of antidepressants each have advantages and disadvantages.

Mean Score Differences

The timepoint at which statistically significant differences in mean depression scores between groups emerge is an apparently straightforward measure. However, as a method for evaluating differences between treatments, it has a number of flaws.

The measure is based on a statistically significant difference without reference to the clinical context. Sample size rather than the onset of response is likely to exert a greater influence on this outcome measure. If the data from large numbers of patients are pooled, it is possible for statistical significance to be achieved on the basis of very small differences between treatment groups that would not be considered of clinical relevance. There is also a risk of random findings due to multiple testing, particularly in studies with small samples. If this method is to be used in the context of establishing early response, the sample sizes chosen in the planning stages of the clinical trial must be based on the expected differences in response rates between comparisons of 2 active drugs or an active drug versus placebo.

A further problem is the criterion for responders. A 50% reduction in score on the pivotal severity rating scale is a generally used measure in efficacy studies to define responders. This reduction in score may be difficult to achieve by 2 weeks, particularly in a partially resistant population.

The introduction of the element of clinical significance is an advance over the original method. It makes use of a combination of statistical significance and clinical relevance in determining the timepoint for onset of action. This method avoids spurious claims of early onset of action based on small, but statistically significant, differences generated by the pooling of large numbers of patients. It can be applied to both placebo-controlled trials (to determine the early response of an antidepressant) or active-controlled clinical trials (to determine an advantage in time to response of one antidepressant over another). However, this refinement does not entirely overcome the sensitivity to sample size, nor does it avoid the disadvantage of using mean scores taken from patient populations that include early and late responders, nonresponders, and partial responders.

A number of factors may affect this kind of analysis of placebo-controlled data, such as the population studied and the size of the study, and these factors can affect the result and lead to potentially misleading claims. For example, Tollefson and Holman[5] performed a post hoc analysis of the pooled data from a large number of placebo-controlled studies and claimed that fluoxetine was associated with early response because a small but significant difference from placebo was observed at 1 week in this large data set. Imposing the additional hurdle of clinical relevance should reduce the number of claims made optimistically on the basis of small though significant differences in large samples but would still not overcome the essential flaw of relying on historical controls to define when response is normally expected.

Pattern Analysis

Pattern analysis is a powerful tool for which the persistence of response is an issue. For evaluation of onset of action, those with an early or early persistent response are of particular importance. The definitions used by Quitkin and colleagues[11] refer to improvement, although the improved category appears to correspond with definitions generally accepted for a responder. The method was originally conceived to separate true drug response from placebo response (in an area of psychiatry where placebo responses are high) rather than to determine the time of onset of action of antidepressants, and its main use may be in this area. The value of pattern analysis for determining onset of action may be compromised if the exclusion rules reduce the number of patients in the analysis substantially, which can be a particular problem in placebo-controlled studies.

Survival Analysis

Survival analysis can be used to analyze variables representing the time to an event of interest (e.g., onset of action of an antidepressant) by estimating and comparing the corresponding survival (response) functions. Because time to an event is determined for each individual patient, a more accurate and sensitive estimate of time to improvement or response is provided than would be obtained from comparing group mean scores.

In classical survival function analysis, the event (death) does not need definition. However, when the technique is used in the context of a clinical trial of an antidepressant, some consideration needs to be given to the definition of the event; possibilities could include a 20% reduction in HAM-D score (which could be defined as "early improvement") or a 50% reduction in HAM-D score (which could be classified as an "early response"). Differences in such definitions may hamper the comparison of studies.

Since the method was originally developed for evaluating survival, a key property is that only time to the event in question is evaluated. Thus, the method does not differentiate between patients who respond and then relapse and those who respond and stay well. The Stassen modification of sustained improvement is a refinement of the classical Kaplan-Meier approach and puts in place an extra hurdle that requires the response to be sustained. However, as with the classical Kaplan-Meier approach, the comparison of survival curves at any one particular timepoint, unless defined as pivotal within the protocol of the study, can be misleading.

In classical survival analysis, it is assumed that the monitored event (death) will occur in all subjects if the follow-up is long enough. In the case of antidepressant drug trials, it is clear that some patients will never respond to a given medication regardless of how long the trial is continued. Therefore, the actual time to onset in the population of patients who will respond is distorted by a population of patients providing censored data (i.e., those who have not responded at the end of the trial and who may never respond). The modification of Laska and Siegel addresses this issue.

In its various forms, survival analysis probably represents the most rigorous methodology for determining early response. It is the only method in which time is an independent variable and is therefore most suited to determining differences between treatments in time to response. In addition, in this method, each patient's individual data contribute to the overall result. This is in contrast to the majority of other methods in which mean scores are the key variables. The sustained response can be used to address the persistence of the early response, and the technique of Laska and Siegel attempts to deal with the problem of patients who will not respond to a particular antidepressant. However, the incremental utility and discrimination of these adaptations of the classical Kaplan-Meier technique may not be as great in practice as they are in theory, particularly as early improvement and overall response may be closely related. Nevertheless, survival analysis has been used successfully in investigating early response[16,17] and holds promise for the more

sensitive discrimination of differences in onset of action between drugs.

Studies for Evaluating Early Response

Double-blind controlled trials specifically designed to evaluate early response provide the best method for identifying antidepressants with fast onset. The steep reductions in depression severity scores that are a common feature after initiation of active treatment in efficacy studies may be a complication in assessing onset of effect of treatment. This apparent improvement is seen with both active drug and placebo and would be expected to make the identification of early onset of active treatment more difficult. The need to take the nonspecific effects of treatment into account in designing efficacy studies has been the subject of much discussion, and methods to limit the confounding effects, as well as refinements of trial methodology, have been proposed.[20,21] Recommended precautions include the selection of patients with depression of greater severity, selection of centers with proven experience, and selection of experienced investigators and minimizing concomitant interventions that have a nonspecific treatment effect.

The analysis will be considerably enhanced by frequent observations, particularly early in treatment, and at least 2 measures per week are required. Conventionally, depression rating scales are used at weekly intervals. However, ratings at half-weekly intervals have been used successfully in a number of studies.[16,17,22] The timepoints most frequently used for this type of study are 0, 3, 7, 10, and 14 days. Minor variations in the day of assessment, normally permitted in studies, should not be overlooked by rounding up to the day when the assessment should have been made because the data from individual patients contribute directly to the evaluation. The actual rating day should be the assessment point. Thus, a patient assessed at day 9 instead of day 10 should be included as an observation at day 9.

There is some suggestion that survival analysis techniques are more sensitive when the severity of the depression in the population is high at entry, but there is no reason why assessment of early onset should not be made with a lower severity of illness. The size of the study might, however, need to be larger to take account of the higher placebo response rate generally seen with milder depression.

It is likely that the standard techniques discussed here will be applied post hoc to existing data obtained from clinical trials that have been designed to establish efficacy at a particular endpoint (typically 4–8 weeks). Fortunately, these techniques are well suited to such applications and, in some cases, have been designed with this in mind. Nevertheless, the power of such post hoc analyses to resolve differences, in effect, will necessarily be lower than that of specifically designed onset of action studies, and differences between drugs will probably need to be large in order to be revealed by such methods.

Comparisons of test drug against placebo alone are of limited value in establishing whether a drug has an early response since this can be defined relative only to an index drug. An active comparator study will require either a large sample size or, perhaps more commonly, pooling of data from several studies. It is also necessary that the comparator drug be used under conditions of fair comparison. This means, in particular, that aggressive dose schedules should be used where indicated (for both comparator and test drug) to ensure the earliest possible improvement in the greatest number of patients.

CONCLUSION

The techniques discussed in this article can be used for establishing early response in absolute terms compared with placebo and relative to a standard comparator. To make a claim of efficacy for an antidepressant, comparisons against placebo are necessary, and these have the advantage that they require somewhat lower numbers of patients to demonstrate significant differences than do direct comparisons between active antidepressants. However, early response is a relative term, and the demonstration of earlier advantage in comparison with an established reference antidepressant, given under appropriate conditions in the correct optimum dose, provides evidence of both efficacy and relative efficacy and may be of greater interest to a clinician.

All the techniques discussed have something to offer, but on balance it would appear that the survival analysis methods have the greatest sensitivity to detect differences in time to improvement or response. Survival analysis is therefore the most appropriate for determining whether a particular drug has an early response, and consensus is emerging that this is the preferred method.

Drug names: fluoxetine (Prozac and others), paroxetine (Paxil).

Disclosure of off-label usage: The authors of this article have determined that, to the best of their knowledge, pindolol is not approved by the U.S. Food and Drug Administration for the treatment of depression.

In the spirit of full disclosure and in compliance with all ACCME Essential Areas and Policies, the faculty for this CME activity were asked to complete a full disclosure statement. The information received is as follows: Dr. Blier is a consultant for Organon, Solvay, and Pierre Fabre Medicament; has received research grant support from Organon; and is a member of the speakers advisory board for Lilly and SmithKline; Dr. Möller is a member of the speakers advisory board for a number of pharmaceutical companies; Dr. Nierenberg is a consultant for Lilly, Organon, and Glaxo; has received research grant support from Lilly, Pfizer, Organon, and Bristol-Myers; and is a member of the speakers advisory board and received honoraria from Lilly, Pfizer, Organon, Glaxo, and Wyeth; Dr. Quitkin is a consultant for Organon and is a major stockholder for Pfizer and American Home Products; Dr. Reimitz is an employee of Organon; Dr. Rosenbaum is a member of the speakers advisory board for Bristol-Myers, Lilly, Forest, Organon, Wyeth, and Pharmacia, and has received research grant support from LichtwerPharmaGmbH and Millennium Pharmaceuticals; Dr. Rush has received research grant support from Abbott, Bristol-Myers, Cyberonics, Lilly, Forest, GlaxoSmithKline, Janssen, Robert Wood Johnson, Meadows Foundation, National Institute

of Mental Health, Novartis, Organon, Pfizer, Pharmacia & Upjohn, SmithKline, Stanley Foundation, Wyeth, and Zeneca; is a consultant for Bristol-Myers, Cyberonics, Lilly, Forest, GlaxoSmithKline, Janssen, Merck, Mitsubishi Pharmaceuticals, National Institute of Mental Health, Organon, Pfizer, Pharmacia & Upjohn, Stanley Foundation, Wyeth, and Yamanouchi; and is a member of the speakers bureau for Abbott, Bristol-Myers, Cyberonics, Lilly, Forest, GlaxoSmithKline, Organon, Pfizer, Pharmacia & Upjohn, and Wyeth; Dr. Thase is a consultant for Bristol-Myers, Cephalon, Cyberonics, Lilly, Forest, GlaxoSmithKline, Novartis, Organon, Pfizer, Pharmacia & Upjohn, and Wyeth; has received research grant support from Cyberonics, Pharmacia & Upjohn, and Wyeth; and is a member of the speakers/advisory board for Bristol-Myers, Lilly, Forest, GlaxoSmithKline, Organon, Pfizer, Pharmacia & Upjohn, Solvay, and Wyeth; Drs. Bech, Montgomery, Pinder and Stassen have no significant commercial relationships to disclose relative to the presentation.

REFERENCES

1. Montgomery SA. Fast-onset antidepressants. Int Clin Psychopharmacol 1997;12:1–6
2. Stassen HH, Angst J, Delini-Stula A. Severity of baseline and onset of improvement in depression: metanalysis of imipramine and moclobemide versus placebo. Eur Psychiatry 1994;9(suppl 3):129–136
3. Stassen HH, Delini-Stula A, Angst J. Time course of improvement under antidepressant treatment: a survival-analytical approach. Eur Neuropsychopharmacol 1993;3:127–135
4. Entsuah AR, Derivan A, Kikta D. Early onset of antidepressant action of venlafaxine: pattern analysis in intent-to-treat patients. Clin Ther 1998; 20:517–526
5. Tollefson GD, Holman SL. How long to onset of antidepressant action: a metanalysis of patients treated with fluoxetine or placebo. Int Clin Psychopharmacol 1994;9:245–250
6. Derivan A. Antidepressants: can we determine how quickly they work? Psychopharmacol Bull 1995;31:23–28
7. De Paula AJM, Omer LMO. Diclofensine (RO8-4650): a new psychoactive drug: its efficacy and safety under double-blind placebo-controlled conditions. Curr Therap Res 1980;28:837–844
8. Huitfeldt B, Montgomery SA. Comparison between zimeldine and amitriptyline of efficacy and adverse symptoms: a combined analysis of four British clinical trials in depression. Acta Psychiatr Scand 1983;68 (suppl 308):55–69
9. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;23:56–62
10. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979;134:382–389
11. Quitkin FM, Rabkin J, Ross D, et al. Identification of true drug response to antidepressants: use of pattern analysis. Arch Gen Psychiatry 1984;44: 782–786
12. Quitkin FM, Rabkin J, Markowitz JM, et al. Use of pattern analysis to identify true drug response: a replication. Arch Gen Psychiatry 1987;44: 259–264
13. Guy W. ECDEU Assessment Manual for Psychopharmacology. US Dept Health, Education, and Welfare publication (ADM) 76-338. Rockville, Md: National Institute of Mental Health; 1976:218–222
14. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958;53:457–481
15. Angst J, Delini-Stula M, Stassen HH. Is a cut-off score a suitable measure of treatment outcome in short-term trials in depression? a methodological meta-analysis. Hum Psychopharmacol 1993;8:311–317
16. Perez V, Gilaberte I, Faries D, et al. Randomised, double-blind, placebo-controlled trial of pindolol in combination with fluoxetine antidepressant treatment. Lancet 1997;349:1594–1597
17. Tome de la Granja MB, Isaac MT, Harte R, et al. Paroxetine and pindolol: a randomised trial of serotonergic autoreceptor blockade in the reduction of antidepressant latency. Int Clin Psychopharmacol 1997;12:81–90
18. Berman RM, Darnell AM, Anand A, et al. Effect of pindolol in hastening response to fluoxetine in the treatment of major depression: a double-blind placebo-controlled trial. Am J Psychiatry 1997;154:37–43
19. Laska EM, Siegel C. Characterizing onset in psychopharmacological clinical trials. Psychopharmacol Bull 1995;31:29–35
20. Montgomery SA. Alternatives to placebo-controlled trials in psychiatry? Eur J Psychopharmacol 1999;9:271–276
21. Schatzberg AF, Kraemer HC. Use of placebo control groups in evaluating efficacy of treatment of unipolar major depression. Biol Psychiatry 2000; 47:736–744
22. Guelfi JD, White AC, Hackett D, et al. Effectiveness of venlafaxine in hospitalized patients with major depression and melancholia. J Clin Psychiatry 1995;56:450–458

For the CME Posttest for this article, see pages 749–750.