

Report by the ACNP Task Force on Response and Remission in Major Depressive Disorder

A John Rush*¹, Helena C Kraemer², Harold A Sackeim³, Maurizio Fava⁴, Madhukar H Trivedi¹, Ellen Frank⁵, Philip T Ninan⁶, Michael E Thase⁵, Alan J Gelenberg⁷, David J Kupfer⁵, Darrel A Regier⁸, Jerrold F Rosenbaum⁴, Oakley Ray^{9,10,11} and Alan F Schatzberg²

¹Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, TX, USA; ²Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, USA; ³New York State Psychiatric Institute and the Department of Psychiatry, College of Physicians and Surgeons of Columbia University, New York, NY, USA; ⁴Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; ⁵Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; ⁶Emory University School of Medicine, Atlanta, GA, USA; ⁷Department of Psychiatry, University of Arizona College of Medicine, Tucson, AZ, USA; ⁸American Psychiatric Institute for Research and Education and Division of Research at the American Psychiatric Association, Arlington, VA, USA; ⁹Department of Psychology, Vanderbilt University, Nashville, TN, USA; ¹⁰Department of Psychiatry, Vanderbilt University, Nashville, TN, USA; ¹¹Department of Pharmacology, Vanderbilt University, Nashville, TN, USA

This report summarizes recommendations from the ACNP Task Force on the conceptualization of remission and its implications for defining recovery, relapse, recurrence, and response for clinical investigators and practicing clinicians. Given the strong implications of remission for better function and a better prognosis, remission is a valid, clinically relevant end point for both practitioners and investigators. Not all depressed patients, however, will reach remission. Response is a less desirable primary outcome in trials because it depends highly on the initial (often single) baseline measure of symptom severity. It is recommended that remission be ascribed after 3 consecutive weeks during which minimal symptom status (absence of both sadness and reduced interest/pleasure along with the presence of fewer than three of the remaining seven DSM-IV-TR diagnostic criterion symptoms) is maintained. Once achieved, remission can only be lost if followed by a relapse. Recovery is ascribed after at least 4 months following the onset of remission, during which a relapse has not occurred. Recovery, once achieved, can only be lost if followed by a recurrence. Day-to-day functioning and quality of life are important secondary end points, but they were not included in the proposed definitions of response, remission, recovery, relapse, or recurrence. These recommendations suggest that symptom ratings that measure all nine criterion symptom domains to define a major depressive episode are preferred as they provide a more certain ascertainment of remission. These recommendations were based largely on logic, the need for internal consistency, and clinical experience owing to the lack of empirical evidence to test these concepts. Research to evaluate these recommendations empirically is needed.

Neuropsychopharmacology (2006) **31**, 1841–1853. doi:10.1038/sj.npp.1301131; published online 21 June 2006

Keywords: response; remission; recovery; relapse; recurrence; depression

INTRODUCTION

In 1989, the MacArthur Foundation sponsored a set of meetings with the goal of achieving consensus in defining outcomes for clinical studies of depression. The resulting report (Frank *et al*, 1991) articulated the key elements inherent in applying the concepts of response, remission, recovery, relapse, and recurrence to major depressive disorder (MDD). The report recommended that these

definitions be based on observable phenomena and include a temporal focus reflecting symptom change over the patient's lifetime. The proposed definitions do not imply a specific cause of symptom change, as symptom change can be due to specific treatment effects, the natural waxing and waning of depressive symptoms, or nonspecific effects of treatment.

Specifically, Frank *et al* recommended that *response* refer to a clinically significant degree of depressive symptom reduction following treatment initiation. When used clinically, response implies that the treatment has caused the response. Responders traditionally include both patients with clinically significant reduction in depressive symptoms, whether or not remission has been ascertained. *Remission* referred to the virtual absence of depressive symptoms. The period of remission may end with either

*Correspondence: Dr AJ Rush, Department of Psychiatry, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75390-9086, USA, Tel: +1 214 648 4601, Fax: +1 214 648 4612, E-mail: john.rush@utsouthwestern.edu
Received 5 May 2005; revised 9 May 2006; accepted 15 May 2006
Online publication: 23 May 2006 at <http://www.acnp.org/citations/Npp052306050297/default.pdf>

relapse or recovery. *Relapse* was viewed as a return of the index major depressive episode (MDE) following the onset of remission but before fulfilling the criteria for recovery. *Recovery* was ascribed when the period of remission had been sufficiently sustained that continued well being may be expected (with or without continuing treatment). A patient was considered recovered from the MDE (but not necessarily recovered from the illness) not simply when syndromal episode criteria were no longer met, but rather when a remitted state had been achieved for a sufficient period of time that a subsequent MDE was viewed as the onset of a new MDE—not simply the reappearance of the index episode. (Another concept of recovery is the successful integration of a mental disorder into the consumer's life and involves rebuilding meaningful lives, hope and optimism, self-empowerment, effective collaboration and direction in clinical care decisions, and decreasing dependence on the mental health system. This paradigm has been developed with increasing participation by recipients of mental health services and has been put forward by the President's New Freedom Commission on Mental Health (2003). The implications for delivery of mental health services are important. In the present report, the concept of recovery is framed within the medical model and relates more narrowly to clinical assessment relevant to a physician's therapeutic decisions and to meaningful clinical assessment in the context of clinical trials.) *Recurrence* referred to the development of a new MDE following the recovery.

Relapse, recovery, and recurrence included not only the level of symptomatic severity but also required that certain criteria be met over defined time intervals. For example, the patient may begin in the well state. After some time, a sufficient number of depressive symptoms are present for a sufficient duration (≥ 2 weeks by DSM-IV-TR) that the patient can be said to have entered an MDE. The patient is in an episode until it ends with remission (another time point), which is defined by a maximal number of symptoms for a sufficient time period. Remission may lead to recovery (another time point) or a relapse. Recovery may persist or be followed by a new episode (a recurrence).

The Task Force was formed to evaluate the available empirical evidence aimed at testing the concepts and to make recommendations for revising the definitions/concepts if called for.

THE INCREASING IMPORTANCE OF REMISSION

Clinical Implications of Remission

The importance of remission rests on evidence that following acute treatment trials remitters, as compared to those who have responded but who have some residual depressive symptoms, have better function (Riso *et al*, 1997; Miller *et al*, 1998; Hirschfeld *et al*, 2002), a better prognosis (Thase *et al*, 1992; Paykel *et al*, 1995, 1999; Judd *et al*, 1998a,b, 1999; Simon *et al*, 2001; Fava *et al*, 2002; Kanai *et al*, 2003), and a more stable, enduring state (Koran *et al*, 2001). Thus, symptoms that are present in the context of a response that falls short of full remission portend a greater vulnerability to greater symptomatology in the future. Indeed, the extent of residual symptoms following the acute treatment of the MDE is one of the few consistent predictors

of relapse during continuation treatment (Priest and Kupfer, 1986). Given these implications of remission for function and prognosis, remission is the accepted goal of acute treatment of MDD (Rush and Ryan, 2002; Depression Guideline Panel, 1993; American Psychiatric Association, 2000b; Anderson *et al*, 2000; Canadian Psychiatric Association Network for Mood and Anxiety Treatments (CANMAT), 2001; Reesal and Lam, 2001; Bauer *et al*, 2002a,b; Hirschfeld *et al*, 1997; Ballenger, 1999; Trivedi *et al*, 1998; Crismon *et al*, 1999; Rush *et al*, 2003a, 2004; Fava *et al*, 2003b; the American Academy of Child and Adolescent Psychiatry, 1998). However, not all depressed patients may reach remission, even with multiple treatment attempts.

There has been a remarkable increase in the reporting of 'remission' as an outcome in acute phase efficacy trials of MDE (eg, Keller *et al*, 1998; Sackeim *et al*, 2000; Alpert *et al*, 2002, 2004; Rush *et al*, 2003a; Bielski *et al*, 2004; Fava *et al*, 2004; Goldstein *et al*, 2004; Montgomery *et al*, 2004; Nelson *et al*, 2004; Perlis *et al*, 2004; Prudic *et al*, 2004; Trivedi *et al*, 2004a) and in pooled or meta-analyses (Lepola *et al*, 1998; Beasley *et al*, 2000; Entsuah *et al*, 2001; Thase *et al*, 2001; Dawson *et al*, 2004). In fact, reports of differential medication effects have resulted from some *post hoc* analyses using pooled samples in short-term (8-week) acute phase medication trials. These reports have relied on a variety of reasonably chosen remission end points at study exit (eg, 17-item Hamilton Rating Scale for Depression ≤ 7 or ≤ 6 (HRSD₁₇; Hamilton, 1960, 1967) or Montgomery-Åsberg Depression Rating Scale ≤ 10 (MADRS; Montgomery and Åsberg, 1979)). These studies, however, raise several questions addressed in this report by the sparsely available data or by consensus opinion.

Remission also provides an easily understood metric of the differences between treatments under study (eg, drug vs placebo). The percent remitted (and the concomitant reporting of the number needed to treat—see below) is certainly clinically more relevant and understandable than differences between groups in terms of baseline to exit changes in overall symptom severity. Treatments with a greater likelihood of attaining remission, a more rapid onset of remission, or a higher probability of sustaining remission have clear therapeutic advantages, given the implications of remission for function and prognosis.

In addition, the definition of remission has implications for how one characterizes the course of depressive illness. For example, the offset of the current MDE should be ascribed when patients (a) have fully recovered (ie, has achieved remission and sustained it long enough to declare recovery), (b) no longer meet MDE criteria but still have residual symptoms (ie, are not in remission), or (c) have met the definition of remission but not recovery. This decision affects whether patients are said to have a single episode or recurrent course, how the number of episodes is estimated, whether one declares the patient to be in or out of an episode, etc.

Evaluations of Remission, Recovery, Relapse, Recurrence

Surprisingly, few studies were found that empirically evaluated the concepts of response, remission, relapse, recovery, and recurrence. Riso *et al* (1997) used data from a

study of depressed outpatients treated acutely for 16 weeks with cognitive behavior therapy and followed up for 3 years. They defined *partial remission* as occurring when a $\geq 50\%$ reduction in baseline HRSD₁₇ and an HRSD₁₇ ≤ 10 had been met. *Remission* was defined as HRSD₁₇ ≤ 6 over a ≥ 3 -week period. *Relapse* was declared after subjects met MDE criteria for ≥ 2 weeks and the HRSD₁₇ was ≥ 14 and both occurred before recovery had begun. *Recovery* was said to have begun after at least 6 consecutive months with all HRSD₁₇ ratings ≤ 6 . Note that in this report, a relapse could occur from 'a state of partial remission' (ie, full remission need not have occurred).

Several measures, including the Beck Depression Inventory (BDI) (Beck *et al*, 1961), Global Assessment Scale (GAS) (Endicott *et al*, 1976), Dysfunctional Attitudes Scale (DAS) (Weissman, 1979), and the Automatic Thoughts Questionnaire (ATQ) (Hollon and Kendall, 1980), gathered at acute trial exit, distinguished remitters from those with a partial remission (response with residual symptoms), which is not surprising, as one group is more symptomatic (overall) than the other. Most measures also distinguished those with and without relapse, and those with and without recovery. Recurrence, but not relapse, was associated with history of prior MDEs.

This study suggests that remission is distinct from response with residual symptoms or partial remission. However, the features distinguishing remission and partial remission were largely correlates of symptom severity (ie, HRSD₁₇ total scores), which in turn was used to define these states. Furthermore, the definitions advanced by Riso *et al* (1997) have not been further evaluated in subsequent research.

Another study examined the impact of using various criterion symptom levels to define the end of an MDE on MDE duration (Philipp and Fickinger, 1993). When a stringent threshold was used, reflecting the virtual absence of symptoms (eg, only 0 or 1 criterion symptoms present), the median episode duration was 26.5 weeks. If the presence of 2–3 symptoms was used to establish an end to an MDE, the median episode duration was 10.0 weeks. As expected, the more stringent the remission criterion, the longer it takes to achieve remission. Thus, use of different operational definitions of remission can lead to radically different descriptions of the course of illness, including both the number and duration of MDEs.

In sum, while the clinical importance of remission is widely accepted, only minimal empirical evidence is available to validate a specific operationalization of this concept.

Ascertainment of Remission

Clinical ratings quantify symptoms, categorize illness severity, and demarcate thresholds for screening purposes. Scores on symptom measures are often used to define response, remission, and relapse. A careful evaluation of the most valid and useful rating instruments for both research and clinical use is imperative. Patient self-reports or clinician ratings of depressive symptoms can be affected by cultural context, alliance, personality style, age, and prior or current life experiences. For example, before treatment initiation, the relationship between patient and clinician

ratings has repeatedly been shown to be modest, sharing on average only 25% common variance (Sayer *et al*, 1993). The degree of discrepancy between patient and clinician ratings of depression severity may be affected by incentives, depression subtype, and the patient's cognitive capacity. Patients with psychotic features are more likely to report minimal symptomatology, while being rated by others as severely depressed. However, clinician and patient rating scales that rate identical symptoms substantially increase agreement between these two perspectives (Rush *et al*, 2006; Trivedi *et al*, 2004b).

Furthermore, clinician assessment of depressive symptoms can have remarkable reliability, especially when a structured interview guide is used. In clinical trials, interrater reliabilities are often at least 0.90 (Baca-Garcia *et al*, 2001). This degree of agreement is within the range of reliability values found for many laboratory tests, such as cardiopulmonary measures during treadmill stress tests (Dobrovolsky *et al*, 2003), use of commercial kits to detect autoimmune disorders (Fritzler *et al*, 2003), and plasma assays for various substances (Fears *et al*, 2002; Wilson *et al*, 2002).

Remission and Recovery

Patients who have remitted and those who have recovered have no or minimal symptoms; thus, they may be clinically indistinguishable. Consequently, the distinction between remission and recovery depends on the interval following symptom reduction that reflects the resolution of the underlying neurobiology of the MDE. Presently, beyond symptoms, there are no validated biomarkers to distinguish remission from recovery. A corollary is that the probability of return to a symptomatic state is much higher for patients who have only achieved a brief period of remission as compared to those who have reached recovery. Consequently, operationalization of these concepts hinges on requiring a longer interval of sustained remission and minimal symptom expression to ascribe recovery relative to remission. It is possible that the remission/recovery distinction may not be valid if the vulnerability does not change over time.

Panel Focus

The panel met on several occasions and conducted literature reviews to identify studies to inform these issues. Given the relative paucity of studies aimed at validating the concepts put forth by Frank *et al* (1991), the panel chose to reach consensus on the specifics of the relevant concepts using the literature, clinical experience, and logic. The recommendations are largely not evidence-based. Rather, they are put forth as recommendations that should be viewed largely as hypotheses in need of empirical study.

The panel focused on the implications of remission for: (1) the conceptualization of remission as well as of response, recovery, relapse, and recurrence; (2) methods needed to operationalize these concepts; (3) the design of efficacy and effectiveness trials with remission as a primary end point; and (4) clinical practice and future research.

Specific questions considered by the Task Force included the following:

(1) What symptom criteria should be used to define remission? (2) Should the definition of remission include the requirement of a return to normal or to premorbid functional capacity? (3) What symptoms, if any, may still be present in remission? (4) Should a minimal duration be required to ascertain remission? (5) When and how should one decide that the patient has 'lost' or 'left' the remitted state (even if return to a full MDE has not occurred), given the well-known symptom fluctuation in 'remitted' patients? (6) What are the implications of these considerations for recovery, relapse, and recurrence?

FACTORS AFFECTING REMISSION

The Task Force recognized that several factors affect the likelihood of attaining remission, the time to remission, and the durability of the remission. Such factors likely include the type, dose, and duration of treatment; baseline symptom severity; the degree of treatment resistance; the presence of concurrent Axis I, II, or III conditions; environmental supports and stressors; the prior course of illness (eg, chronic vs acute illness); and individual genetic vulnerability. It is likely that these factors also influence the likelihood, time to, and durability of recovery.

Electroconvulsive therapy (ECT), for example, typically results in remission in 1–3 weeks (Daly *et al*, 2001), whereas Interpersonal Psychotherapy (IPT) (Klerman *et al*, 1984) or Cognitive Therapy (CBT) (Beck *et al*, 1979) may take 6–10 weeks. With medications, remission may begin within 4–12 weeks (or longer) after beginning treatment (O'Leary *et al*, 2000; Chilvers *et al*, 2001; Trivedi *et al*, 2001, 2006; Quitkin *et al*, 2003).

Higher baseline depressive symptom severity is typically associated with longer times to the onset of remission (Tedlow *et al*, 1998; O'Leary *et al*, 2000), and therefore, to a lower probability of remission at earlier time points (eg, at 6, 8, or 12 weeks). Thus, remission rates in a fixed time period will vary across studies that differ in baseline severity.

The degree of treatment resistance affects the likelihood of and time to achieving remission (Prudic *et al*, 1990, 1996; Sackeim *et al*, 2001), as well as the stability of the remitted state once achieved (Nierenberg *et al*, 1994; Sackeim *et al*, 1990, 1993, 2000, 2001). That is, patients with treatment-resistant as opposed to nonresistant depressions may be more likely to suffer symptom fluctuation (ie, roughening) during the remitted state (ie, there are brief exacerbations of symptoms that do not qualify as relapse or recurrence), and the duration of the remitted state may be shorter, although further research is called for.

The co-occurrence with depression of certain Axis I disorders (Fava *et al*, 1997), Axis II disorders (Ezquiaga *et al*, 1999; Viinamaki *et al*, 2002; Prudic *et al*, 2004), or Axis III disorders (Keitner *et al*, 1991, 1992; Iosifescu *et al*, 2003) may prolong the time needed to reach remission or render remission less likely than is the case with uncomplicated depressions. In addition, it is suggested but not unequivocally established that long-standing (ie, chronic) depressions may take longer to remit or are less likely to remit (Prudic *et al*, 1996; Fava *et al*, 1997; Keller *et al*, 1998, 2000).

RESPONSE REVISITED

Concept of Response

The Task Force agreed with the extant literature that response implies a clinically meaningful degree of symptom reduction, which is usually accompanied by an improvement in the patient's mood, daily function, and/or pain/distress. The Task Force also recognized that identification of a 'response' is clearly useful to clinicians and patients, who must decide ultimately whether to continue, adjust the dose of, add to, or discontinue current treatment. These clinical decisions are inherently categorical and legitimately call for an outcome that provides a yes/no answer for each patient. The concept of response is temporally linked to the onset or change during treatment (even if only watchful waiting) even though response, however defined, does not imply a causal relationship to the treatment itself.

The Task Force identified a number of limitations of using response as a predefined goal of treatment or as a primary outcome criterion in clinical trials. Response strongly depends on the initial pretreatment symptom severity value and its ascertainment requires the systematic assessment of symptoms *before* and *during* treatment. Any unreliability in assessing initial symptom severity, therefore, directly affects the reliability of recognizing a response. Regression to the mean may further foster the invalid impression of symptomatic improvement (Fava *et al*, 2003a).

Furthermore, the recognition of a 'clinically significant' benefit depends on the initial state from which change is measured, the clinical purpose in ascribing response, and the clinical context. For example, a modest benefit in a highly treatment-resistant depression may be more clinically significant than a greater benefit in a nontreatment-resistant depression. Specifically, while a convention of a $\geq 50\%$ reduction in baseline severity is commonly accepted, it may not be adequate for defining clinically significant benefits in more severely ill or highly treatment-resistant patients (Rush *et al*, 2003b).

Finally, patients have a range of initial values and a differential propensity for regression to the mean (see below). There are serious problems in comparing one subject with another in randomized controlled trials (RCTs) using response because it depends highly on the initial level of symptom severity. For example, one more severely depressed patient at baseline may respond, yet still be worse off at treatment exit (in terms of symptoms, behavior, functioning, or pain/distress) than another who does not respond, but who began with a less severe baseline depression (Tedlow *et al*, 1998).

For this reason, response is better used to monitor changes within a subject (for whom initial severity is fixed) for individual clinical decision-making, than in comparing response rates between subjects for whom initial values range widely (as is the case in RCTs).

Should Response Refer Only to Criterion Symptoms?

Noncriterion symptoms that commonly are associated with MDD include anxiety, panic attacks, irritability (in adults), hopelessness, avoidance, or cognitive dysfunction. The Task Force recommended that response should be defined solely

by the nine core criterion depressive symptoms specified in DSM-IV-TR (American Psychiatric Association, 2000a) to define an MDE because these associated symptoms may be a function of other commonly concurrent Axis I, II, or III conditions (Rush *et al*, 2005b), in which case these noncriterion symptoms may or may not respond to the treatment under study for MDD.

The Task Force recommended that evaluation of associated noncriterion symptoms (as secondary outcomes) be conducted by both clinicians and researchers. For example, such studies could determine whether the same antidepressant medication could impact both core depressive and associated anxiety symptoms. As several anxiety disorders and MDD may share similar genetic background vulnerabilities (Kendler, 1996), the lack of response of anxiety symptoms in treatment trials may have implications for identifying different pathophysiological and diagnostic subtypes. The prognostic relevance of residual noncriterion-associated symptoms deserves study.

Should the Definition of Response Include Daily Function?

The Task Force recommended that the definition of response should not include an assessment of function for several reasons. Response is usually associated with improved function (Miller *et al*, 1998), and when remission or recovery is reached in depressed patients with no other concurrent psychiatric or general medical conditions, function typically returns to the premorbid levels (Small *et al*, 1996; Miller *et al*, 1998; Hirschfeld *et al*, 2002; Iosifescu *et al*, 2003; Ormel *et al*, 2004). However, the level of day-to-day functioning is affected by both the depressive disorder and associated Axis I and Axis III conditions. For example, elderly or medically fragile patients, or patients with severe concurrent Axis I or Axis II conditions, often suffer functional impairment for reasons substantially unrelated to MDD. In addition, time lags between depressive symptom response and functional improvement have been reported (Mintz *et al*, 1992). Furthermore, premorbid functioning is highly predictive of posttreatment functioning in depression (Ormel *et al*, 2004). Thus, while depression affects day-to-day function, function may also be relatively independent of the outcome of the treatment for depression. Therefore, in the assessment of the efficacy of antidepressant treatment, change in function is likely to be a potentially less sensitive indicator of short-term, acute improvement than the core MDE criterion symptoms. On the other hand, function should be assessed and reported as a secondary outcome as it informs us about the clinical significance of the symptom changes achieved.

How should Response be Operationalized?

The Task Force recommended that the definition of response be chosen to define a clinically meaningful benefit in the context of the population under study, taking into account treatment resistance, initial severity, and other clinical factors. Typically, response has been defined as a $\geq 50\%$ reduction in pretreatment symptom severity (eg, with the HRSD₁₇, the MADRS, or the Inventory of Depressive Symptomatology). On the other hand, different

degrees of symptom reduction (eg, $\geq 25\%$) in patients with highly treatment-resistant depression may reflect significant relief in those populations. The specific percent reduction chosen should reflect a clinically significant benefit, which, as noted above, depends on the initial severity level, on the context (eg, treatment resistant or not), and on the rating scale.

What is the Minimal Time Period for Ascertaining Response?

The Task Force recommended that response criteria be met for 3 consecutive weeks to take into account error in the assessment of symptomatology and unstable symptomatic fluctuations. Requiring that response criteria be met for a reasonable period of time guards against miscategorizing transient improvement as a clinically significant benefit (ie, a response). In practice, weekly assessments to ascertain response may be unpractical, however.

One might consider identifying a provisional response when the response criterion is first met, then identifying a definite response when the response criterion is still met after an additional 2 weeks (Sackeim *et al*, 1987, 1993, 2000, 2001).

Research Recommendations

Alternative definitions of minimal time periods to declare response deserve study, as does the recommendation to exclude day-to-day function in the definition of response. Studies are needed to better define the norms for linking different levels of symptom reduction with different degrees of functional improvement. As these associations are imperfect, it is important to know whether discrepancies in the degree of symptom improvement and functional improvement have prognostic relevance in general, or in specific for particular groups of depressed patients (eg, the chronically vs acutely ill).

REMISSION REVISITED

Concept of Remission

Remission implies that the signs and symptoms of the illness must be absent or close to it. Remitted patients have a better prognosis and better function than those with only a response without remission. In fact, remission is typically associated with a return to the day-to-day function that was typical for the patient before the onset of any depressive symptoms. Remission, unlike response, entails an absolute allowable ceiling level in symptom expression. Remission may be ascribed whether or not patients are receiving treatment.

Should Remission Refer Only to the Core Criterion Symptoms?

The Task Force recommended that remission refer only to the nine criterion symptom domains identified in DSM-IV-TR to diagnose a major depressive episode. (Should the definition of MDE change (eg, core criterion symptoms added or deleted), operationalizing remission will require

revised methods.) This recommendation is consistent with the recommendation (above) for response (to use solely the nine criterion symptoms), and is based on the evidence to date that demonstrates the relevance of remission to function and prognosis (ie, most studies have focused on core depressive symptoms). Noncriterion associated symptoms may be of use as secondary outcomes, although there are insufficient data to date on this issue.

Should the Definition of Remission Include Daily Function?

The Task Force recommended that daily function should not be part of the definition of remission for the same reasons noted for response. It is recognized that remission is typically associated with a return to premorbid day-to-day function. Daily function may provide an important independent validation of symptom remission. Again, function should typically be measured and reported as a secondary outcome.

What is the Minimal Time Period Required to Ascertain Remission?

The Task Force recommended that 3 consecutive weeks must pass, during which each week is characterized by the virtual absence of depressive symptoms, before remission can be ascribed. It was felt that remission should be ascribed once a sufficient time has passed, such that the remitted state is likely to persist in many patients. It was our estimate that one might expect remission to be sustained if it was present for at least 3 consecutive weeks (to ensure that transient fluctuations were not designated as remission).

What Symptoms may be Present in the Remitted State?

The Task Force recommended that neither sad mood nor loss of interest/pleasure may be present in the remitted state and, further, that fewer than three of the seven additional core criterion symptoms may be present (eg, poor concentration, disturbed appetite/weight, disturbed sleep, etc.). The Task Force recognizes the highly specific nature of this recommendation deserves study even though it does have face validity. We felt that the presence of either essential symptom (reduced interest/pleasure, sad mood) would likely be associated with a worse prognosis than if both were absent, and that a simple count of symptoms (eg, presence of three or four as opposed to five of the nine criterion symptoms) provided an incomplete description of the remitted state. The basic notion underlying this recommendation was that depression at its core represents a hedonic deficit that is best captured by these two depressive symptoms. Thus, if either symptom was present, the disorder would not be truly remitted.

When is Remission Lost?

The Task Force recommended that remission can end only (a) with a return of the index MDE (ie, a relapse) or (b) with a new MDE (ie, a recurrence). In sum, the state of remission is not lost until either relapse or recurrence occurs. That is,

once remission has begun, remitted patients may display roughening of the remitted state (some subsyndromal symptoms, insufficient to qualify for an MDE diagnosis, may be encountered without loss of the remitted state). This symptomatic roughening should be called subsyndromal symptoms following remission or partial remission.

How Should Remission be Measured?

The proposed definition of remission logically requires that rating scales used to operationalize remission must include all nine core criterion symptom domains used to diagnose an MDE by DSM-IV-TR. Rating scales that identify all nine criterion domains include the nine-item self-reported Patient Health Questionnaire (PHS; Kroenke *et al*, 2001), the 16-item Quick Inventory of Depressive Symptomatology available as a clinician rating (QIDS-C₁₆) or self-report (QIDS-SR₁₆) (Rush *et al*, 2003c; Trivedi *et al*, 2004b), and the BDI-II (Beck *et al*, 1961, 1996)—a self-report. The BDI-II does not include weight gain, but does otherwise include all other criterion symptom domains.

Note, however, that total scores on selected rating scales are insufficient to ascertain remission if one uses our proposed definition that rests on the nine core criterion symptoms. Therefore, it is not recommended that total score thresholds be used alone to declare remission, if our proposed definition of remission is adopted. For example, the HRSD₁₇ does not include oversleeping, weight and appetite increase, or concentration/decision-making. The MADRS (Montgomery and Åsberg, 1979) does not include oversleeping and overeating, as well as interest (though it assesses inability to feel), and energy (though it assesses lassitude), self-criticism (guilt), and psychomotor changes. There are many ways, however, to arrive at the same total score.

The field has estimated remission using total score thresholds on these and other rating scales, without reference to the above-recommended definition. If one chooses the HRSD₁₇ to estimate remission, the Task Force suggested that an HRSD₁₇ score of ≤ 5 (Ghatavi *et al*, 2002) or ≤ 7 (based on the precedent in the literature) be used. For example, Nierenberg *et al* (1999) found that only 17.6% of patients with an HRSD₁₇ ≤ 7 had no symptoms of MDD.

An HRSD₁₇ ≤ 7 corresponds to an MADRS score ≤ 9 (Carmody *et al*, in press), or a 30-item Inventory of Depressive Symptomatology—Clinician-Rated (IDS-C₃₀) score ≤ 12 , an Inventory of Depressive Symptomatology—Self-Report (IDS-SR₃₀) (Rush *et al*, 1996, 2003c) score ≤ 14 , or a QIDS-C₁₆ or QIDS-SR₃₀ score of ≤ 5 . The corresponding PHS-9 score is likely ≤ 5 (Kroenke *et al*, 2001). Alternatively, Zimmerman *et al* (2004a, b, c) have recommended an MADRS total score of ≤ 5 to define remission.

Research Recommendations

The Task Force recognizes that the above recommendations by which to define remission in terms of 'minimal' symptoms and the 3-week duration are somewhat arbitrary. These and alternative definitions call for empirical studies that relate different symptom and duration criteria to prognosis and function. For example, the survival and

hazard curves of the time from onset of remission (variously defined) to relapse could be examined, and a criterion (symptom level and duration) at which the hazard function stabilizes at a low level could be identified to validate the proposed or alternative operationalizations of remission. Alternatively, validation of a definition might be tested against a criterion such as maximal differentiation between 'remitted' and 'not remitted' patients on an independent measure of daily function.

Whether remitted patients who do not achieve normal function have a worse prognosis, and whether treatments that target this postremission functional impairment result in better prognoses deserve study. Data analyses that provide a 'crosswalk' between total scores on commonly used symptom measures to establish 'equivalent' symptom severity thresholds, including remission (eg, see Rush *et al*, 2003c), are encouraged because these conversion tables help in the comparison of studies that use different symptom measures.

Research to identify specific factors that affect the likelihood of, time to, and duration of remission would provide much needed data for the design of more efficient and clinically informative trials. Secondary end points, such as function or the status of commonly associated, but noncriterion, symptoms should be reported in efficacy and effectiveness studies for the reasons noted above. Given the potential limitation of excluding noncriterion symptoms in the proposed definition of remission, studies to examine this recommendation and studies to identify the most common continuing noncriterion symptoms in the remitted state are recommended.

RECOVERY REVISITED

Concept of Recovery

Recovery implies an extended period of remission such that an MDE is unlikely to occur in the near future. That is, recovery implies that the remitted state persists long enough and has sufficient consistency that many future months of remission can be anticipated for most patients. As is the case with remission, recovery may be ascribed while the patient is either on or off treatment. Recovery, once present, can only be lost if a recurrence occurs (ie, subsyndromal symptoms may occur without loss of the 'recovered' status). In theory, recovery implies that those disease processes that are immediately involved in the expression of the syndrome are arrested or corrected such that the syndromal expression is no longer present. On the other hand, underlying vulnerability to subsequent syndromal episodes may remain. Thus, recovery is recovery not from the illness but from the last MDE. As with response and remission, the Task Force recognized that some patients may not be able to enter a period of recovery given the limitations of current treatment options.

Once recovery is ascribed, it is logical to consider discontinuing treatment (depending on the past history of individual patients). For example, for individuals with single episode MDD, several treatment guidelines (Depression Guideline Panel 1993; APA, 2000b) recommend that treatment be discontinued after 4–9 months of continuation phase treatment following recovery from the index MDE.

Should Symptoms be Used to Define Recovery?

The Task Force recommended that recovery should be defined only by symptomatic status for the same reasons recommending symptom status alone be used to define remission. As with remission, recovery does not require normalization of day-to-day function, although it often occurs.

When should Recovery be Ascribed?

The Task Force recommended that recovery be ascribed after at least 4 months of remission. Riso *et al* (1997) used a 6-month duration to define recovery with evidence of validation based on the prior course of illness. By definition, recovery can only occur after remission has been ascribed. The main reason for the 4-month recommendation is that placebo-controlled trials of continuation therapy and naturalistic studies of relapse in remitted depressed patients indicate that the great majority of relapses occur within the first 4 months of the year following the onset of remission (Reimherr *et al*, 1998). To ensure that recovery has occurred, frequent enough measurements must be made to detect a return of the index MDE (ie, every 2 weeks). The other rating scale recommendations for defining remission apply to recovery. When symptoms appear during or following recovery that are insufficient to meet criteria for an MDE, the term 'subsyndromal symptoms following recovery' is recommended.

Research Recommendations

Research to empirically test these recommended definitions (eg, 4-month requirement) is needed. In addition, research to identify neurobiological or other clinical markers to assess the presence of recovery (ie, episode-dependent markers) is needed. Finally, whether different treatments differ in the durability of recovery, once achieved, is unknown and deserves study.

RELAPSE AND RECURRENCE REVISITED

Concepts of Relapse and Recurrence

Both relapse and recurrence refer to the return to an MDE, rather than the reappearance of selected symptoms that are insufficient in number, duration, or intensity to diagnose an MDE. Relapse and recurrence differ, however, with respect to the time at which this episode occurs following remission (relapse) or recovery (recurrence). Relapse occurs *before* recovery but after remission is ascertained. Recurrence occurs only *after* recovery is ascribed. The Task Force recommended that *relapse* be designated as the time point at which the syndrome of depression returns with sufficient manifestation of core criterion symptoms to meet the criteria for an MDE by DSM-IV-TR (ie, ≥ 5 of the nine criterion symptoms for ≥ 2 weeks). Similarly, recurrence should denote the end of a period of recovery by the return of the depressive symptoms of sufficient severity to meet criteria for a new MDE by DSM-IV-TR (ie, following recovery).

'Roughening,' 'depressive breakthroughs,' and 'symptomatic blips' are all terms that refer to subsyndromal

depressive symptoms (ie, symptoms that are not sufficient to meet MDE criteria) that occur following the onset of remission or recovery. Such fluctuations should be monitored and, where appropriate, described in research reports. The Task Force recommended that neither remission nor recovery, once ascribed, can be lost without the onset of a relapse or recurrence (ie, onset of an MDE defined by DSM-IV). In this scheme, there is a well-defined sequence of events beginning with onset of the first MDE, which may or may not be followed by remission. Remission may be followed by either relapse or recovery. Relapse may be followed only by remission. Recovery may be followed by recurrence. Recurrence may be followed by remission. No other transitions are possible.

Research Recommendations

Investigations are needed to define empirically the optimal duration of remission sufficient to determine the onset of recovery. Whether states of stability *vs* instability (ie, with or without roughening) in periods of remission or recovery have prognostic relevance or significantly impact day-to-day function deserves study.

IMPLICATIONS FOR CLINICAL TRIAL DESIGNS AND REPORTS

The above Task Force recommendations have important implications for the design and execution of efficacy and effectiveness trials, particularly when remission is the primary outcome.

Clinical Trial Durations

Remission typically follows response by at least several weeks (O'Leary *et al*, 2000; Chilvers *et al*, 2001; Koran *et al*, 2001; Trivedi *et al*, 2001; Quitkin *et al*, 2003; Trivedi *et al*, 2006). Consequently, trials with remission as an end point may need to be longer. If remission is the primary outcome, the trial should be of sufficient duration that remission can occur in most, if not all, subjects.

When trial duration is extended, the difference between two treatments (or a treatment and placebo) should increase over time if they differ in efficacy (at least in theory). In such cases, the longer the duration of the trial, the greater the effect sizes, and, consequently, the smaller the samples needed to detect treatment-control differences. In prolonged trials, the chances of spontaneous remission increase, and furthermore some patients will initially remit, but over time these same patients may subsequently suffer a relapse. The latter will confound acute phase outcomes (remission) with longer term outcomes (relapse). In addition, both ethical and feasibility issues, especially with a placebo control group, will be encountered with prolonged trials, although both issues may be addressed by predefined triage points (see below).

Optimal trial durations will likely depend upon several of the factors noted above that affect the likelihood of or the time to remission (eg, initial symptom severity, the type and delivery of the treatment, degree of treatment resistance, concurrent disorders, etc.). Thus, the optimal acute treatment trial duration is the time at which the treatment-

control effect size is maximized. We found no randomized comparisons of two different acute phase medication trial durations (eg, 8 *vs* 16 weeks) in terms of time to and probability of remission in depressed patients. It is possible that the time to response or remission may be affected by trial length, as that may affect patient expectations and efforts to resolve life problems that precipitated or that maintain the depression. Thus, while longer trials are more likely to detect potential remission rates, the actual trial duration must be guided by when one expects remission to occur. That timing is highly dependent on the multiple clinical factors noted above. For example, remission is unlikely after more than 8–10 ECT treatments (Husain *et al*, 2004), yet it may increase following even several months of treatment with medication in chronic depression (Koran *et al*, 2001) or vagus nerve stimulation (Rush *et al*, 2005a).

The Task Force recommended that for studies of currently available medications, when remission is the primary outcome, acute trials should be of at least 12 weeks duration. Shorter duration (eg, 8 weeks) may be satisfactory to differentiate two treatments (eg, drug/placebo), with the caveat that neither treatment will have been used long enough to provide a full picture of the actual remission rates achievable. An upper limit to the duration of acute phase trials with remission as an end point should be chosen such that those who have not yet remitted are unlikely to remit in the short term. Absent this information, the Task Force recommended an upper limit of 20 weeks for acute phase trials (with current medications or depression-targeted psychotherapies) when remission is the primary outcome. A total of 20 weeks is recommended because we believe that such a time period will maximize the opportunity for most subjects to reach remission. Again, this issue deserves empirical study. These recommendations for a 12–20-week trial duration are neither evidence-based nor may they be fully applicable to a wide range of treatments and patient population.

The Task Force recognizes the need to balance the desire for sufficient treatment exposure to ensure that remission can occur in those able to reach remission with the need to protect those for whom remission will not occur even with prolonged, ineffective treatment. Consequently, the traditional fixed duration design may not be ideal for dealing with these contradictory aims.

Development of Triage Points

A duration adaptive design (DAD) is one means to balance the needs for longer trial duration for potential remitters and shorter durations for those assessed as not likely to benefit (Agras *et al*, 2000). For both practical and ethical reasons, longer trial durations are acceptable only if there are early exit rules that can be executed at specific points in time (triage points) to remove subjects who are unlikely to remit with further treatment. That is, triage points are points in time during the course of treatment that clinicians or investigators may decide to remove (ie, triage out) patients from the study treatment. In practice, patients are typically triaged out early in the case of intolerable side effects (ie, the clinician decides that even if effective, the patient cannot safely continue the treatment). Later

triage points occur when one decides that the ultimate goal of the treatment (ie, remission) will not be achieved without a change in the type of treatment (eg, switching or augmenting).

Triage points may occur at various times, have varying degrees of reliability, and can be defined by various rules (eg, percent change from baseline, absolute severity score at some fixed time point or interval). Triage points and the relevant rules will vary for different groups of patients, types of depression, types of treatment, etc., as noted above.

If triage points specific to the type of patient or condition, such as comorbidities, course of illness, gender, age, and severity, could be empirically identified and appropriate thresholds by which to identify those patients unlikely to remit could be developed, future study designs would be remarkably advanced. With the empirical definition of triage points, those who are removed from the trial at the triage point would be declared 'failures' (ie, not achieving remission). Some suggestions have been made for patients with eating disorders (Agras *et al*, 2000), depression (Nierenberg *et al*, 1995; Quitkin *et al*, 2003; Trivedi *et al*, 2006), or bipolar disorder (Frank *et al*, 2001). We do not recommend the use of unproven triage points as this would compromise both the power of the trials and bias effect sizes.

Clinical Trial Procedures

As the ascertainment of remission based on the above Task Force recommendations requires 3 consecutive weeks with minimal to no symptoms, research assessments should be obtained weekly with psychometrically acceptable symptom measures that assess all nine criterion symptom domains. While measures of daily function or quality of life do not define remission given the Task Force recommendations, such assessments are recommended for secondary analyses (eg, patient satisfaction measured by the Quality of Life Enjoyment and Satisfaction Questionnaire (QLESQ) (Endicott *et al*, 1993), or patient perception of mental or physical health measured by the MOS 36-item short-form health survey (SF-36) (Ware and Sherbourne, 1992)).

Analyzing and Reporting Results

The statistical testing of clinical trial results can be accomplished with categorical outcomes (eg, % remitted at exit or at preselected times in the trial). Alternatively, one can use time to remission in a survival analysis (using all remission times up to the preselected total trial duration with remission at that time as a 'censored' observation). The power to detect treatment-control differences is always less with a categorical outcome than with survival analysis (Cohen, 1983). The longer the preselected trial duration, the greater the power to detect differences.

Thus, with limited sample sizes and longer trial durations, survival analysis, especially in populations unlikely to show spontaneous remissions, has significant advantages. Whether the analysis is based on comparing categorical outcomes at a fixed time, or on survival analysis, the Task Force recommended that survival curves for both groups should be presented to facilitate the evaluation of

the clinical significance of any statistically significant differences.

The Task Force also strongly recommended that trials report the number to treat (NTT) at various times after treatment initiation to better evaluate clinical significance. The NTT is the number of subjects that need to be treated in order to obtain one additional subject reaching remission at each follow-up time in the treatment over what would have been achieved in the control group (Cook and Sackett, 1995). To illustrate, $NTT = 1$ means that every subject in the treatment group and none in the control group remits—a very unlikely finding. $NTT = 5$ means that for every five patients treated, one would expect one more success in the treatment group than in the control group. The lower the NTT, the more meaningful and clinically effective the experimental treatment compared to the control. NTT is easily computed from the survival curves, for at each follow-up time, NTT equals $1/(\% \text{ remitted in the experimental treatment group minus } \% \text{ remitted in the control group})$. This provides a clear benchmark of clinical significance of the between-group differences. NTT will inevitably be very large at 4 weeks (with a criterion of remission requiring 3 weeks of being essentially symptom free), and will decrease (improve) as the follow-up time increases with effective treatment.

Following completion of the clinical trial, the Task Force recommended that additional moderator analyses be conducted to try to identify baseline features that identify remitters vs nonremitters (Kraemer *et al*, 2002). These results may sharpen the evidence by which to select an agent or to select among agents, and begin to address the question of whether agents might differ in their spectra of action. If such moderators are identified, they could suggest eligibility criteria or stratification factors for future studies (thereby increasing the power to detect treatment effects without increasing sample sizes).

If researchers adopted these Task Force recommendations in reporting trial results, clinicians would directly benefit. First, if remission was to become the primary outcome in most trials and if survival curves and the NTT at each time point for most trials were published, the ensuing greater cross-study consistency would assist clinicians in evaluating the efficacy or effectiveness of different treatments, and help them to assess the clinical significance of trial results. Second, the identification of triage points would be of immense help to clinicians in deciding when to discontinue or modify a current treatment that is unlikely to work fully. Moreover, the relative success rates at various triage points might provide a more clinically useful approximation of 'response.' Third, if moderators could be identified, clinicians would better know which treatments are more likely to 'work' with which types of patients.

IMPLICATIONS FOR CLINICAL PRACTICE

Remission is the desired goal of acute treatment, and sustained remission is the desired goal of long-term treatment. While currently available treatments do not uniformly result in remission, clinicians should endeavor to ensure that each patient is treated as optimally as possible to achieve this outcome, given the adverse implications of

not achieving remission (in terms of function and prognosis). Thus, clinicians must decide for each patient whether further treatment changes are likely or unlikely to increase the chances of remission and at what cost (eg, side-effect burden).

If these recommendations were adopted for daily practice, clinicians would need to (1) specifically and repeatedly measure core criterion depressive symptom severity to guide the implementation and timely modification of treatment, (2) conduct sufficient visits or measurements to establish that 3 consecutive weeks of minimal to no symptoms (ie, remission) has or has not been achieved, (3) systematically inquire about the magnitude and types of side effects and overall side-effect burden, so as to accurately gauge whether the dose or type of treatment needs modification in order to achieve remission in a time-efficient fashion, and (4) follow the trajectory of

Table 1 Summary of Task Force Recommendations

Remission

- Should become a primary end point in acute phase clinical trials.
- Is present if for ≥ 3 weeks there is an absence of both sad mood and reduced interest and no more than three of the remaining seven symptoms of a major depressive episode (MDE) are present.
- May be followed by recovery or recurrence only.
- Day-to-day function is not part of the formal definition of remission, but is an important secondary outcome.
- Symptoms that occur following remission that are insufficient to diagnose an MDE are 'subsyndromal symptoms following remission.'
- Trials that use remission as a primary outcome need to be of longer duration (12–20 weeks), with sufficiently frequent symptom assessments to establish the presence of remission.
- Remission is not achievable for all depressed patients.

Relapse

- Can only occur following remission but before recovery.
- Requires that DSM-IV-TR diagnostic criteria for an MDE be met.

Recovery

- Can only be declared after ≥ 4 months following the onset of remission.
- Symptoms that do not meet DSM-IV-TR criteria for an MDE that occur following recovery are 'subsyndromal symptoms following recovery.'

Recurrence

- Requires the development of an MDE defined by DSM-IV-TR.
- Can only occur following the onset of recovery.

Research reports should

- Empirically evaluate these recommendations.
- Routinely provide survival curves to describe or compare outcomes across groups.
- Report the number to treat.
- Conduct and report moderator analyses.
- Establish triage points to facilitate longer duration trials.
- Compare different symptom scales.
- Report secondary outcomes based on associated (noncriterion) symptoms, function, and quality of life.

symptom change (or lack of change) such that treatments (dose, type) can be modified in a timely fashion, hopefully informed by empirically defined triage points. The use of a depressive symptom measure to assess the nine criterion symptom domains that define an MDE by DSM-IV-TR (American Psychiatric Association, 2000a) would become routine.

SUMMARY

Table 1 summarizes the Task Force recommendations.

Virtually all of these recommendations are based largely on logic, clinical impression, and consensus. These recommendations and judgments must be evaluated empirically and compared to alternate conceptualizations (ie, an empirical refinement of these recommendations is essential). Both *post hoc* data analyses and prospective studies are strongly recommended. Only with such investigations can we define the best methods and means to operationalize the concepts of remission, recovery, relapse, and recurrence. For practitioners, the regular measurement of core criterion depressive symptoms at frequent enough intervals to facilitate timely treatment changes is recommended to improve the quality of care and outcomes for depressed patients.

REFERENCES

- Agras WS, Crow SJ, Halmi KA, Mitchell JE, Wilson GT, Kraemer HC (2000). Outcome predictors for the cognitive behavior treatment of bulimia nervosa: data from a multisite study. *Am J Psychiatry* 157: 1302–1308.
- Alpert JE, Franznick DA, Hollander SB, Fava M (2004). Gepirone extended-release treatment of anxious depression: evidence from a retrospective subgroup analysis in patients with major depressive disorder. *J Clin Psychiatry* 65: 1069–1075.
- Alpert JE, Mischoulon D, Rubenstein GE, Bottonari K, Nierenberg AA, Fava M (2002). Folinic acid (Leucovorin) as an adjunctive treatment for SSRI-refractory depression. *Ann Clin Psychiatry* 14: 33–38.
- American Academy of Child and Adolescent Psychiatry (1998). Practice parameters for the assessment and treatment of children and adolescents with depressive disorders. *J Am Acad Child Adolesc Psychiatry* 37(Suppl 10): 63S–83S.
- American Psychiatric Association (2000a). *Diagnostic and Statistical Manual for Mental Disorders, 4th edn, Text Revision (DSM-IV-TR)*. American Psychiatric Press: Washington, DC.
- American Psychiatric Association (2000b). Practice guideline for the treatment of patients with major depressive disorder (revision). *Am J Psychiatry* 157(Suppl 4): 1–45.
- Anderson IM, Nutt DJ, Deakin JFW (2000). Evidence-based guidelines for treating depressive disorders with antidepressants: a revision of the 1993 British Association for Psychopharmacology guidelines. *J Psychopharmacol* 14: 3–20.
- Baca-Garcia E, Blanco C, Saiz-Ruiz J, Rico F, az-Sastre C, Cicchetti DV (2001). Assessment of reliability in the clinical evaluation of depressive symptoms among multiple investigators in a multi-center clinical trial. *Psychiatry Res* 102: 163–173.
- Ballenger JC (1999). Clinical guidelines for establishing remission in patients with depression and anxiety. *J Clin Psychiatry* 60(Suppl 22): 29–34.
- Bauer M, Whybrow PC, Angst J, Versiani M, Miller H-J, WFSBP Task Force on Treatment Guidelines for Unipolar Depressive Disorders (2002a). World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for biological treatment of

- unipolar depressive disorders, Part 1: Acute and continuation treatment of major depressive disorder. *World J Biol Psychiatry* 3: 5–43.
- Bauer M, Whybrow PC, Angst J, Versiani M, Moller H-J, WFSBP Task Force on Treatment Guidelines for Unipolar Depressive Disorders (2002b). World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for biological treatment of unipolar depressive disorders, Part 2: Maintenance treatment of major depressive disorder and treatment of chronic depressive disorders and subthreshold depressions. *World J Biol Psychiatry* 3: 69–86.
- Beasley Jr CM, Nilsson ME, Koke SC, Gonzales JS (2000). Efficacy, adverse events, and treatment discontinuations in fluoxetine clinical studies of major depression: a meta-analysis of the 20-mg/day dose. *J Clin Psychiatry* 61: 722–728.
- Beck AT, Rush AJ, Shaw BF, Emery G (1979). *Cognitive Therapy of Depression*. Guilford Press: New York.
- Beck AT, Steer RA, Brown GK (1996). *Beck Depression Inventory*, 2nd edn. Manual, The Psychological Corporation: San Antonio, TX.
- Beck AT, Ward CH, Mendelson M, Mock JE, Erbaugh JK (1961). An inventory for measuring depression. *Arch Gen Psychiatry* 4: 561–571.
- Bielski RJ, Ventura D, Chang CC (2004). A double-blind comparison of escitalopram and venlafaxine extended release in the treatment of major depressive disorder. *J Clin Psychiatry* 65: 1190–1196.
- Canadian Psychiatric Association Network for Mood and Anxiety Treatments (2001). Clinical Guidelines for the Treatment of Depressive Disorders. *Can J Psychiatry* 46(Suppl 1): 5S–90S.
- Carmody T, Rush AJ, Bernstein I, Warden D, Brannan S, Burnham D et al (in press). The Montgomery Åsberg and the Hamilton Ratings of depression: a comparison of measures. *Eur Neuropsychopharmacol*.
- Chilvers C, Dewey M, Fielding K, Grettton V, Miller P, Palmer B et al (2001). Counseling versus antidepressants in Primary Care Study Group. Antidepressant drugs and generic counseling for treatment of major depression in primary care: randomised trial with patient preference arms. *BMJ* 322: 772–775.
- Cohen J (1983). The cost of dichotomization. *Appl Psychol Meas* 7: 249–253.
- Cook RJ, Sackett DL (1995). The number needed to treat: a clinically useful measure of treatment effect. *Br Med J* 310: 452–454.
- Crismon ML, Trivedi M, Pigott TA, Rush AJ, Hirschfeld RM, Kahn DA et al, Texas Consensus Conference Panel on Medication Treatment of Major Depressive Disorder (1999). The Texas Medication Algorithm Project: report of the Texas Consensus Conference Panel on Medication Treatment of Major Depressive Disorder. *J Clin Psychiatry* 60: 142–156.
- Daly JJ, Prudic J, Devanand DP, Nobler MS, Lisanby SH, Peyser S et al (2001). ECT in bipolar and unipolar depression: differences in speed of response. *Bipolar Disord* 3: 95–104.
- Dawson MY, Michalak EE, Waraich P, Anderson JE, Lam RW (2004). Is remission of depressive symptoms in primary care a realistic goal? A meta-analysis. *BMC Fam Pract* 5: 19.
- Depression Guideline Panel (1993). *Clinical Practice Guideline, Number 5: Depression in Primary Care: Volume 2. Treatment of Major Depression*. US Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, AHCPR Publication No. 93-0551: Rockville, MD.
- Dobrovolsky CL, Ivey FM, Rogers MA, Sorkin JD, Macko RF (2003). Reliability of treadmill exercise testing in older patients with chronic hemiparetic stroke. *Arch Phys Med Rehabil* 84: 1308–1312.
- Endicott J, Nee J, Harrison W, Blumenthal R (1993). Quality of life enjoyment and satisfaction questionnaire: a new measure. *Psychopharmacol Bull* 29: 321–326.
- Endicott J, Spitzer RL, Fleiss JL, Cohen J (1976). The global assessment scale: a procedure for measuring the overall severity of psychiatric disturbance. *Arch Gen Psychiatry* 33: 766–771.
- Entsuaeh AR, Huang H, Thase ME (2001). Response and remission rates in different subpopulations with major depressive disorder administered venlafaxine, selective serotonin reuptake inhibitors, or placebo. *J Clin Psychiatry* 62: 869–877.
- Ezquiaga E, Garcia A, Pallares T, Bravo MF (1999). Psychosocial predictors of outcome in major depression: a prospective 12-month study. *J Affect Disord* 52: 209–216.
- Fava GA, Fabbri S, Sonino N (2002). Residual symptoms in depression: an emerging therapeutic target. *Prog Neuropsychopharmacol Biol Psychiatry* 26: 1019–1027.
- Fava M, Evins AE, Dorer DJ, Schoenfeld DA (2003a). The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. *Psychother Psychosom* 72: 115–127.
- Fava M, Mallinckrodt CH, Detke MJ, Watkin JG, Wohlreich MM (2004). The effect of duloxetine on painful physical symptoms in depressed patients: do improvements in these symptoms result in higher remission rates? *J Clin Psychiatry* 65: 521–530.
- Fava M, Rush AJ, Trivedi MH, Nierenberg AA, Thase ME, Sackeim HA et al (2003b). Background and rationale for the sequenced treatment alternatives to relieve depression (STAR*D) study. *Psychiatr Clin North Am* 26: 457–494, x.
- Fava M, Uebelacker LA, Alpert JE, Nierenberg AA, Pava JA, Rosenbaum JF (1997). Major depressive subtypes and treatment response. *Biol Psychiatry* 42: 568–576.
- Fears TR, Ziegler RG, Donaldson JL, Falk RT, Hoover RN, Stanczyk FZ et al (2002). Reproducibility studies and interlaboratory concordance for androgen assays of male plasma hormone levels. *Cancer Epidemiol Biomarkers Prev* 11: 785–789.
- Frank E, Kupfer DJ, Gerebtzoff A, Meya U, Laghrissi-Thode F, Grochocinski VJ et al (2001). The development of study exit criteria for evaluating antimanic compounds. *J Clin Psychiatry* 62: 421–425.
- Frank E, Prien RF, Jarrett JB, Keller MB, Kupfer DJ, Lavori P et al (1991). Conceptualization and rationale for consensus definitions of terms in major depressive disorder: response, remission, recovery, relapse, and recurrence. *Arch Gen Psychiatry* 48: 851–855.
- Fritzler MJ, Wiik A, Tan EM, Smolen JS, McDougal JS, Chan EK et al (2003). A critical evaluation of enzyme immunoassay kits for detection of antinuclear autoantibodies of defined specificities. III. Comparative performance characteristics of academic and manufacturers' laboratories. *J Rheumatol* 30: 2374–2381.
- Ghatavi K, Nicolson R, MacDonald C, Osher S, Levitt A (2002). Defining guilt in depression: a comparison of subjects with major depression, chronic medical illness and healthy controls. *J Affect Disord* 68: 307–315.
- Goldstein DJ, Lu Y, Detke MJ, Wiltse C, Mallinckrodt C, Demitrack MA (2004). Duloxetine in the treatment of depression: a double-blind placebo-controlled comparison with paroxetine. *J Clin Psychopharmacol* 24: 389–399.
- Hamilton M (1960). A rating scale for depression. *J Neurol Neurosurg Psychiatry* 23: 56–62.
- Hamilton M (1967). Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol* 6: 278–296.
- Hirschfeld RM, Dunner DL, Keitner G, Klein DN, Koran LM, Kornstein SG et al (2002). Does psychosocial functioning improve independent of depressive symptoms? A comparison of nefazodone, psychotherapy, and their combination. *Biol Psychiatry* 51: 123–133.
- Hirschfeld RM, Keller MB, Panico S, Arons BS, Barlow D, Davidoff F et al (1997). The National Depressive and Manic-Depressive Association consensus statement on the undertreatment of depression. *JAMA* 277: 333–340.
- Hollon SD, Kendall PC (1980). Cognitive self-statements in depression: development of an automatic thoughts questionnaire. *Cognitive Ther Res* 4: 383–397.
- Husain MM, Rush AJ, Fink M, Knapp R, Petrides G, Rummans T et al (2004). Speed of response and remission in major

- depressive disorder with acute electroconvulsive therapy (ECT): a consortium for research in ECT (CORE) report. *J Clin Psychiatry* 65: 485–491.
- Iosifescu DV, Nierenberg AA, Alpert JE, Smith M, Bitran S, Dording C et al (2003). The impact of medical comorbidity on acute treatment in major depressive disorder. *Am J Psychiatry* 160: 2122–2127.
- Judd LL, Akiskal HS, Maser JD, Zeller PJ, Endicott J, Coryell W et al (1998a). A prospective 12-year study of subsyndromal and syndromal depressive symptoms in unipolar major depressive disorders. *Arch Gen Psychiatry* 55: 694–700.
- Judd LL, Akiskal HS, Maser JD, Zeller PJ, Endicott J, Coryell W et al (1998b). Major depressive disorder: a prospective study of residual subthreshold depressive symptoms as predictor of rapid relapse. *J Affect Disord* 50: 97–108.
- Judd LL, Paulus MP, Zeller P (1999). The role of residual subthreshold depressive symptoms in early episode relapse in unipolar major depressive disorder. *Arch Gen Psychiatry* 56: 764–765.
- Kanai T, Takeuchi H, Furukawa TA, Yoshimura R, Imaizumi T, Kitamura T et al (2003). Time to recurrence after recovery from major depressive episodes and its predictors. *Psychol Med* 33: 839–845.
- Keitner GI, Ryan CE, Miller IW, Kohn R, Epstein NB (1991). 12-month outcome of patients with major depression and comorbid psychiatric or medical illness (compound depression). *Am J Psychiatry* 148: 345–350.
- Keitner GI, Ryan CE, Miller IW, Norman WH (1992). Recovery and major depression. factors associated with twelve-month outcome. *Am J Psychiatry* 149: 93–99.
- Keller MB, Gelenberg AJ, Hirschfeld RM, Rush AJ, Thase ME, Kocsis JH et al (1998). The treatment of chronic depression part 2. a double-blind randomized trial of sertraline and imipramine. *J Clin Psychiatry* 59: 598–607.
- Keller MB, McCullough JP, Klein DN, Arnow B, Dunner DL, Gelenberg AJ et al (2000). A comparison of nefazodone the cognitive behavioral-analysis system of psychotherapy and their combination for the treatment of chronic depression. *N Engl J Med* 342: 1462–1470.
- Kendler KS (1996). Major depression and generalized anxiety disorder. Same genes (partly) different environments—revisited. *Br J Psychiatry Suppl* 30: 68–75.
- Klerman GL, Weissman MM, Rounsaville BJ, Chevron RS (1984). *Interpersonal Psychotherapy of Depression*. Basic Books: New York.
- Koran LM, Gelenberg AJ, Kornstein SG, Howland RH, Friedman RA, DeBattista C et al (2001). Sertraline versus imipramine to prevent relapse in chronic depression. *J Affect Disord* 65: 27–36.
- Kraemer HC, Wilson GT, Fairburn CG, Agras WS (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Arch Gen Psychiatry* 59: 877–883.
- Kroenke K, Spitzer RL, Williams JB (2001). The PHQ-9. Validity of a brief depression severity measure. *J Gen Intern Med* 16: 606–613.
- Lepola UM, Wade AG, Leinonen EV, Koponen HJ, Frazer J, Sjodin I et al (1998). A controlled, prospective, 1-year trial of citalopram in the treatment of panic disorder. *J Clin Psychiatry* 59: 528–534.
- Miller IW, Keitner GI, Schatzberg AJ, Klein DN, Thase ME, Rush AJ et al (1998). The treatment of chronic depression. Part 3: Psychosocial functioning before and after treatment with sertraline or imipramine. *J Clin Psychiatry* 59: 608–619.
- Mintz J, Mintz LI, Arruda MJ, Hwang SS (1992). Treatments of depression and the functional capacity to work (published erratum appears in *Arch Gen Psychiatry* 1993 Mar; 50(3): 241). *Arch Gen Psychiatry* 49: 761–768.
- Montgomery SA, Åsberg M (1979). A new depression scale designed to be sensitive to change. *Br J Psychiatry* 134: 382–389.
- Montgomery SA, Huusom AK, Bothmer J (2004). A randomised study comparing escitalopram with venlafaxine XR in primary care patients with major depressive disorder. *Neuropsychobiology* 50: 57–64.
- Nelson JC, Mazure CM, Jatlow PI, Bowers Jr MB, Price LH (2004). Combining norepinephrine and serotonin reuptake inhibition mechanisms for treatment of depression: a double-blind, randomized study. *Biol Psychiatry* 55: 296–300.
- Nierenberg AA, Feighner JP, Rudolph R, Cole JO, Sullivan J (1994). Venlafaxine for treatment-resistant unipolar depression. *J Clin Psychopharmacol* 14: 419–423.
- Nierenberg AA, Keefe BR, Leslie VC, Alpert JE, Pava JA, Worthington III JJ et al (1999). Residual symptoms in depressed patients who respond acutely to fluoxetine. *J Clin Psychiatry* 60: 221–225.
- Nierenberg AA, McLean NE, Alpert JE, Worthington JJ, Rosenbaum JF, Fava M (1995). Early nonresponse to fluoxetine as a predictor of poor 8-week outcome. *Am J Psychiatry* 152: 1500–1503.
- O’Leary D, Costello F, Gormley N, Webb M (2000). Remission onset and relapse in depression. An 18-month prospective study of course for 100 first admission patients. *J Affect Disord* 57: 159–171.
- Ormel J, Oldehinkel AJ, Nolen WA, Vollebergh W (2004). Psychosocial disability before, during, and after a major depressive episode: a 3-wave population-based study of state, scar, and trait effects. *Arch Gen Psychiatry* 61: 387–392.
- Paykel ES, Ramana R, Cooper Z, Hayhurst H, Kerr J, Barocka A (1995). Residual symptoms after partial remission. an important outcome in depression. *Psychol Med* 25: 1171–1180.
- Paykel ES, Scott J, Teasdale JD, Johnson AL, Garland A, Moore R et al (1999). Prevention of relapse in residual depression by cognitive therapy. A controlled trial. *Arch Gen Psychiatry* 56: 829–835.
- Perlis RH, Iosifescu DV, Alpert J, Nierenberg AA, Rosenbaum JF, Fava M (2004). Effect of medical comorbidity on response to fluoxetine augmentation or dose increase in outpatients with treatment-resistant depression. *Psychosomatics* 45: 224–229.
- Philipp M, Fickinger MP (1993). The definition of remission and its impact on the length of a depressive episode. *Arch Gen Psychiatry* 50: 407–408.
- President’s New Freedom Commission on Mental Health (2003). *Achieving the Promise: Transforming Mental Health Care in America, Final Report* (Pub. No. SMA-03-3832). DHHS: Rockville, MD.
- Prien RF, Kupfer DJ (1986). Continuation drug therapy for major depressive episodes. How long should it be maintained? *Am J Psychiatry* 143: 18–23.
- Prudic J, Haskett RF, Mulsant B, Malone KM, Pettinati HM, Stephens S et al (1996). Resistance to antidepressant medications and short-term clinical response to ECT. *Am J Psychiatry* 153: 985–992.
- Prudic J, Olfson M, Marcus SC, Fuller RB, Sackeim HA (2004). Effectiveness of electroconvulsive therapy in community settings. *Biol Psychiatry* 55: 301–312.
- Prudic J, Sackeim HA, Devanand DP (1990). Medication resistance and clinical response to electroconvulsive therapy. *Psychiatry Res* 31: 287–296.
- Quitkin FM, Petkova E, McGrath PJ, Taylor B, Beasley C, Stewart J et al (2003). When should a trial of fluoxetine for major depression be declared failed? *Am J Psychiatry* 160: 734–740.
- Reesal RT, Lam RW, CANMAT Depression Work Group (2001). Clinical guidelines for the treatment of depressive disorders. II. Principles of management. *Can J Psychiatry* 46(Suppl 1): 21S–28S.
- Reimherr FW, Amsterdam JD, Quitkin FM, Rosenbaum JF, Fava M, Zajecka J et al (1998). Optimal length of continuation therapy

- in depression: a prospective assessment during long-term fluoxetine treatment. *Am J Psychiatry* 155: 1247–1253.
- Riso LP, Thase ME, Howland RH, Friedman ES, Simons AD, Tu XM (1997). A prospective test of criteria for response remission relapse recovery and recurrence in patients treated with cognitive behavior therapy. *J Affect Disord* 43: 131–142.
- Rush AJ, Bernstein IH, Trivedi MH, Carmody TJ, Wisniewski S, Mundt JC et al (2006). 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* 59: 493–501 (Epub 2005 Sep 30).
- Rush AJ, Crismon ML, Kashner TM, Toprac MG, Carmody TJ, Trivedi MH et al (2003a). Texas Medication Algorithm Project, phase 3 (TMAP-3): rationale and study design. *J Clin Psychiatry* 64: 357–369.
- Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi MH, Sackeim HA et al (2004). Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Control Clin Trials* 25: 119–142.
- Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH (1996). The Inventory of Depressive Symptomatology (IDS). Psychometric properties. *Psychol Med* 26: 477–486.
- Rush AJ, Ryan ND (2002). Current and emerging therapeutics for depression. In: Davis KL, Charney D, Coyle JT, Nemeroff C (eds). *Neuropsychopharmacology. The Fifth Generation of Progress*. Lippincott Williams & Wilkins: Philadelphia. pp 1081–1095.
- Rush AJ, Sackeim HA, Marangell LB, George MS, Brannan SK, Davis SM et al (2005a). Effects of 12 months of vagus nerve stimulation in treatment-resistant depression: a naturalistic study. *Biol Psychiatry* 58: 355–363.
- Rush AJ, Thase ME, Dube S (2003b). Research issues in the study of difficult-to-treat depression. *Biol Psychiatr* 53: 743–753.
- Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN et al (2003c). The 16-item Quick Inventory of Depressive Symptomatology (QIDS) Clinician Rating (QIDS-C) and Self-Report (QIDS-SR). A psychometric evaluation in patients with chronic major depression. *Biol Psychiatry* 54: 573–583.
- Rush AJ, Zimmerman M, Wisniewski SR, Fava M, Hollon SD, Warden D et al (2005b). Comorbid psychiatric disorders in depressed outpatients: Demographic and clinical features. *J Affect Disord* 87: 43–55.
- Sackeim HA, Decina P, Prohovnik I, Malitz S (1987). Seizure threshold in electroconvulsive therapy. Effects of sex, age, electrode placement, and number of treatments. *Arch Gen Psychiatry* 44: 355–360.
- Sackeim HA, Haskett RF, Mulsant BH, Thase ME, Mann JJ, Pettinati HM et al (2001). Continuation pharmacotherapy in the prevention of relapse following electroconvulsive therapy: a randomized controlled trial. *JAMA* 285: 1299–1307.
- Sackeim HA, Prudic J, Devanand DP, Decina P, Kerr B, Malitz S (1990). The impact of medication resistance and continuation pharmacotherapy on relapse following response to electroconvulsive therapy in major depression. *J Clin Psychopharmacol* 10: 96–104.
- Sackeim HA, Prudic J, Devanand DP, Kiersky JE, Fitzsimons L, Moody BJ et al (1993). Effects of stimulus intensity and electrode placement on the efficacy and cognitive effects of electroconvulsive therapy (see comments). *N Engl J Med* 328: 839–846.
- Sackeim HA, Prudic J, Devanand DP, Nobler MS, Lisanby SH, Peyser S et al (2000). A prospective, randomized, double-blind comparison of bilateral and right unilateral electroconvulsive therapy at different stimulus intensities. *Arch Gen Psychiatry* 57: 425–434.
- Sayer NA, Sackeim HA, Moeller JR, Prudic J, Devanand DP, Coleman EA et al (1993). The relations between observer-rating and self-report of depressive symptomatology. *Psychol Assess* 5: 350–360.
- Simon GE, Manning WG, Katzelnick DJ, Pearson SD, Henk HJ, Helstad CS (2001). Cost-effectiveness of systematic depression treatment for high utilizers of general medical care. *Arch Gen Psychiatry* 58: 181–187.
- Small GW, Birkett M, Meyers BS, Koran LM, Bystritsky A, Nemeroff CB (1996). Impact of physical illness on quality of life and antidepressant response in geriatric major depression. Fluoxetine Collaborative Study Group. *J Am Geriatr Soc* 44: 1220–1225.
- Tedlow J, Fava M, Uebelacker L, Nierenberg AA, Alpert JE, Rosenbaum J (1998). Outcome definitions and predictors in depression. *Psychother Psychosom* 67: 266–270.
- Thase ME, Entsuah AR, Rudolph RL (2001). Remission rates during treatment with venlafaxine or selective serotonin reuptake inhibitors. *Br J Psychiatry* 178: 234–241.
- Thase ME, Simons AD, McGeary J, Cahalane JF, Hughes C, Harden T et al (1992). Relapse after cognitive behavior therapy of depression. potential implications for longer courses of treatment. *Am J Psychiatry* 149: 1046–1052.
- Trivedi MH, DeBattista C, Fawcett J, Nelson C, Osser DN, Stein D et al (1998). Developing treatment algorithms for unipolar depression in cyberspace. International Psychopharmacology Algorithm Project (IPAP). *Psychopharmacol Bull* 34: 355–359.
- Trivedi MH, Rush AJ, Crismon ML, Kashner TM, Toprac MG, Carmody TJ et al (2004a). Clinical results for patients with major depressive disorder in the Texas Medication Algorithm Project. *Arch Gen Psychiatry* 61: 669–680.
- Trivedi MH, Rush AJ, Ibrahim HM, Carmody TJ, Biggs MM, Suppes T et al (2004b). 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* 34: 73–82.
- Trivedi MH, Rush AJ, Pan JY, Carmody TJ (2001). Which depressed patients respond to nefazodone and when? *J Clin Psychiatry* 62: 158–163.
- Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L et al, for the STAR*D Study Team (2006). Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry* 163: 28–40.
- Viinamaki H, Hintikka J, Tanskanen A, Honkalampi K, Antikainen R, Koivumaa-Honkanen H et al (2002). Partial remission in major depression. a two-phase, 12-month prospective study. *Nord J Psychiatry* 56: 33–37.
- Ware Jr JE, Sherbourne CD (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 30: 473–483.
- Weissman AN (1979). The dysfunctional attitude scale. A validation study. *Diss Abst Int* 40: 1389B–1390B.
- Wilson JF, Watson ID, Williams J, Toseland PA, Thomson AH, Sweeney G et al (2002). Primary standardization of assays for anticonvulsant drugs: comparison of accuracy and precision. *Clin Chem* 48: 1963–1969.
- Zimmerman M, Chelminski I, Posternak M (2004a). A review of studies of the Montgomery-Åsberg Depression Rating Scale in controls: implications for the definition of remission in treatment studies of depression. *Int Clin Psychopharmacol* 19: 1–7.
- Zimmerman M, Posternak MA, Chelminski I (2004b). Defining remission on the Montgomery-Åsberg depression rating scale. *J Clin Psychiatry* 65: 163–168.
- Zimmerman M, Posternak MA, Chelminski I (2004c). Implications of using different cut-offs on symptom severity scales to define remission from depression. *Int Clin Psychopharmacol* 19: 215–220.