Notice of Correction—This article has been amended to include a list of conference participants on page 1525.

The History and Current State of Antidepressant Clinical Trial Design: A Call to Action for Proof-of-Concept Studies

Alan J. Gelenberg, M.D.; Michael E. Thase, M.D.; Roger E. Meyer, M.D.; Frederick K. Goodwin, M.D.; Martin M. Katz, Ph.D.; Helena Chmura Kraemer, Ph.D.; William Z. Potter, M.D.; Richard C. Shelton, M.D.; Maurizio Fava, M.D.; Arif Khan, M.D.; Madhukar H. Trivedi, M.D.; Philip T. Ninan, M.D.; John J. Mann, M.D.; Susan Bergeson, M.B.A.; Jean Endicott, Ph.D.; James H. Kocsis, M.D.; Andrew C. Leon, Ph.D.; Husseini K. Manji, M.D.; and Jerrold F. Rosenbaum, M.D.

Background: The development of new antidepressant drugs has reached a plateau. There is an unmet need for faster, better, and safer medications, but as placebo-response rates rise, effect sizes shrink, and more studies fail or are negative, pharmaceutical companies are increasingly reluctant to invest in new drug development because of the risk of failure. In the absence of an identifiable human pathophysiology that can be modeled in preclinical studies, the principal point of leverage to move beyond the present dilemma may be improving the information gleaned from well-designed proof-of-concept (POC) studies of new antidepressant drugs with novel central nervous system effects. With this in mind, a group of experts was convened under the auspices of the University of Arizona Department of Psychiatry and Best Practice Project Management, Inc.

Participants: Forty-five experts in the study of antidepressant drugs from academia, government (U.S. Food and Drug Administration and National Institute of Mental Health), and industry participated.

Evidence/Consensus Process: In order to define the state of clinical trials methodology in the antidepressant area, and to chart a way forward, a 2-day consensus conference was held June 21–22, 2007, in Bethesda, Md., at which careful reviews of the literature were presented for discussion. Following the presentations, participants were divided into 3 workgroups and asked to address a series of separate questions related to methodology in POC studies. The goals were to review the history of antidepressant drug trials, discuss ways to improve study design and data analysis, and plan more informative POC studies.

Conclusions: The participants concluded that the federal government, academic centers, and the pharmaceutical industry need to collaborate on establishing a network of sites at which small, POC studies can be conducted and resulting data can be shared. New technologies to analyze and measure the major affective, cognitive, and behavioral components of depression in relationship to potential biomarkers of response should be incorporated. Standard assessment instruments should be employed across studies to allow for future meta-analyses, but new instruments should be developed to differentiate subtypes and symptom clusters within the disorder that might respond differently to treatment. Better early-stage POC studies are needed and should be able to amplify the signal strength of drug efficacy and enhance the quality of information in clinical trials of new medications with novel pharmacologic profiles.

(J Clin Psychiatry 2008;69:1513–1528)

Received April 8, 2008; accepted Aug. 1, 2008. *From Healthcare Technology Systems, Madison, Wis. (Dr. Gelenberg); Western Psychiatric Institute and Clinic, Pittsburgh, Pa. (Dr. Thase); Best Practice Project Management, Inc., Bethesda, Md. (Dr. Meyer); George Washington University Medical Center, Washington, D.C. (Dr. Goodwin); University of Texas, San Antonio (Dr. Katz); Stanford University, Palo Alto, Calif. (Dr. Kraemer); Merck & Co., Inc., North Wales, Pa. (Dr. Potter); Mood Disorders Research Group, Department of Psychiatry, Vanderbilt University Medical Center, Nashville, Tenn. (Dr. Shelton); Massachusetts General Hospital, Harvard Medical School, Boston (Drs. Fava and Rosenbaum); Northwest Clinical Research Center, Bellevue, Wash., and Duke University Medical Center, Durham, N.C. (Dr. Khan); University of Texas Southwestern Medical School, Dallas (Dr. Trivedi); the Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, Ga. (Dr. Ninan); Wyeth Pharmaceuticals, Collegeville, Pa. (Dr. Ninan); the Department of Neuroscience, New York State Psychiatric Institute, New York (Dr. Mann); Depression and Bipolar Support Alliance, Chicago, Ill. (Ms. Bergeson); New York State Psychiatric Institute, New York (Dr. Endicott); the Department of Psychiatry, Weill Medical College of Cornell University, New York, N.Y. (Drs. Kocsis and Leon); and the Laboratory of Molecular Pathophysiology, National Institute of Mental Health, Bethesda, Md. (Dr. Manji). Ms. Bergeson is currently affiliated with Consumer Affairs, United Behavioral Health.*

Supported by unrestricted grants from AstraZeneca Pharmaceuticals LP, Bristol-Myers Squibb Company, Forest Research Institute, Johnson & Johnson, Merck & Co., Inc., Pfizer Inc., sanofi-aventis, USA, Inc., Sepracor, and Wyeth Pharmaceuticals.

This article represents a consensus of the participants in the conference "Advancing Signal Strength in Proof of Concept Studies in Major Depression," which was convened by the University of Arizona Department of Psychiatry and Best Practice Project Management, Inc. It took place at the Clark Building in Bethesda, Md., on June 21–22, 2007.

The authors acknowledge Heather Hopkins, Desert Editors, for writing assistance. Ms. Hopkins reports no financial or other relationship relevant to the subject of this article.

Financial disclosure, disclaimers, and conference participants are listed at end of the article.

Corresponding author and reprints: Alan J. Gelenberg, M.D., Healthcare Technology Systems, 7617 Mineral Point Rd., Suite 300, Madison, WI 53717 (e-mail: agelenberg@Healthtechsys.com).

Reprinted with correction to page 1525.

Antidepressant medications have provided a major benefit to countless patients since the mid-20th century. However, if there is to be further progress in developing antidepressant drugs, we must seriously rethink and reconceptualize future study designs. A draft report from the Agency for Healthcare Research and Quality concluded that current antidepressants produce symptomatic remission in only one-third of depressed patients with their first course of medication.*[1–4] Some meta-analyses of large data sets suggest small differences in efficacy among the classes of antidepressants.†[2,5–8] Yet, none of these differences can be considered clinically meaningful, and there is a great unmet need for better medications to treat major depressive disorder (MDD).

The primary end points in clinical trials have not changed substantially over the past 40 years. There are no generally accepted surrogate markers of disease or treatment response, and there is no consensus on the moderating or mediating factors that might affect response to active drug or placebo.[10,11] At least half of recent antidepressant clinical trials have been negative due to a high placebo response rate.[12]

For more than a decade, the selective serotonin reuptake inhibitors (SSRIs) and the serotonin-norepinephrine reuptake inhibitors (SNRIs) have been the most widely used antidepressants. There have been no recent breakthrough developments in novel mechanisms of action or enhanced effect size. For example, since 2002 the only "new" antidepressants to be introduced in the United States are escitalopram (a stereoisomer of an already marketed drug, citalopram), duloxetine (a relatively older molecule that resurfaced to become the second member of the SNRI class in the United States), and a patch-delivered formulation of selegiline (an irreversible monoamine oxidase inhibitor [MAOI] first synthesized in the 1970s). New drugs emerging since 1991 have fallen by half each year, even though there has been a doubling of research supported by the National Institute of Mental Health (NIMH) and an abundance of new biotechnical tools.[105] The same animal models that were used 40 years ago to screen for potential antidepressants are still being used today. Drug development costs are increasing, yet the number of drugs resulting in new drug applications is decreasing (see U.S. Department of Health and Human Services,[13] Figure 1).

Certainly, most antidepressant drug discovery and development efforts fail because MDD is a heterogeneous illness, and the pathophysiology of the disorder (or any potential subcategory) is incompletely understood. As long as MDD is defined phenomenologically rather than biologically, it will be impossible to specify pharmacodynamic effects that reliably predict improvements in clinical outcome. Clinically, we need a more precise diagnosis, more detailed analysis of the behavioral profile of depression, and more effective and safer treatments. There is a great need for antidepressants that (1) work faster, (2) have higher rates of response and remission, (3) continue working longer, (4) work in partially responsive or refractory patients, and (5) have fewer side effects and potential drug-drug interactions. The situation is approaching an impasse, and, in an era in which relatively safe and inexpensive SSRIs have become the standard for comparison, fewer companies may choose to invest in developing new drugs until there are more robust predictors of clinical efficacy. If payers will not reimburse for the extra cost of a novel, patent-protected medication because of the lack of evidence of a differentiating advantage, the market for new antidepressants will decrease. Moreover, until new clinical trial methods are developed and validated, hypothetical new drugs that potentially could be more efficacious for a subset of patients with MDD might not be discovered.

Bottlenecks in the overall drug development process highlighted by the U.S. Food and Drug Administration (FDA) include a limited number of biomarkers and surrogate end points for predicting clinical efficacy in animal models and early-stage clinical testing.[14] This is certainly true for antidepressants. At this time, the primary point of leverage to advance antidepressant drug development is to enhance the quality of information in early-stage proof-of-concept (POC) studies to amplify the signal of drug efficacy and to control better for the impact of the placebo effect. Progress in this area might suggest more relevant animal models for screening antidepressant drugs, as well as improved methodology for pivotal phase 2 and phase 3 trials. Most if not all currently available preclinical models of depression were based on effects of monoaminergic compounds (MAOIs, tricyclic antidepressants [TCAs], SSRIs, and SNRIs), suggesting the potential for a monoaminergic bias. Finally, while these preclinical models have been useful in screening for different classes of monoaminergic compounds, their utility with novel agents needs to be demonstrated.‡[15–20] Importantly, no extant preclinical behavioral paradigm qualifies as a validated animal model for the pathophysiology of MDD, or for a specific disaggregated component of the syndrome.

In an attempt to encourage informed innovation in the design of POC studies that might amplify signal strength of drug efficacy in antidepressant trials, the University of Arizona Department of Psychiatry and Best Practice Project Management, Inc., convened a consensus development conference on June 21–22, 2007, in Bethesda, Md.

*In a publication that appeared 6 months after this conference, Turner et al.[4] reported that the "true" effect size of antidepressant medications based on all published as well as unpublished results is smaller than in the published literature.

†Other studies offer some evidence that various drug types affect different symptoms.[9]

‡Interestingly, many novel mechanisms may also be effective in these preclinical behavioral paradigms, including mGlur5 negative allosteric modulators, vasopressin antagonists, and multiple other neuropeptidergic mechanisms.[15–20]

Reprinted with correction to page 1525.

Attendees included 45 participants from academia, the National Institutes of Health (NIH), consumer advocacy, the FDA, and the pharmaceutical industry. This article represents a consensus of the participants.

HISTORY OF ANTIDEPRESSANT CLINICAL TRIALS

Antidepressant clinical trials began in the 1960s with small samples of severely ill inpatients and large effect sizes of drugs versus placebo. In the past 2 decades, antidepressant clinical trials have moved to large outpatient populations with resultant small to moderate average effect sizes.[21] Tricyclic antidepressants and MAOIs dominated the early years of antidepressant drug development, leading to monoaminergic-based theories of the pathophysiology of MDD. The middle generation of antidepressants included maprotiline, amoxapine, nomifensine, bupropion, and trazodone, with TCAs as the reference standard. When SSRIs were first tested, TCAs were used as the active comparator. Most recently, the SSRIs have become the standard of comparison for other novel antidepressants. While monoaminergic-focused pharmacotherapy still dominates the antidepressant pharmacopoeia and drug pipeline, extant data no longer support a monoaminergic "deficit" as an adequate explanation for MDD.[22]

Most early inpatient trials found TCAs to be superior to placebo. Since then, the capacity to demonstrate superior efficacy of a new putative antidepressant compared with placebo has steadily declined. From 1985 to 1997, 55% of new and 42% of standard antidepressant trials from the FDA Summary Basis of Approval (SBA) data sets failed to demonstrate significant drug versus placebo differences.[23] In a review of published studies between 1980 and 2000, the percentage of patients assigned to TCAs or SSRIs who had a greater than 50% improvement rose each year, but the percentage of patients assigned to placebo who showed this level of improvement rose as well, with a relatively greater increase in placebo response rates.[21] The placebo response rates in these trials varied from 10% to 50%, with a disproportionate number of the more recent studies at the higher end of this distribution. One consequence of these temporal trends is that sample sizes of 30 in studies conducted in the 1960s produced the same power we now get from sample sizes in the hundreds, due to shrinking drug-placebo differences (i.e., the between-group effect size).

What explains the growing number of failed antidepressant drug trials? First was a secular trend in the population of patients available for studies. Changing care delivery in the community and growing awareness of the availability of treatments for depression have altered the pool of subjects for antidepressant research. Although rating-scale scores look comparable across the decades, current study populations are less severely ill than earlier inpatient populations in other respects, such as level of global or functional impairment. Fewer severely ill patients are available for studies now—unless they are nonresponders to prior treatment. Existing randomized, controlled trial (RCT) designs are inadequately sensitive to drug-induced changes in the less severely ill (see, for example, Thase et al.[8]). Also, the placebo response rate is greater in patients who are less severely ill. Thus, larger studies are now required to achieve the statistical power of earlier studies. However, larger studies are as likely to fail due to a proportional offsetting increase in error variance resulting from the greater number of sites and/or faster pace of sample accrual. Factors at both ends of the severity spectrum diminish the ability to detect meaningful drug-placebo differences: mildly ill patients have a higher placebo response rate, and if multiple treatments have failed in severely ill patients before they come to clinical trials, they are less likely to respond to the next therapy.

Second, the diagnosis of depression has expanded. In the early days of modern psychopharmacology, the emphasis was more dimensional, with a focus on core features of functionally impaired subjects. The move to categorical diagnostic criteria allowed diverse presentations to qualify for MDD. Two patients can now fully meet the MDD criteria without a single overlapping symptom. For example, one patient could have depressed mood, weight gain, hypersomnia, fatigue or loss of energy, and recurrent thoughts of death, while another exhibits markedly diminished interest or pleasure, significant weight loss, insomnia, psychomotor retardation, feelings of worthlessness or guilt, and diminished ability to think or concentrate. A drug with a specific biochemical or neuroanatomic mechanism may not produce the same result in these 2 different individuals.

Rating Scales, Measuring Symptoms, and End Points

The first studies of antidepressant effects used global ratings, even before the Clinical Global Impressions scale was developed. The publication of the Hamilton Rating Scale for Depression (HAM-D) in 1960 was a watershed event.[24] The original had 21 total items, 17 of which were to be scored. Hamilton recommended that 2 experienced psychiatrists rate the scale independently, and their ratings were summed to get a total score. The recommendations changed over time to an average of the 2 scores and were then further streamlined to a single rater with established reliability.[25] The HAM-D has been criticized for not fully covering the diagnostic criteria for MDD.[25–27] It was developed more than a decade before the introduction of the first sets of operationalized diagnostic criteria, and fully 20 years before the publication of the *Diagnostic and Statistical Manual of Mental Disorders,* Third Edition (DSM-III). The HAM-D was intended to give a broad appraisal of the overall severity of depression, not to replace a diagnostic evaluation or clinician's global impression, and it was supposed to be used by experienced clinicians

Reprinted with correction to page 1525.

only. It measured symptoms most commonly seen in Hamilton's inpatient unit; that is, those of older, more severely ill, and less treatment-resistant patients. For this reason, the HAM-D is more circumscribed and inconsistent in its coverage of the symptoms that often characterize milder depressive episodes, and it is less useful in teens and children than in adults. The rating scale is at its sharpest in moderate to severe depression, losing precision as severity increases.[28]

In addition to the above, the HAM-D is heavily weighted toward anxiety and somatic symptoms. There is no measure of hostility, even though hostility was once considered a defining characteristic of major depression[29–31] and early studies of antidepressants highlighted reduction in hostility as the initial indicator of antidepressant efficacy.[32–34] There also is little coverage of symptoms related to motivation and enjoyment, key elements of the depressive spectrum. Many of its items are extraneous to treatment outcome. The HAM-D asks just 1 question about functionality, but functionality is the issue of greatest concern to mental health consumers (Depression and Bipolar Support Alliance, unpublished survey results). All of these criticisms notwithstanding, the HAM-D continues to be widely used.

There have been many revisions to the HAM-D. In 1967, Hamilton[35] added the cognitive triad of hopelessness, helplessness, and worthlessness to create a 24-item scale. Guy[36] developed brief standardized diagnostic probes, although these too were global ratings and largely impressionistic. Miller and colleagues[37] were the first to publish a structured interview for the HAM-D. They also were the first to attempt to add reverse vegetative symptoms, such as hypersomnolence, hyperphagia, and weight gain.[37] An even more detailed structured interview guide was developed in 1988 by Janet Williams (Structured Interview Guide for the HAM-D [SIGH-D]),[38] and subsequently this was supplemented for the reverse vegetative symptoms (Structured Interview Guide for the HAM-D–Seasonal Affective Disorder version).[39] Thase and colleagues[40,41] at the University of Pittsburgh Medical Center also developed supplemental HAM-D items rating reverse vegetative symptoms for use in studies assessing the efficacy of MAOIs, although ultimately they suggested using it as a "stand alone" for studies of patients with prominent vegetative reversal.[40] While there are several alternate versions of the HAM-D, the 17-item version remains the most widely used. On different versions, some items are scored differently.[42]

After almost 50 years, the HAM-D retains its primacy as a measure of the overall severity of MDD. Although sometimes referred to in jest as the "lead standard" because of its imperfections, the HAM-D contains a valid, single dimensional "core" subscale of 5 or 6 items.[43–46] The structured interview guides improve the precision of administration, and outstanding reliability coefficients can be achieved if raters are trained properly.[38] However, what is possible under optimal circumstances and what happens in large multicenter trials may be highly divergent: rater training for industry-sponsored trials can be variably effective, with less-than-adequate attention to establishing and maintaining reliability over time. The impact of imprecise ratings of the primary dependent measure can have a devastating effect on RCTs: with a reduction in reliability, there is a corresponding drop in power and design sensitivity.[47] That translates into requiring even larger sample sizes to assess the efficacy of a new drug.

Beyond the HAM-D, the Beck Depression Inventory (BDI)[48] and the Montgomery-Asberg Depression Rating Scale (MADRS)[49] have the next longest track records. The BDI, now used exclusively as a self-report measure and more likely to be included in a study of psychotherapy than pharmacotherapy, differs from the HAM-D in that it is dominated by cognitive symptoms of depression. Self-report assessments such as the BDI may be vulnerable to score inflation by a subset of patients with more marked subjective distress, a potential limitation that may be amplified by the emphasis on cognitive symptoms of depression (i.e., pessimism and hopelessness). For example, in one study patients who had disproportionately high BDI scores (as compared with HAM-D scores) were relatively nonresponsive to pharmacotherapy.[50] The BDI has minimal coverage of neurovegetative symptoms of depression. The self-report is more useful for some purposes, like screening or office practice, than others, and historically, self-report measures have not been preferred for use in studies conducted for regulatory purposes. However, FDA officials have recently indicated that self-report is a valid measure in assessing outcomes.[106,107]

The symmetric structure and equal weighting of items on the MADRS give it a slight edge over the HAM-D in psychometrics. The MADRS is less "imbalanced" than the HAM-D with respect to ratings of anxiety and, as such, may be less affected by sedating or nonsedating effects of antidepressants. The unidimensional core of MADRS also consists of a larger proportion of the total scale items. Despite these strengths, however, the MADRS has not shown obvious superiority to the HAM-D in demonstrating antidepressant drug effects.[51]

Additional alternatives to the HAM-D, BDI, and MADRS include the Inventory of Depressive Symptomatology (IDS; Clinician [C] and Self-Report [SR] and regular and Quick [Q] versions)[26] and the Patient Health Questionnaire. Each offers unique strengths over the HAM-D (assessment of the full DSM syndrome criteria), but for this small improvement, one loses continuity across nearly 50 years of research with the HAM-D, and, for the most part, investigators have selected the older standards, using newer scales as secondary measures. Some experts believe it is now time to break with tradition, while others are more conservative, favoring

Reprinted with correction to page 1525.

retention of the HAM-D. Some people have tried to adapt the HAM-D for self-report, because self-report, in concept, can take out the clinician variable. It is more cost-effective to capture self-report in an electronic format, such as Interactive Voice Response or Web-based tools.

The results of RCTs in depression reflect not only rating scales, but how the scales are administered. As mentioned previously, reliability and standardization of ratings are seldom addressed or monitored. Expectations and demand characteristics insidiously and pervasively distort pretreatment and posttreatment ratings. Depression ratings drift downward over the study's course as staff become more familiar with the patient and provide interpersonal support and as patients and raters expect improvement.

In addition, most studies offer financial incentives that may inflate initial ratings. If excluding a patient results in a $500 payment, while enrollment produces up to $14,000 in fees for a multistage trial, one can understand how an on-site evaluator may be more likely to obtain a score of 21 (i.e., above the enrollment threshold) than an independent, off-site evaluator, particularly when the distinction is relatively subtle (e.g., the difference between a score of 21 and a score of 19—i.e., just below the threshold). As intake severity threshold scores have increased in an effort to improve signal detection, a countervailing bias (deliberate or unconscious) toward elevating scores to get more patients into the study may be operating. Another secular trend that may affect the sensitivity of the measures could be related to the fact that less-well-trained personnel perform these clinician-rated measures in clinical trials.

One more issue is the difference between how clinicians define outcomes (i.e., symptom reduction or response or remission rates) and what patients define as improvement. Clinicians define recovery by decreases in irritability and suicidal thoughts and increases in sleep and appetite. Patients, on the other hand, want to have meaningful work; satisfying relationships; safe, stable housing; and hope (Zimmerman et al.[52] and Depression and Bipolar Support Association, unpublished survey results). The severity of depressive symptoms is only partially and variably related to functional impairment.[53] With current depression end points, symptomatic improvement does not necessarily result in improved functional capacity.[53] The breadth of depression assessments could be expanded by adding global functioning, quality of life, social and vocational functioning, sexual functioning, and selected comorbidities, such as anxiety, pain, or insomnia. This breadth is most easily attained by the addition of brief, but comprehensive self-report scales.

Given the heterogeneity of depressed patients, the clinician-patient differences in expectations, and the lack of concordance between symptomatic improvement and function, some antidepressants that appear to perform well in research studies have been disappointing in practice—and in the market (such as trazodone, nefazodone, moclobemide, and reboxetine). This poor performance has been related, in part, to side effects and adverse events. While statistical significance in RCT demonstrations of efficacy is important, it may not translate to clinical significance, effectiveness, or tolerability.

Site Selection

When modern psychopharmacology was "born" in the 1950s, clinical investigation was conducted by scientists at academic centers or research sites in public and private mental hospitals. A majority of patients in contemporary RCTs are enrolled by investigators at private for-profit outpatient sites. Typically, operational groups within pharmaceutical companies choose trial sites. The decision makers usually focus on throughput (performance metrics) based on prior experience. Quality may be sacrificed for speed of enrollment. Since speed is critical in the cost of drug development, many large pharmaceutical companies have developed their own databases on site performance and whether a site can discriminate an active drug from placebo. Site differences in differentiating drug and placebo response in phase 3 antidepressant trials have been well documented.[54] Information on quality and performance metrics is considered competitive and not shared between companies. While site metrics for pivotal trials are reasonably well understood, site requirements for POC studies with novel antidepressants need to be determined, including access to specific populations or subpopulations, quality/reliability of clinician ratings, and potential access to cutting-edge technology that may suggest surrogate end points. Because of their access to research resources, academic sites should be considered prime locations for POC studies, provided that sponsors can be certain that the responsibilities of investigators are not delegated to residents and inexperienced staff and that the sites can meet reasonable performance (speed of recruitment) requirements and cost considerations. Drug company sponsors and contract research organizations (CROs) should not automatically use the metrics appropriate for pivotal trials in evaluating sites for POC studies.

Placebo Response Rates

More than many other central nervous system disorders, depression is sensitive to the placebo effect, which presents a major challenge to the development of novel antidepressant drugs. An analysis of the FDA SBA reports for 11 approved antidepressants between 1985 and 2005 found that the magnitude of placebo response was the single most powerful predictor of the outcome of an antidepressant trial.[55] If the magnitude of symptom reduction among depressed patients assigned to placebo was more than 30%, the chance of significant separation from antidepressant was only about 21%. By contrast, among the studies with lower placebo response rates, the chance of a "positive" study (i.e., a statistically significant effect

Reprinted with correction to page 1525.

favoring drug over placebo) was 74%. It is therefore noteworthy that placebo response rates in contemporary RCTs typically range between 40% and 50%, which almost single-handedly explains the high rate of failed studies.

One early attempt to improve signal detection involved use of a 1- or 2-week lead-in period of placebo therapy prior to beginning the RCT. The logic of this strategy is straightforward: if the patients who were most responsive to nonspecific factors could be identified before randomization, the subsequent placebo response rate would be expected to be lower. However, results of one meta-analysis indicated that the placebo lead-in strategy did not improve signal detection,[56] perhaps because of both the brevity of the intervention and the fact that placebo has typically been administered single blind. As fewer than 10% of patients typically respond during a single-blind placebo lead-in,[56] it may be that the "unblinded" clinicians inadvertently convey low expectations of benefit to their study patients. It remains to be seen if alternate lead-in strategies, such as a longer period of double-blind placebo therapy or using nonresponse to an initial course of psychotherapy as a means to identify depressed people more likely to "require" pharmacotherapy,[57,58] are cost-effective ways to enhance signal detection.

It is almost axiomatic that more severely depressed patients are relatively more responsive to antidepressants than to placebo. Regrettably, attempts to capitalize on this observation by requiring that studies only enroll patients with moderate-to-high levels of symptom severity appear to engender inflation of pretreatment severity ratings, which can offset or even reverse the predictive value of pretreatment severity ratings. For example, Khan et al.[59] recently found that higher pretreatment symptom severity ratings only predicted larger effect sizes when the studies did not utilize high severity inclusion criteria.

Acuity and chronicity are also important factors. Whereas short episodes may be the most likely to remit spontaneously or respond to placebo-expectancy interventions, patients presenting with more protracted depressive episodes may be less likely to respond to standard interventions. Although prospective empirical confirmation is lacking, some of the experts in attendance at the meeting suggested that the best signal detection occurs in depressed people who have been ill for more than 6 months but fewer than 3 years.

Placebo response rates in antidepressant trials also have been shown to be linked to several methodologic or study design facts. For example, studies that use flexible dosing protocols are more likely to detect significant drug versus placebo differences (as compared to studies using fixed-dose protocols), as are studies that compare fewer treatment arms (vs. those with 4 or more arms). These particular findings are interrelated: flexible-dose studies typically have fewer arms than those using fixed-dose protocols. One practical implication of these observations is to emphasize simpler 2- or 3-arm studies and flexible dosing protocols in POC trials.

Finally, there are ethical issues to consider when designing a study involving administrating placebo to some patients with MDD. There is reasonable justification for the use of placebo in short-term (i.e., up to 12-week) trials, based on the high response and remission rates observed with placebo and the opportunity for responders to be spared an unnecessary medication exposure. Additionally, nonresponders may subsequently be offered active treatment. But it is much harder to justify longer-term use of placebo in MDD trials. In such cases, for ethical reasons nonresponders must have easy opportunities for removal from protocol and rapid treatment with an established therapy.

Trial Duration

Initial response to antidepressants occurs over a variable time period. Some trials have suggested that early improvement can be detected within days.[33] Symptoms may continue to improve over 8 to 12 weeks or longer. Some argue that longer clinical trials allow time for more subjects to respond to treatment and for placebo response to dissipate, strengthening the treatment signal. In 1984, Quitkin et al.[60] concluded that improvement during the first 2 weeks was due to the placebo effect and that the "true" clinical action of antidepressants was delayed for several weeks. However, more recent studies[61,62] and meta-analyses[63–65] suggest that clinical efficacy can be detected in treatment-responsive patients within 1 to 2 weeks. Indeed, the failure of an antidepressant to bring about ≥ 20% improvement in symptoms during the first 2 weeks is a good predictor of patient nonresponse at 6 to 8 weeks.[9,65–68]

There are statistically significant differences between changes in mood and interest symptom clusters in early responders, late responders, and nonresponders in the first 4 weeks of treatment.[69] Subjects who are going to have a late response continue to improve in the mood symptom cluster, but those who are not going to respond tend to "flatten out" at 3 weeks. If a patient has less than a 20% improvement by 6 weeks, there is little point in continuing treatment with the same medication.[70] The strength of the evidence argues for shortening clinical trials for new antidepressants to 3 to 4 weeks from the current 6 to 8. In addition, prolonged placebo treatment periods raise issues about feasibility and dropouts that may be fatal to a study.

ALTERNATIVE MODELS AND OUTCOME METHODS FOR ANTIDEPRESSANT CLINICAL TRIALS

The established model for the design of an antidepressant clinical trial is "disorder-specific." This model is

Reprinted with correction to page 1525.

based on the theory that depression is a unitary disorder. The antidepressant drug is presumed to target the specific pathophysiology underlying the illness. A test of the drug's efficacy then requires measuring its capacity to reduce the severity of the "whole" syndrome. An alternate model is "component-specific." It is based on the premise that depressive disorder is not unitary but multifaceted, comprised of major behavioral and affective elements that interact to create the depressed state. Thus, the therapeutic effects of a given drug may be differentially effective on specific subcomponents of the disorder, such as anxiety, hostility, and motor functioning, or on items other than the "core" pathology of depression. Indeed, these changes may not be detected if only broad rating instruments are used. In theory, one could tailor new treatments to the different pathophysiology underlying different depression symptoms or components, but data supporting this linkage are limited. Some investigators report that the initial actions of serotonergic drugs, such as most of the SSRIs and TCAs, are on anxiety and hostility, while the noradrenergic drugs (for example, desipramine) act first on motor activity and anxiety.[9,59] The findings of earlier action on components of the disorder (vs. total severity or global functioning)[9] have been confirmed in clinical studies and in meta-analyses of clinical trial results.[63] Going forward, a multifaceted approach to assessing the components of depression in POC studies could apply several different types of measurements[67]—including novel clinical ratings, self-report measures (including new electronic versions), psychomotor performance, cognitive function, personalized outcomes, functional outcomes, video evaluation, genetics, and brain imaging.

Adaptive treatment clinical trial designs are becoming popular in some therapeutic areas.[71] In these designs, clinical decisions, and adjustments in clinical care, are based on adaptive threshold-dependent algorithms (the individual's needs), as well as the mechanism of action of the drug being tested and the expected response time. Responses and side effects are monitored over time. Certain benchmarks are established in advance that meet temporal and symptomatic criteria for clinical decision-making, such as when to raise or lower the dose, augment, or switch treatments. Recent literature advocates designing trials to assess optimal adaptive strategies.[72,73] In a shorter trial using an adaptive trial design, time points for decision-making should be fixed (in advance of the start of the trial) at 3, 4, and 6 weeks. Given response heterogeneity and delayed effects in clinical trial and "real-world" populations, it is necessary to individualize treatment decisions. These measurement-based care approaches have been used recently in practical clinical trials like the NIMH-funded Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial.[72,73] Such measurement-based care approaches have an added advantage of identifying maximally tolerated doses and not settling for minimal effective doses as is commonly done in industry-sponsored studies.* Adaptive treatment strategies provide repeated adjustments based on past and present treatment-related effects and may enable shorter trials.

POTENTIAL TARGETS FOR ANTIDEPRESSANT ACTION

Moderators and Mediators

A treatment moderator is a pretreatment variable that identifies patients who respond differentially to treatments in terms of specific treatment outcomes. Potential moderators include clinical cross-sectional measures at baseline, such as anxiety, hostility, cognitive functioning, and severity of depressed mood; historical measures like age at onset, duration of the current episode, degree of familial aggregation, and recurrence; biological measures; and even programmatic measures (for example, whether family intends to be involved in the treatment program or not).

Treatment mediators, on the other hand, are possible mechanisms through which a specific treatment might work. Mediators are variables reflecting change or events during treatment that might explain the degree of treatment effect. Possible mediators include adherence to treatment, the degree of family involvement during treatment, and changes in biological or behavioral response that presage changes in outcome. Mediators may suggest how the treatment might be improved or made more cost-effective for future studies.

If moderator variables could be identified, patients could be selected for whom a specific treatment is most likely to be effective. Existing data sets with contrasting treatments and a large enough sample could be used to try to identify moderators. Well-documented data sets from industry, NIMH, and individual studies at academic centers might be useful for this purpose. Unfortunately, many of the available data sets have limited coverage or evaluate a limited subset of the psychological, behavioral, and physical symptoms of depression. A moderator can be identified only if it has been defined and quantified prior to initiation of treatment. Access to the richest data sets, including some large studies on chronic depression, is limited.

Another problem in trying to identify moderators is that many variables lack a standard definition. For example, age at onset could mean age at first major

*Unless critical decision points with required dosage escalation (in the absence of dose-limiting side effects) are specified, the mean dose of the medication may be too low, leading to a reduced response rate. In fact, the desire to minimize reported side effects, which may adversely affect marketability, may actually lead companies to use lower doses in clinical trials than might be ideal. This is especially critical in POC trials, which may need to use aggressive dosing strategies in order to maximize the chance to detect drug-placebo differences.

Reprinted with correction to page 1525.

Table 1. Subtypes of Depression[a]

Subtype	Characteristics	Suggested Treatment
Melancholic depression	More common in inpatients and elderly Greater illness severity and impairment	Antidepressants with dual reuptake or combination of SSRI and SNRI[76]
Atypical depression	Younger age at onset Longer index episode More chronic course Two of the following: hypersomnia hyperphagia intense lethargy/fatigue (leaden paralysis) hypersensitivity to rejection/criticism	MAOIs[77] Augmentation with T_3[78]
Depression with anger attacks[79,80]	High hostility, somatization, and anxiety Blunted neuroendocrine responses to serotonergic challenges[81] Structural brain differences in white matter hyperintensities[78]	Serotonergic antidepressants[80]
Depression with prominent insomnia	Lower remission and response rates[82]	Combination of antidepressant and hypnotic[83]
Anxious depression	Lower remission and response rates[72,80,84,85]	

[a]Other subtypes such as bipolar depression, late-life depression, depression related to stroke or other brain disease, or depression associated with concomitant medical illness were not addressed at this conference.
Abbreviations: MAOI = monoamine oxidase inhibitor, SNRI = serotonin-norepinephrine reuptake inhibitor, SSRI = selective serotonin reuptake inhibitor, T_3 = triiodothyronine.

depression, first symptoms of depression, or current episode. The same is true for duration of disorder. There may be differences in how these variables are defined between sites or even between raters at the same site. Unless it is clear what is being analyzed, it will be difficult to obtain accurate results and impossible to pool data from different studies. Identifying potential moderators can generate hypotheses, not conclusions. These hypotheses would have to be cross-validated with other data sets or with new, prospective studies.

Mediators are even less well studied than moderators. In concept, the therapeutic alliance between patient and therapist could be a mediator in psychotherapy outcomes. Improvement in maternal depression may mediate improvement in depressive symptomatology in their at-risk offspring.[74] The tryptophan depletion paradigm could help to clarify serotonin's role as a possible mediator of antidepressant effects.[75]

Subtypes

Another approach to increase detectable effect sizes is to examine subtypes of depression (Table 1). Subtypes may be characterized by distinctive biological and clinical features, or, at the least, symptom differences may have therapeutic and research relevance. Some may moderate response, and there may be differential therapeutic effects on specific symptoms.

A history of early life adversity may affect treatment response.[86] Other research is investigating the possibility that genetic polymorphisms—e.g., brain-derived neurotrophic factor, serotonin transporter, 5-HT_{1A}—may be moderators of drug response.[87,88]

Depressed medically ill patients may not constitute a subtype, but they need to be thought of differently in terms of a potential therapeutic target. Depression is more common in the medically ill than the non–medically ill, and some biological mechanisms may be congruent between being ill and being depressed. Patients with more severe medical comorbidity have lower response and remission rates compared to patients with less medical comorbidity.[78]

Industry has historically been reluctant to target a subtype of depression that could restrict the total number of potential patients for a drug. Yet, identifying a subtype that comprises, for example, 20% to 30% of the depressed population, for which a new drug will reliably work could allow focus on a solid market for that drug and an important niche. Further, identifying a discrete subtype could improve signal detection in clinical antidepressant trials, thereby reducing the overall costs of drug development.

BRAIN IMAGING AND GENOMIC ANALYSIS

Someday, imaging may explain some of the variance in clinical outcome on antidepressants and contribute to a biological classification of mood disorders that complements current phenomenology. Positron emission tomography (PET) could help investigators to understand disease etiology, identify treatment targets, choose the right dose range, monitor antidepressant action, provide biological surrogates or biomarkers for efficacy, identify subtypes, and eventually aid in selecting the best medication for individual patients.*[89–91] Brain-imaging–based biological classification would have substantial implications for drug

*Beyond neuroimaging, some studies of slow-wave activity look promising with respect to identifying antidepressant responsive profiles as well as placebo responsive profiles.[89–91] Localized slow-wave activity may represent a trace of synaptic potentiation.[92] This has been demonstrated with learning paradigms and visual motor tasks.[93]

Reprinted with correction to page 1525.

development, including the ability to identify patients who are more likely to respond to antidepressants in general or to specific classes of antidepressants, thereby increasing the signal strength in antidepressant trials. Examples are evidence that low transporter binding and high 5-HT_{1A} autoreceptor binding (related to a higher expressing promoter variant in the 5-HT_{1A} gene) detected on PET scanning predict a poorer antidepressant medication response or remission rate.[94]

Genomic information also could point the way to new moderators of treatment response in POC trials. Depression is undoubtedly heterogeneous, begging for genetic dissection. Pharmacogenetics is the study of genetic variations and how they relate to variable drug response or alternatively variable adverse effects. It includes the genetic polymorphism of drug transporters, drug metabolizing enzymes, and drug receptors. Pharmacogenomics is a subset of genetics that identifies individual differences in drug pharmacokinetics and pharmacodynamics related to host genes. It aims at identifying disease genes and new drug response markers at levels of drug absorption and metabolism, drug target, or disease pathway.[95] Pharmacogenomics usually focuses on sequence variations in particular genes, such as single nucleotide polymorphisms (SNPs). Pharmacodynamic predictors and adverse event prediction can help either select a certain patient for a certain treatment because they will respond to that treatment or select them out of a treatment because they will have an adverse outcome. An intronic variation of the 5HT2A polymorphism is probably the marker most ready for use. Data from the STAR*D pharmacogenomic study suggest that this variation is associated with poor response to the SSRI citalopram.[96]

CONSENSUS FINDINGS AND RECOMMENDATIONS

Based on the review of the state of antidepressant drug development and clinical trials design in this field, conference participants developed a set of recommendations to advance discussion of, and planning for, future POC studies. The recommendations focused on 4 general areas: study design, study conduct, data analysis, and future research.

Improving Study Design

1. Relevant designs for POC antidepressant trials should be restricted to simple, small studies that test a single hypothesis.
2. Trials should be shorter and include triage to drop nonresponders in the first several weeks of the study.
3. POC studies should be double blind. Open POC trials would be misleading.
4. A series of POC studies can start with a small pilot study to establish feasibility. A small single-site trial can inform the design and power calculations for a larger multisite POC study. The latter would involve more than 1 but fewer than 6 sites.
5. POC for a new agent and scale validation cannot be carried out concurrently. POC trials can provide exploratory analysis, but not the comprehensive validation that is needed before including a scale in the phase 2, phase 2B, and phase 3 trials.
6. The analysis of subtypes and symptom clusters will require adaptations of existing designs. For example, enrichment strategies may be needed to increase the inclusion of patients with specific characteristics—e.g., suicidality, anger, anergy, chronicity, and treatment resistance.
7. If patients would not receive treatment if they were not in the study, then a placebo is an appropriate control.
8. If there is a best standard of treatment, then that should be used as the comparison. The use of an active comparator can vastly increase sample size requirements.[97]
9. "Treatment as usual" can also be used as the control. Treatment as usual varies across settings and populations, thus limiting the generalizability of study results.
10. The HAM-D should be used at least at intake and endpoint in all POC trials so the data can be included in meta-analyses. However, the MADRS and the QIDS are also highly sensitive to patients' symptoms and to symptomatic change and highly correlated and can enhance assessments.
11. Self-reports, including electronic versions, can be used to supplement clinical ratings, and the FDA is currently considering their use as primary outcomes for some disorders.
12. The use of currently available cognitive measurements, such as neurocognitive testing, in POC trials would be beneficial for future studies.
13. Patients' symptoms and side effects should be measured on a routine basis every visit.
14. Since onset of action might occur in days, rather than weeks or months, twice-weekly or more frequent evaluations should be performed and should include established rating scales as well as self-report scales. A combination of rating scales and self-report methods would provide measures of the components in addition to overall severity.
15. Self-report measures should be continuous (e.g., daily or more frequent) and repeated measures should be conducted over time, which would help with issues of reliability, attrition, and missing data.
16. Methods such as Interactive Voice Response System technology or the Depression and Bipolar Support Alliance (DBSA) mood-tracking online questionnaire could be used to track mood and measure onset of

Reprinted with correction to page 1525.

action. The DBSA scale might also be helpful for informing trial design.

17. Video evaluation methods provide a permanent visual record of patient interviews, thus eliminating the memory factor in judging severity of pathology at admission to the study. They can facilitate reliability testing of behavioral measures and diagnosis across clinical settings and allow for centralization of clinician ratings. Video evaluation should be considered in multisite POC studies.
18. Critical decision points should be used to make new treatment decisions, such as when to escalate the dose of a treatment, declare a treatment failure, and deal with partial improvement.
19. Biomarkers that can be used in POC trials should be adopted on the basis of how easy they are to use.
20. The federal government, academic centers, and industry should collaborate to create a network of sites to focus on POC studies and translational research in the treatment of MDD. Such a network should share data across POC studies and identify moderator and mediator variables related to antidepressant and placebo response.
21. DNA from all clinical studies should be banked now for future use in genomic analysis. Ideally, DNA would be collected from all willing patients in clinical trials for all studies (including industry) and banked centrally.
22. Cerebrospinal fluid (CSF) samples also represent a valuable pooled resource. Again, the federal government, industry, and academia should collaborate to create a data bank of CSF samples and integrate them with newer methodologies, such as PET scanning of receptors and transporters from the same neurotransmitter systems and relevant functional gene variants. Future imaging methods will extend to indices of second messenger systems and trophic effects on the brain that have been identified with antidepressant action.[88–100] Targeted CSF studies might be very helpful in POC trials to identify elements associated with treatment responses. Promising methods for CSF include proteomic and metabolic analyses.[101–103]
23. Power calculations should consider the sampling, measurement, design, and outcome decisions actually made in the study, not generically. A critical component of sample size determination is the value of the effect size below which clinicians are unlikely to prefer treatment over control. Until standards are established, a trial should have sufficient statistical power to detect a moderate effect size (d = 50). Most studies require no fewer than about 65 patients per treatment group, a few more in a well-designed multisite trial. Conceivably, in highly targeted studies, fewer patients per treatment group would be needed.

Improving Study Conduct

1. Site considerations
   a. Companies should develop a common tool for site evaluation for POC studies and enable the creation of a common data pool of quality POC sites.
   b. The criteria for high-quality sites for POC studies in depression will not be the same as criteria for "high-throughput" sites for phase 3 studies. These distinct criteria need to be spelled out.
   c. Sites should demonstrate the ability to discriminate active treatment from placebo.
   d. Formalized training systems for achieving standards and competency of raters across POC study sites should be developed, validated, and applied.
   e. Requirements for sites should be standardized. While academic sites will likely be preferred for studies using new technology or tools for surrogate markers, they should be held to quality and performance (speed) metrics consistent with good clinical practice and good business practice.
2. The growing placebo response in RCTs demands more sensitive behavioral methods and ethical solutions, the creative use of new technologies, and practical methods to "blind" raters to the expectations of the study, phase of study, and treatment assignments.
3. Since there are as yet no biological markers for mental disorders, behavioral evaluation is the only way to determine what the drug is doing, so reliable and valid assessments are very important. Sites need to be compensated for the additional costs related to supplemental assessments involved in POC studies. While adding methodology increases costs, when the reliability of the rating procedure is improved, sample size can be reduced and costs reduced.
4. The relevant population for a POC study should be the one upon which clinicians are called to treat, except for those unwilling or unable to participate, those who might be harmed by one of the treatments, or those who previous moderator analyses indicate would not be benefited by inclusion.

Improving Data Analysis

1. A study needs to be analyzed as it was designed, in terms of stratification, matching, use of covariates, and multisite designs. In a multisite trial, the analysis should be done centrally and must include site and site-by-treatment interaction. One of the major reasons for doing a multisite trial is to test the generalizability of the treatment effect, so it must be included in the analysis.
2. Stratification or matching often complicates the execution of the trial. If ignored, it may increase Type I error. If dealt with properly, it may necessitate a large sample size to control Type II error. Stratification and matching should not be performed unless there is a

Reprinted with correction to page 1525.

strong indication from prior trials that it is necessary, and the design should be kept as simple as possible to answer the primary research question.
3. The analysis should be performed before the blinding is removed, so that the interpretation of the results is not conditioned by what the researchers want to say about their data. Only after final clinical conclusions are drawn should blinds be lifted.
4. "Intention to treat" analysis should include all patients randomized in the study, even if they do not appear for the first treatment. With this in mind, it would be best to delay randomization to the last possible moment, in order to minimize the "no-shows." The FDA generally uses a modified intent-to-treat analysis that includes all randomized subjects who received at least 1 dose of assigned treatment and who had a baseline and at least 1 follow-up assessment.
5. When interpreting the results, the only p value that matters is the one for the primary outcome. Other results are publishable, but naked p values should not be given. Area under the curve, number needed to treat, and success rate differences are acceptable effect sizes.
6. Even if the p value is greater than or equal to .05, which means that the study is inadequate to find a difference between drug and placebo, investigators should continue analyzing the results in multiple different ways, not to change the conclusions but to improve the design of future studies. Reasons that an effect may not have shown include inadequate sample size, unreliable measures, and too much or not enough stratification.
7. After each RCT, moderators should be assessed. In future studies, effect size can be increased by targeting the intervention to patients who have the best potential of responding to it.
8. Although much of the effort to identify possible treatment moderators will be exploratory, when a possible moderator is identified, it must be formally tested prospectively with enough subjects included in the study to provide sufficient power for the analysis.
9. Site differences should be examined to determine if some characteristics affect outcome. An example might be involving family members in treatment.
10. The FDA has promoted last-observation-carried-forward (LOCF) methodology for years, but various repeated-measures models typically provide less biased estimates of the treatment effect. The FDA will accept mixed-effects regression models as an alternative to LOCF for the primary analysis model. However, investigators need to seek advice from the FDA on the analytical plans: model justification (e.g., the acceptable level of missing data), sensitivity analyses, interim analyses, and adjustment for multiple comparisons. Since there will always be missing data, it is important to understand the reasons for the missing data.[104]
11. Reasons for discontinuation of treatment should be clearly defined: is the cause of discontinuation a particular side effect, a group of side effects, or simply that the treatment is not working? Patients may discontinue in large part because they are not feeling significantly better, rather than because of an intolerable side effect.
12. Conclusions should not generalize beyond the limits of the sample studied. If 90% of a population is excluded from a study, the results apply only to the 10% who were included in the trial.

Future Research

1. More research is needed to define subtypes within the depressive spectrum and discrete symptomatic elements that may more accurately reflect the underlying biology.
2. While validated assessment instruments, such as the Symptom Checklist-90 or its brief form and selected mood scales as well as the IDS and QIDS, have been in use for decades and are applicable in new studies, novel assessment instruments are sorely needed.
3. Although currently used scales cover a wide range of depression symptoms, other symptoms are left out—e.g., anger, hostility, agitation, stress coping, and perceived stress, as well as the cognitive aspects of depression, including attention, concentration, decision making, and mood regulation. Scales for specific measurements could expand horizons.
4. Neuroimaging and genomic analysis offer great hope for biomarker development to subtype depression and predict treatment response. Critical for research in this area is to avoid following research paradigms of the past that sought to link a biological (e.g., neurochemical) variable to the overall diagnosis of depression. Biomarkers should be identified in relationship to major affective, cognitive, and behavioral components of depressive disorder and in assessing state changes related to treatment in POC studies.
5. Biomarkers may eventually identify patient subgroups for preferential treatment response and side effects.

Issues for Further Discussion

1. If trials use symptomatic volunteers (self-declared patients), are the results applicable to the patient population being treated in real clinical practice settings? Is the use of symptomatic volunteers an inclusion criteria issue or a site issue?
2. Should bridging studies be used going into phase 1 with patients rather than with nonpatients? Bridging studies have been routinely done in studies of cognitive enhancing drugs in Alzheimer disease and of

Reprinted with correction to page 1525.

antipsychotic drugs in schizophrenia. Drugs act differently in nonpatients and in patients. A downside of bridging studies in phase 1 is the increased cost of conducting studies in patients (versus nonpatients). The benefits of perhaps picking up an early signal of efficacy need to be weighed against the increased costs of the study.

3. Two other issues have not been well defined in antidepressant clinical trials:
   a. What does "treatment as usual" mean as a control condition rather than placebo for patients previously nonresponsive to antidepressants? Treatment as usual may differ significantly between sites in ways that may not be well documented.
   b. Nonresponsive patients make up two-thirds of the patient population. Previous treatment nonresponse might be a moderator variable, but it is unclear how prior history of nonresponse should be factored into placebo-controlled, and active treatment–controlled, antidepressant clinical trials; perhaps stratification could be used.

   These issues of "treatment as usual" and previously nonresponsive patients as a distinct moderator in antidepressant trials need to be addressed in future studies.

CONCLUSION

The primary endpoints in clinical trials have not changed substantially over the past 40 years. There are no generally accepted surrogate markers of disease or treatment response, and there is no consensus on the moderating or mediating factors that might affect response to active drug or placebo. A high percentage of antidepressant clinical trials have been negative due to a high placebo response rate.[15] The situation is approaching an impasse, and, in an era in which relatively safe and inexpensive SSRIs have become the standard for comparison, one can imagine that fewer companies will be willing to invest in the development of new drugs until there are more robust predictors of clinical efficacy. There are unused validated trial methods available to complement established ones, and new methods are currently being developed. Until a new model for clinical trials is agreed upon, though, new drugs that might be more efficacious for a subset of MDD patients could be overlooked.

Bottlenecks in the overall drug development process highlighted by the FDA include a limited number of biomarkers and surrogate endpoints for predicting clinical efficacy in animal models and early-stage clinical testing. This is certainly true in the antidepressant area. At this time, the primary point of leverage to advance antidepressant drug development is to enhance the quality of information in early-stage POC studies to amplify the signal of drug efficacy and better control for the impact of the placebo effect. Progress in this area might suggest more relevant animal models for screening antidepressant drugs, as well as improved methodology for pivotal phase 2 and phase 3 trials.

The 45 participants from academia, the NIH, patient advocacy, the FDA, and the pharmaceutical industry at our consensus development conference agreed on the need to improve the state of the art in pivotal clinical trials through focused POC studies of novel, putative antidepressant drugs. In terms of study design, the conferees recommended small, double-blind trials involving 2 to 6 sites that test a single hypothesis. Outcome should be measured using HAM-D at intake and endpoint to allow data to be included in future meta-analyses, but the use of other scales and methods for documenting patients' progress should be incorporated as well to enhance assessments. Newer approaches to achieving a common standard of clinical assessment include video evaluations and frequent patient self-report via Interactive Voice Response or other electronic technology.

The federal government, academic centers, and industry need to collaborate in the creation of a network of sites that can focus on POC studies and translational research on MDD. Such a network should optimally enable the sharing of data across POC studies and the identification of robust moderator and mediator variables related to antidepressant and placebo response. Drug study sites should be evaluated using a common tool relevant to POC studies. Quality metrics should be emphasized over throughput, but academic and commercial sites should be held to equivalent standards on issues of recruitment rates, quality of clinical assessment, and the ability to differentiate active treatment from placebo. Academic sites should have the advantage of access to state-of-the-art technology, but consortia of academic and commercial sites might be developed to help meet recruitment and scientific targets.

The data analytic strategy should be chosen such that it corresponds to the study design. Stratification and matching should be kept simple and only used if necessary. Analysis should be performed before blinding is removed. Even if a study does not find a significant difference between drug and placebo, exploratory analyses could provide information to improve the design and conduct of future studies. Moderators should be assessed after each RCT, so that future studies can target the intervention to patients who are most likely to respond to it.

With the advent of new technologies in brain imaging and new tools for assessing CSF, this is the time in which detailed analysis and measurement of the major affective, cognitive, and behavioral components of the depressive disorder should be studied in relationship to potential biomarkers in small POC studies. Biomarkers should be incorporated into POC trials on the basis of their promising theoretical interest and eventual applicability in later

Reprinted with correction to page 1525.

pivotal trials. DNA from all POC (and other) antidepressant clinical trials should be banked (preferably at a central repository or as a shared resource) for future genomic analysis to improve genetic predictors of drug response and side effect profile. Cerebrospinal fluid samples also represent a valuable pooled resource for proteomic and metabolic analyses that might clarify factors that mediate treatment response.

Neuroimaging represents a promising area for biomarker development of disease (disorder) state, depression severity, subtyping of the disorder, disaggregation of symptoms, and monitoring of global and specific symptomatic response to treatment. For research in this area to succeed, it is critical that we avoid the mistakes of the past that sought to find a simple biochemical explanation of major depression (e.g., relative depletion of monoamines norepinephrine and serotonin) and focus instead on uncovering basic behavioral and biological components of the depressive disorders. Novel assessment instruments should be developed to differentiate subtypes and symptom clusters within MDD.

While a number of issues remain about how best to design and conduct POC antidepressant trials, the overview of the state of the field and the consensus recommendations offered by the conference suggest possible short-term actions and longer-term strategies to advance the development of more efficacious drugs to treat MDD.

Drug names: bupropion (Aplenzin, Wellbutrin, and others), citalopram (Celexa and others), desipramine (Norpramin and others), duloxetine (Cymbalta), escitalopram (Lexapro and others), norepinephrine (Levophed and others), selegiline (EMSAM, Eldepryl, and others).

Financial disclosure: **Dr. Gelenberg** is a consultant for Eli Lilly, Pfizer, Best Practice Project Management, AstraZeneca, Wyeth, Cyberonics, Novartis, Forest, GlaxoSmithKline, ZARS Pharma, Jazz, Lundbeck, and Takeda; has received research grant funding from Eli Lilly; is a member of the speakers bureaus for Pfizer, GlaxoSmithKline, and Wyeth; and is a major stock shareholder of Healthcare Technology Systems. **Dr. Thase** is a consultant for AstraZeneca, Bristol-Myers Squibb, Cephalon, Cyberonics, Eli Lilly, Forest, GlaxoSmithKline, Janssen, MedAvante, Neuronetics, Novartis, Organon, Sepracor, Shire, Supernus, and Wyeth; has received grant/research support from Eli Lilly and Sepracor; is a member of the speakers bureaus for AstraZeneca, Bristol-Myers Squibb, Cyberonics, Eli Lilly, GlaxoSmithKline, Organon, sanofi-aventis, and Wyeth; has given expert testimony for Jones Day (Wyeth litigation), Phillips Lytle (GlaxoSmithKline litigation), and Pepper Hamilton LLP (Eli Lilly litigation); has equity holdings (exclude mutual funds/blinded trusts) in MedAvante; and has received royalties from American Psychiatric Publishing, Inc., Guilford Publications, Herald House, and W.W. Norton & Company. His spouse/partner is a senior medical director at Advogent (formerly Cardinal Health). **Dr. Meyer** is an employee and part owner of Best Practice Project Management. In the past 12 months, Best Practice has received research/grant support from Bristol-Myers Squibb, MedRespond, Forest, and sanofi-aventis and has received support for consulting activities from EnVivo, Ovation, Neurim, Janssen, and Takeda. Dr. Meyer has been involved in each of these programs. Dr. Meyer has also served as a consultant for the National Institute on Drug Abuse and has (or will) receive honoraria from the Department of Psychiatry, Northwestern University. **Dr. Goodwin** is a consultant for GlaxoSmithKline, Eli Lilly, Pfizer, Wyeth, and AstraZeneca; has received honoraria from GlaxoSmithKline; and is a member of the speakers/advisory boards for GlaxoSmithKline and Eli Lilly. **Dr. Potter** is an employee of Merck. **Dr. Shelton** has received grant/research support from Eli Lilly, GlaxoSmithKline, Janssen, Pfizer, Sanofi, Wyeth-Ayerst, AstraZeneca, and Abbott; is a consultant for Pfizer and Janssen; and is on the speakers bureaus for Bristol-Myers Squibb, Eli Lilly, Janssen, Pfizer, GlaxoSmithKline, Solvay, Wyeth-Ayerst, and Abbott. **Dr. Fava** has received grant/research support from Abbott, Alkermes, Aspect Medical Systems, AstraZeneca, Bristol-Myers Squibb, Cephalon, Eli Lilly, Forest, GlaxoSmithKline, Johnson & Johnson, Lichtwer Pharma GmbH, Lorex, Novartis, Organon, PamLab, Pfizer, Pharmavite, Roche, sanofi-aventis, Solvay, Synthelabo, and Wyeth-Ayerst; is a consultant for Amarin, Aspect Medical Systems, AstraZeneca, Bayer AG, Best Practice Project Management, Biovail, BrainCells, Bristol-Myers Squibb, Cephalon, CNS Response, Compellis, Cypress, Dov, Eli Lilly, EPIX, Fabre-Kramer, Forest, GlaxoSmithKline, Grunenthal GmbH, Janssen, Jazz, Johnson & Johnson, Knoll, Lundbeck, MedAvante, Merck, Neuronetics, Novartis, Nutrition 21, Organon, PamLab, Pfizer, PharmaStar, Pharmavite, Precision Human Biolaboratory, Roche, sanofi-aventis, Sepracor, Solvay, Somaxon, Somerset, Synthelabo, Takeda, Tetragenex, Transcept, Vanda, and Wyeth-Ayerst; is a member of the speakers bureaus for AstraZeneca, Boehringer Ingelheim, Bristol-Myers Squibb, Eli Lilly, Forest, GlaxoSmithKline, Novartis, Organon, Pfizer, PharmaStar, Primedia, Reed-Elsevier, and Wyeth-Ayerst; has equity holdings in Compellis and MedAvante; has submitted patent applications for SPCD and for a combination of azapirones and bupropion in major depressive disorder; and has received copyright royalties for the Massachusetts General Hospital Cognitive and Physical Functioning Questionnaire, Discontinuation Emergent Signs and Symptoms scale, and SAFER criteria. In the last 2 years, **Dr. Trivedi** has received grant/research support from the Agency for Healthcare Research and Quality, the National Institute of Mental Health, the National Institute on Drug Abuse, and Targacept and has been a consultant for AstraZeneca, Bristol-Myers Squibb, Cephalon, Fabre-Kramer, Forest, GlaxoSmithKline, Janssen, Johnson & Johnson, Eli Lilly, Neuronetics, Pfizer, and Wyeth-Ayerst. **Dr. Ninan** is an employee of Wyeth. **Dr. Mann** is a consultant for Wyeth and sanofi-aventis and has received grant/research support from GlaxoSmithKline and Novartis. **Ms. Bergeson** is an employee of the Depression and Bipolar Support Alliance. **Dr. Kocsis** is a consultant for Wyeth; has received grant/research support from National Institute of Mental Health, Wyeth, AstraZeneca, Novartis, National Institute on Drug Abuse, Forest, and sanofi-aventis; and is a member of the speakers/advisory boards for Wyeth and AstraZeneca. **Dr. Leon** is a consultant for the U.S. Food and Drug Administration and the National Institute of Mental Health and is a member of the Data and Safety Monitoring Boards for Pfizer and Organon. **Dr. Rosenbaum** is a member of the advisory boards for Forest, Eli Lilly, MedAvante, Novartis, Sanofi, and Wyeth; is a consultant for Neuronetics, Organon, Somaxon, and Supernus; is a member of the speakers bureau for Boehringer Ingelheim Italia; is a member of the speakers bureau and involved with educational content development for Primedia Health Care; and has equity holdings in Compellis, MedAvante, and Somaxon. **Drs. Katz**, **Kraemer**, **Khan**, **Endicott**, and **Manji** report no additional financial or other relationship relevant to the subject of this article.

Disclaimers: Neither **Dr. Gelenberg**, the journal's editor in chief, nor **Drs. Thase**, **Meyer**, **Fava**, **Kocsis**, **Leon**, and **Rosenbaum**, editorial board members, were involved in the editorial review or decision to publish this article.

Conference participants: In addition to the authors of this article, the following individuals participated in the conference "Advancing Signal Strength in Proof-of-Concept Studies in Major Depression," June 21–22, 2007: *Government and academia*: **Linda L. Carpenter, M.D.**, Butler Hospital/Brown University; **Ian A. Cook, M.D.**, Semel Institute for Neuroscience and Human Behavior at University of California, Los Angeles; **Wayne C. Drevets, M.D.**, National Institute of Mental Health; **Ronald S. Duman, Ph.D.**, Yale University School of Medicine; **Joel B. Greenhouse, Ph.D.**, Carnegie Mellon University; **John K. Hsiao, M.D.**, National Institute of Mental Health; **Anthony J. Rothschild, M.D.**, University of Massachusetts Medical School;

Reprinted with correction to page 1525.

Holly A. Swartz, M.D., Western Psychiatric Institute and Clinic; **Howard Tennen, Ph.D.**, University of Connecticut Health Center; **Philip Sung-En Wang, M.D., Dr.P.H.**, National Institute of Mental Health; **Thomas P. Laughren, M.D.**, Food and Drug Administration. *Industry:* **Kathryn M. Connor, M.D., M.H.S.**, Merck Research Laboratories, Merck & Co., Inc.; **Aroon Datta, M.Sc.**, Forest Research Institute; **Judith Dunn, Ph.D.**, Sepracor; **Judith Jaeger, Ph.D.**, AstraZeneca; **Ronald Marcus, M.D.**, Bristol-Myers Squibb Company; **Randall L. Morrison, Ph.D.**, Ortho-McNeil Janssen Scientific Affairs, L.L.C., Johnson & Johnson; **Jorge A. Quiroz, M.D.**, Johnson & Johnson; **Tanya Ramey, M.D., Ph.D.**, Pfizer Inc.; **Brigitte A. Robertson, M.D.**, Sepracor; **Miqun Robinson, M.D., Ph.D.**, Sanofi-Aventis; **Sharon Rosenzweig-Lipson, Ph.D.**, Wyeth Research; **Mark A. Smith, M.D., Ph.D.**, AstraZeneca; **Gary Tong, M.D., Ph.D.**, Bristol-Myers Squibb Company.

REFERENCES

1. RTI International-University of North Carolina Evidence-Based Practice Center. Comparative Effectiveness of Second-Generation Antidepressants in the Pharmacologic Treatment of Adult Depression. Rockville, Md: Agency for Healthcare Research and Quality; 2007. Publication 07-EHC007-EF
2. Mann JJ. The medical management of depression. N Engl J Med 2005; 353:1819–1834
3. Montgomery SA, Baldwin DS, Blier P, et al. Which antidepressants have demonstrated superior efficacy? a review of the evidence. Int Clin Psychopharmacol 2007 Nov;22(6):323–329
4. Turner EH, Matthews AM, Linardatos E, et al. Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med 2008;358:252–260
5. Nemeroff CB, Entsuah R, Benattia I, et al. Comprehensive analysis of remission (COMPARE) with venlafaxine versus SSRIs. Biol Psychiatry 2008;63:424–434
6. Papakostas GI, Thase ME, Fava M, et al. Are antidepressant drugs that combine serotonergic and noradrenergic mechanisms of action more effective than the selective serotonin reuptake inhibitors in treating major depressive disorder? a meta-analysis of studies of newer agents. Biol Psychiatry 2007 Dec;62(11):1217–1227
7. Thase ME, Haight BR, Richard N, et al. Remission rates following antidepressant therapy with bupropion or selective serotonin reuptake inhibitors: a meta-analysis of original data from 7 randomized controlled trials. J Clin Psychiatry 2005 Aug;66(8):974–981
8. Thase ME, Pritchett YL, Ossanna MJ, et al. Efficacy of duloxetine and selective serotonin reuptake inhibitors: comparisons as assessed by remission rates in patients with major depressive disorder. J Clin Psychopharmacol 2007;27:672–676
9. Katz MM, Tekell JL, Bowden CL, et al. Onset and early behavioral effects of pharmacologically different antidepressants and placebo in depression. Neuropsychopharmacology 2004;29:566–579
10. Kraemer HC, Wilson GT, Fairburn CG, et al. Mediators and moderators of treatment effects in randomized clinical trials. Arch Gen Psychiatry 2002;59:877–883
11. Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and trouble with averages. Milbank Q 2004;82: 661–687
12. Thase ME. Comparing the methods used to compare antidepressants. Psychopharmacol Bull 2002;36(suppl 1):1–17
13. US Health and Human Services. Challenge and opportunity on the critical path to new medical products. March 2004. US Food and Drug Administration Web site. Available at: http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html. Accessed Jan 31, 2008
14. US Department of Health and Human Services. Critical Path Opportunities Report. March 2006. US Food and Drug Administration Web site. Available at: http://www.fda.gov/ohrms/dockets/ac/06/briefing/2006-4241B1-02-25-FDA-Critical%20Path%20report.pdf. Accessed Jan 31, 2008
15. Malberg JE, Platt B, Lucki I, et al. Increasing the levels of insulin-like growth factor-1 by an IGF binding protein inhibitor produces anxiolytic and antidepressant-like effects. Neuropsychopharmacology 2007;32: 2360–2368
16. Rajarao SJR, Platt B, Sukoff SJ, et al. Anxiolytic-like activity of the nonselective galanin receptor agonist, galnon. Neuropeptides 2007;41:307–320
17. Ring RH, Malberg JE, Potestio L, et al. Anxiolytic-like activity of oxytocin in male mice: behavioral and autonomic evidence, therapeutic implications. Psychopharmacology 2006;185:218–225
18. Pietraszek M, Sukhanov I, Maciejak P, et al. Anxiolytic-like effects of mGlu1 and mGlu5 receptor antagonists in rats. Eur J Pharmacol 2005; 514:25–34
19. Hodgson RA, Higgins GA, Guthrie DH, et al. Comparison of the V1b antagonist, SSR149415, and the CRF1 antagonist, CP-154,526, in rodent models of anxiety and depression. Pharmacol Biochem Behav 2007 Mar;86(3):431–440
20. Pilc A, Klodzinska A, Branski P, et al. Multiple MPEP administrations evoke anxiolytic- and antidepressant-like effects in rats. Neuropharmacology 2002;43:181–187
21. Walsh BT, Seidman SN, Sysko R, et al. Placebo response in studies of major depression: variable, substantial, and growing. JAMA 2002;287: 1840–1847
22. Manji HK, Drevets WC, Charney DS. The cellular neurobiology of depression. Nat Med 2001;7:541–547
23. Khan A, Leventhal RM, Khan SR, et al. Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. J Clin Psychopharmacol 2002;22:40–45
24. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;23:56–62
25. Bagby RM, Ryder AG, Schuller DR, et al. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? Am J Psychiatry 2004;161:2163–2177
26. Trivedi MH, Rush AJ, Ibrahim HM, et al. The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self-Report (QIDS-SR) in public sector patients with mood disorders: a psychometric evaluation. Psychol Med 2004;34:73–82
27. Rush AJ, Bernstein IH, Trivedi MH, et al. An evaluation of the Quick Inventory of Depressive Symptomatology and the Hamilton Rating Scale for Depression: a Sequenced Treatment Alternatives to Relieve Depression trial report. Biol Psychiatry 2006;59:493–501
28. Bech P. Rating scales in depression: limitations and pitfalls. Dialogues Clin Neurosci 2006;8:207–215
29. Raskin A, Schulterbrandt JG, Reatig N, et al. Depression subtypes and response to phenelzine, diazepam, and a placebo: results of a nine hospital collaborative study. Arch Gen Psychiatry 1974;30:66–75
30. Freud S. Mourning and melancholia. In: Collected Papers, vol 4. New York, NY: Basic Books; 1959
31. Abraham K. Selected Works of Karl Abraham. London, England: Hogarth Press; 1949
32. Raskin A, Schulterbrandt JG, Reatig N, et al. Differential response to chlorpromazine, imipramine, and placebo: a study of subgroups of hospitalized depressed patients. Arch Gen Psychiatry 1970;23:164–173
33. Katz MM, Bowden CL, Berman N, et al. Resolving the onset of antidepressants' clinical actions. J Clin Psychopharmacol 2006;26:549–553
34. Katz MM, Koslow SH, Maas JW, et al. Identifying the specific clinical actions of amitriptyline: interrelationships of behaviour, affect, and plasma levels in depression. Psychol Med 1991;21:599–611
35. Hamilton M. Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 1967;6:278–296
36. Guy W. ECDEU Assessment Manual for Psychopharmacology, revised. US Dept Health, Education, and Welfare publication (ADM) 76-338. Washington, DC: National Institute of Mental Health; 1976:218–222
37. Miller IW, Bishop S, Norman WH, et al. The Modified Hamilton Rating Scale for Depression: reliability and validity. Psychiatry Res 1985;14: 131–142
38. Williams JB. A structured interview guide for the Hamilton Depression Rating Scale. Arch Gen Psychiatry 1988;45:742–747
39. Williams JB. Structured interview guides for the Hamilton Rating Scales. Psychopharmacol Ser 1990;9:48–63
40. Thase ME, Frank E, Mallinger AG, et al. Treatment of imipramine-resistant recurrent depression, 3: efficacy of monoamine oxidase inhibitors. J Clin Psychiatry 1992 Jan;53(1):5–11
41. Thase ME, Mallinger AG, McKnight D, et al. Treatment of imipramine-resistant recurrent depression, 4: a double-blind crossover study of tranylcypromine for anergic bipolar depression. Am J Psychiatry 1992;149:195–198
42. Hooijer C, Zitman FG, Griez E, et al. The Hamilton Depression Rating

Reprinted with correction to page 1525.

Scale (HDRS): changes in scores as a function of training and version used. J Affect Disord 1991;22:21–29
43. Bech P. Rating scales for affective disorders: their validity and consistency. Acta Psychiatr Scand Suppl 1981;295:1–101
44. Bech P. Meta-analysis of placebo-controlled trials with mirtazapine using the core items of the Hamilton Depression Scale as evidence of a pure antidepressive effect in the short-term treatment of major depression. Int J Neuropsychopharmacol 2001;4:337–345
45. Gibbons RD, Clark DC, Kupfer DJ. Exactly what does the Hamilton Depression Rating Scale measure? J Psychiatr Res 1993;27:259–273
46. Maier W. The Hamilton Depression Scale and its alternatives: a comparison of their reliability and validity. Psychopharmacol Ser 1990;9:64–71
47. Leon AC, Marzuk PM, Portera L. More reliable outcome measures can reduce sample size requirements. Arch Gen Psychiatry 1995;52:867–871
48. Beck AT, Ward CH, Mendelson M, et al. An inventory for measuring depression. Arch Gen Psychiatry 1961;4:561–571
49. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979;134:382–389
50. Rush AJ, Roffwarg HP, Giles DE, et al. Psychobiological predictors of antidepressant drug response. Pharmacopsychiatria 1983;16:192–194
51. Bech P, Engelhardt N, Evans KR, et al. Why the Hamilton Depression Rating Scale endures. Am J Psychiatry 2005 Dec;162(12):2396
52. Zimmerman M, McGlinchey JB, Posternak MA, et al. How should remission from depression be defined? the depressed patient's perspective. Am J Psychiatry 2006 Jan;163(1):148–150
53. Zimmerman M, McGlinchey JB, Posternak MA, et al. Remission in depressed outpatients: more than just symptom resolution? J Psychiatr Res 2008 Aug;42(10):797–801
54. Kraemer HC, Robinson TN. Are certain multicenter randomized clinical trial structures misleading clinical and policy decisions? Contemp Clin Trials 2005;26:518–529
55. Khan A, Detke M, Khan SR, et al. Placebo response and antidepressant clinical trial outcome. J Nerv Ment Dis 2003;191:211–218
56. Trivedi MH, Rush H. Does a placebo run-in or a placebo treatment cell affect the efficacy of antidepressant medications? Neuropsychopharmacology 1994;11:33–43
57. Stewart JW, Mercier MA, Agosti V, et al. Imipramine is effective after unsuccessful cognitive therapy: sequential use of cognitive therapy and imipramine in depressed outpatients. J Clin Psychopharmacol 1993 Apr;13(2):114–119
58. Frank E, Kupfer DJ, Buysse DJ, et al. Randomized trial of weekly, twice-monthly, and monthly interpersonal psychotherapy as maintenance treatment for women with recurrent depression. Am J Psychiatry 2007;164:761–767
59. Khan A, Schwartz K, Kolts RL, et al. Relationship between depression severity entry criteria and antidepressant clinical trial outcomes. Biol Psychiatry 2007;62:65–71
60. Quitkin FM, Rabkin JG, Ross D, et al. Identification of true drug response to antidepressants: use of pattern analysis. Arch Gen Psychiatry 1984;41:782–786
61. Stassen HH, Delini-Stula A, Angst J. Time course of improvement under antidepressant treatment: a survival-analytical approach. Eur Neuropsychopharmacol 1993;3:127–135
62. Katz MM, Koslow SH, Maas JW, et al. The timing, specificity and clinical prediction of tricyclic drug effects in depression. Psychol Med 1987;17:297–309
63. Taylor MJ, Freemantle N, Geddes JR, et al. Early onset of selective serotonin reuptake inhibitor antidepressant action: systematic review and meta-analysis. Arch Gen Psychiatry 2006;63:1217–1223
64. Papakostas GI, Perlis RH, Scalia MJ, et al. A meta-analysis of early sustained response rates between antidepressants and placebo for the treatment of major depressive disorder. J Clin Psychopharmacol 2006; 26:56–60
65. Posternak MA, Zimmerman M. Is there a delay in the antidepressant effect? a meta-analysis. J Clin Psychiatry 2005 Feb;66(2):148–158
66. Coryell W, Coppen A, Zeigler VE, et al. Early improvement as a predictor of response to amitriptyline and nortriptyline: a comparison of two patient samples. Psychol Med 1982;12:135–139
67. Katz MM, Houston JP, Brannan S, et al. A multivantaged behavioural method for measuring onset and sequence of the clinical actions of antidepressants. Int J Neuropsychopharmacol 2004;7:471–479
68. Stassen HH, Angst J, Delini-Stula A. Delayed onset of action of antidepressant drugs: survey of results of Zurich meta-analyses. Pharmacopsychiatry 1996;29:87–96
69. Trivedi MH, Morris DW, Grannemann BD, et al. Symptom clusters as predictors of late response to antidepressant treatment. J Clin Psychiatry 2005 Aug;66(8):1064–1070
70. Nierenberg AA, McLean NE, Alpert JE, et al. Early nonresponse to fluoxetine as a predictor of poor 8-week outcome. Am J Psychiatry 1995;152:1500–1503
71. Trivedi MH, Daly EJ. Treatment strategies and tactics for treatment-resistant depression. US Psychiatry 2007;17–20
72. Trivedi MH, Rush AJ, Wisniewski SR, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. Am J Psychiatry 2006;163:28–40
73. Trivedi MH, Rush AJ, Gaynes BN, et al. Maximizing the adequacy of medication treatment in controlled trials and clinical practice: STAR*D measurement-based care. Neuropsychopharmacology 2007;32: 2479–2489
74. Weissman MM, Pilowsky DJ, Wickramaratne PJ, et al. Remissions in maternal depression and child psychopathology: a STAR*D-Child report. JAMA 2006;295:1389–1398
75. Delgado PL, Miller HL, Salomon RM, et al. Tryptophan-depletion challenge in depressed patients treated with desipramine or fluoxetine: implications for the role of serotonin in the mechanism of antidepressant action. Biol Psychiatry 1999;46:212–220
76. Nelson JC, Mazure CM, Jatlow PI, et al. Combining norepinephrine and serotonin reuptake inhibition mechanisms for treatment of depression: a double-blind, randomized study. Biol Psychiatry 2004;55:296–300
77. Quitkin FM, Stewart JW, McGrath PJ, et al. Columbia atypical depression: a subgroup of depressives with better response to MAOI than to tricyclic antidepressants or placebo. Br J Psychiatry Suppl 1993;21: 30–34
78. Iosifescu DV, Renshaw PF, Dougherty DD, et al. Major depressive disorder with anger attacks and subcortical MRI white matter hyperintensities. J Nerv Ment Dis 2007;195:175–178
79. Fava M, Anderson K, Rosenbaum JF. "Anger attacks": possible variants of panic and major depressive disorders. Am J Psychiatry 1990;147: 867–870
80. Fava M, Rosenbaum JF. Anger attacks in patients with depression. J Clin Psychiatry 1999;60(suppl 15):21–24
81. Fava M, Vuolo RD, Wright EC, et al. Fenfluramine challenge in unipolar depression with and without anger attacks. Psychiatry Res 2000;94:9–18
82. Fava M. Moderators of response to next-step strategies in resistant depression. Presented at the 45th annual meeting of the American College of Neuropsychopharmacology (ACNP); 2006; Hollywood, Fla
83. Fava M, McCall WV, Krystal A, et al. Eszopiclone co-administered with fluoxetine in patients with insomnia coexisting with major depressive disorder. Biol Psychiatry 2006;59:1052–1060
84. Rush AJ, Trivedi M, Fava M. Depression, 4: STAR*D treatment trial for depression. Am J Psychiatry 2003 Feb;160(2):237
85. Fava M, Rush AJ, Alpert JE, et al. Difference in treatment outcome in outpatients with anxious versus nonanxious depression: a STAR*D report. Am J Psychiatry 2008;165:342–351
86. Nemeroff CB, Heim CM, Thase ME, et al. Differential responses to psychotherapy versus pharmacotherapy in patients with chronic forms of major depression and childhood trauma. Proc Nat Acad Sci U S A 2003 Nov;100(24):14293–14296
87. Chen ZY, Jing D, Bath KG, et al. Genetic variant BDNF (Val66Met) polymorphism alters anxiety-related behavior. Science 2006;314: 140–143
88. Choi MJ, Kang RH, Lim SW, et al. Brain-derived neurotrophic factor gene polymorphism (Val66Met) and citalopram response in major depressive disorder. Brain Res 2006;1118:176–182
89. Tononi G, Cirelli C. Sleep function and synaptic homeostasis. Sleep Med Rev 2006;10:49–62
90. Quera Salva MA, Vanier B, et al. Major depressive disorder, sleep EEG and agomelatine: an open-label study. Int J Neuropsychopharmacol 2007;10:691–696
91. Hoffmann R, Hendrickse W, Rush AJ, et al. Slow-wave activity during non-REM sleep in men with schizophrenia and major depressive disorders. Psychiatry Res 2000;95:215–225
92. Vyazovskiy VV, Cirelli C, Pfister-Genskow M, et al. Molecular and electrophysiological evidence for net synaptic potentiation in wake and depression in sleep. Nat Neurosci 2008;11:200–208

Reprinted with correction to page 1525.

   © COPYRIGHT 2008 PHYSICIANS POSTGRADUATE PRESS, INC.

93. Huber R, Ghilardi MF, Massimini M, et al. Local sleep and learning. Nature 2004;430:78–81
94. Parsey RV, Olvet DM, Oquendo MA, et al. Higher 5-HT1A receptor binding potential during a major depressive episode predicts poor treatment response: preliminary data from a naturalistic study. Neuropsychopharmacology 2006;31:1745–1749
95. Ingelman-Sundberg M. Pharmacogenetics: an opportunity for a safer and more efficient pharmacotherapy. J Intern Med 2001;250: 186–200
96. McMahon FJ, Buervenich S, Charney D, et al. Variation in the gene encoding the serotonin 2A receptor is associated with outcome of antidepressant treatment. Am J Hum Genet 2006;78:804–814
97. Leon AC. Placebo protects subjects from nonresponse: a paradox of power. Arch Gen Psychiatry 2000;57:329–330
98. Manji HK, Quiroz JA, Sporn J, et al. Enhancing neuronal plasticity and cellular resilience to develop novel, improved therapeutics for difficult-to-treat depression. Biol Psychiatry 2003;53:707–742
99. Carlson PJ, Singh JB, Zarate CA Jr, et al. Neural circuitry and neuroplasticity in mood disorders: insights for novel therapeutic targets. NeuroRx 2006;3:22–41
100. Schloesser RJ, Huang J, Klein PS, et al. Cellular plasticity cascades in the pathophysiology and treatment of bipolar disorder. Neuropsychopharmacology 2008;33:110–133
101. Kaddurah-Daouk R, Kristal BS, Weinshilboum RM. Metabolomics: a global biochemical approach to drug response and disease. Annu Rev Pharmacol Toxicol 2008;48:653–683
102. Wilson KE, Ryan MM, Prime JE, et al. Functional genomics and proteomics: application in neurosciences. J Neurol Neurosurg Psychiatry 2004;75:529–538
103. Lan MJ, McLoughlin GA, Griffin JL, et al. Metabonomic analysis identifies molecular changes associated with the pathophysiology and drug treatment of bipolar disorder. Mol Psychiatry 2008 Feb 5 [epub ahead of print]
104. Leon AC, Demirtas H, Hedeker D. Bias reduction with an adjustment for participants' intent to drop out of a randomized controlled clinical trial. Clin Trials 2007;4:540–547
105. Economist Technol Q. March 13, 2004:37–38
106. US Food and Drug Administration. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes 2006;4:79
107. Burke LB, Kennedy DL, Miskala PH, et al. The use of patient-reported outcome measures in the evaluation of medical products for regulatory approval. Clin Pharmacol Ther 2008;84:281–283

Reprinted with correction to page 1525.