Methodological Issues in Developing New Acute Treatments for Patients with Bipolar Illness

A. John Rush, Robert M. Post, Willem A. Nolen, Paul E. Keck, Jr., Trisha Suppes, Lori Altshuler, and Susan L. McElroy

One important aim of the recent reorganization of the National Institute of Mental Health (NIMH) is to streamline the development of new treatments for patients with severe mental illnesses, such as bipolar disorder. Researching new treatments for patients with bipolar disorder presents specific problems not readily addressed by traditional efficacy trial methodologies that aim to maximize internal validity. This article reexamines several assumptions that have guided the design of these efficacy trials but that also create obstacles for studies of bipolar disorder and suggests potential solutions. This article draws on literature from neurology and psychiatry and discussions at a MacArthur Foundation–sponsored Conference on Longitudinal Methodology in 1992 (David J. Kupfer, M.D., Chair), which brought together investigators to consider alternative designs for patients with severe and persistent mental illness. In addition, we benefited from discussions at two NIMH-sponsored conferences, one held in 1989 (Prien and Potter 1990) and the other in 1994 (Prien and Rush 1996), at which investigators and methodologists discussed issues surrounding the development and conduct of informative efficacy trials for patients with bipolar disorder. Based on these discussions and recent literature reviews, we 1) outline common problems in the development and evaluation of effective acute treatments for bipolar disorder and 2) suggest possible solutions to these impediments. We also discuss alternative designs by which to build a sequence of acute treatment studies from which efficacy, safety, and the comparative value of different treatments can be established. Biol Psychiatry 2000;48:615–624 © 2000 Society of Biological Psychiatry

Key Words: Acute phase treatment, bipolar illness, methodological issues

Defining the Impediments

Studies of new acute phase treatments for bipolar disorder designed to assess efficacy, safety, and tolerability have largely been modeled after traditional methodologies used to develop and evaluate acute phase treatments for patients with nonpsychotic major depressive disorder. This approach usually entails the identification of drug-free, moderately symptomatic, acutely nonsuicidal outpatients with moderate levels of disability. These patients provide informed consent to participate in a randomized, double-blind, acute phase comparison trial of the new treatment to a standard control treatment, pill placebo, or both for 6 to 12 weeks. Clinical outcomes typically include symptom severity (assessed by itemized symptom checklists) and adverse effects, each obtained at frequent intervals.

Limitations of the Classic Acute Phase Efficacy Trial

The traditional approach to evaluating the efficacy and safety of new medications is often extremely difficult to implement for bipolar disorder. Even if successful, it may still preclude informative generalizability, depending on the restrictiveness of inclusion or breadth of exclusion criteria (Licht et al 1997). For example, a recent study of the treatment of acute manic episodes, engaged and randomly assigned drug-free inpatients with bipolar disorder, manic phase, to divalproex sodium, lithium, or pill placebo (Bowden et al 1994). This study, although highly informative about the acute phase efficacy of divalproex, did not answer the clinically important question of where divalproex alone or combined with other treatment(s) fits in the management of patients with bipolar disorder during the acute phase and especially in subsequent longer term treatment. This carefully designed initial trial had high internal validity but limited generalizability.

In addition, the Bowden et al (1994) study would
probably not be feasible in today’s managed health care climate where hospitalization is curtailed—even for patients with mania. In this context, patient safety is of paramount importance. If patients receive placebo and have a reasonably good initial response while in the hospital but are subsequently discharged on placebo only, their relapse would occur as outpatients—which increases risk.

Unlike major depressive disorder, most patients with bipolar disorder will likely need a combination of medications to achieve euthymia (Freeman and Stoll 1998; Post et al 1998a; Solomon et al 1996). Thus, to determine when a combination of mood stabilizers, with or without atypical antipsychotic agents, is required—and therefore where a new therapy fits (not only as a monotherapy but also as a part of combination therapy)—is a key clinical issue typically not addressed by traditional efficacy trials in which monotherapy is evaluated against placebo. The question of where a treatment fits is important because only about half of manic patients taking either lithium or divalproex sodium respond (≥50% improvement) after 3 weeks (Bowden et al 1994; Keck and McElroy 1996). That is, 50% improvement from baseline may still leave the patient symptomatic. Moreover, as emphasized at the 1992 MacArthur Conference on Longitudinal Methodology, these designs are not clinician friendly because they do not answer the question in most need of an answer, namely, if drug A does not work in a given patient, then what are the chances that drug B will?

The following discussion attempts to address several of these current impediments by reconceptualizing approaches to the acute phase efficacy–effectiveness studies of new treatments for bipolar disorder. The discussion is divided into three sections: 1) a reevaluation of sample definition and of study participants, 2) a commentary on the utility of developing consensus-based outcome measures and a time period for outcome sampling, and finally 3) borrowing substantially from the neurology literature, suggestions for alternative study designs that, when taken together, provide a matrix for rapidly developing efficacy data from representative patients while protecting patient safety. These suggestions are offered with the hope that a consensus among investigators evaluating new acute phase treatments for bipolar disorder can be developed to more rapidly evaluate treatments of potential benefits to patients.

**Sample Composition**

Several differences distinguish bipolar disorders from major depressive disorders and impact the composition of patient samples and the nature of outcomes to be measured. Given issues of patient safety, these differences may well recommend designs other than the traditional randomized, double-blind, acute phase, placebo-controlled study of moderately ill outpatients. Among the differences between the disorders are the presence of concurrent conditions, which are common in adults with major depression but are even more frequent in those patients with bipolar disorder. In addition, two types of comorbidity (lifetime and concurrent) occur in those with bipolar disorder. They have a 10 to 20 times higher risk of having three or more lifetime comorbid conditions than any other psychiatric illness (Keck et al 1997; Kessler et al 1994; McElroy et al, in press; Strakowski et al 1998). Given this high rate of concurrent comorbidity, both generalizability and feasibility are reduced, and these patients would be excluded if traditional restrictive criteria were applied.

Patients with major depression who are psychotic, actively suicidal, and judged to be at risk are typically excluded from placebo-controlled efficacy trials; however, patients with bipolar disorder often become suicidal more rapidly and more unpredictably, especially given the high prevalence of psychotic symptoms in this population.

**Outcome Measures**

Turning to the issue of meaningful, agreed-on outcome measures, efficacy studies in major depression generally use itemized depressive symptom severity rating scales completed by clinicians, such as the Hamilton Rating Scale for Depression (HRS-D; Hamilton 1960, 1967), the Montgomery–Åsberg Depression Rating Scale (MADRS; Montgomery and Åsberg 1979), or the Inventory of Depressive Symptomatology—Clinician Rated (IDS-C; Rush et al 1986, 1996) as primary outcome measures. For confirmation or elaboration of treatment effects, symptom self-reports such as the Beck Depression Inventory (BDI; Beck et al 1961, 1979), the Zung Depression Rating Scale (ZDRS; Zung 1965), or the IDS-SR (Rush et al 1986, 1996) are often employed.

For bipolar disorder, the clinical instruments used to assess symptom severity have been borrowed from or modeled after those used for major depression. For example, the depressed phase of the illness is often assessed by a depressive symptom clinical rating (e.g., the HRS-D), whereas the manic phase of illness is often assessed by a similarly designed rating instrument, such as the Young Mania Rating Scale (YMRS; Young et al 1978).

A number of difficulties are encountered when using these simple symptom checklists in bipolar disorder, however. Reliance on patients’ reports of their symptoms and assessments for brief time periods (e.g., 7 days before the rating) carry key limitations, including the following: 1) the different illness phases (depressed, hypomanic,
manic, and mixed) may all occur between measurement occasions; 2) the co-occurrence of manic and depressive symptoms may not be fully captured by symptom assessments; 3) the different illness phases often are associated with psychosis, which can impair ability to identify symptoms; 4) moreover, many patients, when manic or hypomanic, underestimate pathology, rendering self-reports insufficient or misleading; and 5) finally, in rapidly-cycling patients, even monthly cross-sectional symptomatic assessments may miss many periods of severe depression or mania that can be captured by other, more frequent, prospective daily assessments (Denicoff et al 1997b).

Another outcome measurement issue, given the waxing and waning course of bipolar illness, is how to define response, especially because the morbid state has three different and sometimes even apparently opposite elements: depressive symptoms, manic symptoms, and mood lability (i.e., the tendency to shift rapidly from one state to another with an almost infinite variety of accompanying patterns). Additionally, patients with bipolar disorder may experience rapid shifts from euthymic or nearly asymptomatic states into more severe depressive or manic states (even a psychotic condition). Switches in mood states may happen so rapidly that scheduled appointments cannot sufficiently capture acute symptom worsening or improvement.

**Rapid Cycling**

A related difference between major depressive and bipolar disorders is the frequent cycling between illness phases (i.e., from manic to mixed, from manic to depressed, etc.), often referred to as “rapid cycling” (i.e., four or more mood episodes per year). Moreover, rapid cycling may be so rapid (i.e., ultra rapid and ultradian forms; George et al 1998; Kramerling and Post 1996; Pazzaglia et al 1993) that at a particular time, the patient may not qualify for studies in which a minimum duration of an episode (e.g., 2 weeks) or no clinical change during an observational period (e.g., placebo run-in or “washout”) are among the study inclusion criteria.

**Concomitant Medications**

An additional, important distinction between major depressive and bipolar disorders is the use of concomitant medications. Although difficult, it is possible to find samples of depressed patients receiving no specific antidepressant treatment; however, recruitment of medication-free patients with bipolar disorder is extremely difficult.

A central problem is what to do with patients receiving concomitant medications. Is discontinuation of these agents during a washout period appropriate? What ethical and safety issues are involved in such a procedure? What if the medication is a mood stabilizer, yet the overall response has been incomplete? Is the patient excluded? If patients undergo a washout, many will experience a worsening of symptoms or develop an episode, thus becoming an acute clinical and study management problem (i.e., requiring extra care for safety reasons or becoming study ineligible). If, on the other hand, patients requiring concomitant agents are excluded, both generalizability and feasibility are increasingly compromised.

It has been suggested that a partial response to a particular medication should not lead to its discontinuation, but rather (as is also common clinical practice) to continue the treatment and add a second agent in an attempt to convert a partial response to a complete response (i.e., augmentation; for reviews, see Post et al 1997a; Woggon 1997). Is there any way to engage such patients to assist in establishing efficacy and safety, informing us as to where the new agent might fit in the treatment plan?

**Sample Issues**

Concurrent comorbidity, variations in the course of illness, and different variants (e.g., bipolar I, bipolar II, and bipolar disorder, not otherwise specified [NOS]) of the disorder, as well as the high probability that most otherwise study-eligible patients are already taking one or more “therapeutic” agents, all create substantial impediments to the feasibility of acquiring a “clean” sample of patients. Further, the tendency to shift illness phases (depressed, manic, or mixed) with or without treatment, complicates both study design and sample definition.

Consensus opinions (Prien and Rush 1996) recommend broader inclusion and fewer exclusion criteria for patients with bipolar disorder who enroll in trials of safety, tolerability, and efficacy of new mood stabilizing, antimanic, or antidepressant medications. The reasons for this recommendation include the following: 1) generalizability of findings, whether negative or positive, is nearly immediate because populations under study are basically identical to those in routine practice; 2) by including patients with greater comorbidity, greater mood lability, or both, placebo response rates are likely reduced, so that sample sizes needed to identify an effective treatment should be smaller than with less complex patient populations; and 3) feasibility is increased, and cost and time to develop useful clinical information are reduced.

On the other hand, the inclusion of more complex populations in clinical trials potentially threatens internal validity. Randomization may fail in that a disproportionate representation of one as opposed to another type of condition enters...
one treatment cell. For example, patients with a rapid cycling course—who may be less responsive to drug A, but more responsive to drug B—may, by chance, be assigned disproportionately to the cell containing drug A. This may in turn render drug A less effective than it normally would be. The standard solution to this problem is to stratify based on parameters that are well known to affect outcome. Unfortunately, with bipolar disorder, parameters that actually affect outcome are not yet completely known; however, concurrent substance abuse and rapid cycling have been suggested by numerous studies (Aagaard and Vestergaard 1990; Calabrese and Woyshville 1995a; Denicoff et al 1997a; Himmelhoch and Garfinkel 1986; Kusalic and Engelsmann 1998; Post et al 1996a; Swann et al 1999) to affect the acute phase response to lithium. Thus, stratification for each of these parameters, if present at baseline, can allow inclusion of patients with rapid cycling or with active substance abuse. Nonetheless, one must limit the number of stratifying variables as not to substantially reduce feasibility while retaining generalizability.

Another limitation of restricted samples that are less representative is that such patients may preferentially benefit from the new treatment because their illness may be more responsive to treatment overall. Specifically, this strategy leads to a risk that an effective treatment will be found for only a minority (10–25%) of patients and that this treatment will not meaningfully add to what is already available for more representative patients (i.e., a drug with only modest “real-world” utility is inappropriately recommended because of inadequate study design; Licht et al 1997).

Thus, the dilemma encountered in sample selection is whether to study an unrepresentative and difficult-to-obtain sample in an attempt to avoid the potentially confounding affects of variables that may not randomize equally to each treatment cell or whether to “roll the dice,” allowing patients with more complex but representative illnesses into the trial, gambling on the likelihood that randomization will succeed.

To develop new acute phase treatments with real-world generalizability, inclusiveness is preferred by allowing in most concurrent conditions, with the exception of active substance abuse, which almost always requires independent concurrent treatment. Patients with past histories of substance abuse not currently abusing are, of course, at risk for recidivism, even in the course of a new treatment trial. Using this approach, those patients randomized to different treatment cells are assumed to be equally likely to have a recurrence of their substance abuse.

In addition, because the treatment implications (i.e., all of these patients are likely to require mood stabilizers) have not been established for the division between types of bipolar disorder (bipolar I vs. bipolar II), including both diagnostic groups in efficacy trials can be justified. Because rapid cycling appears to differentially affect outcome, stratification for this single variable can be planned, or this group can be completely excluded. The advantage of a more inclusive, representative sample is especially relevant to maintenance trials designed to determine whether acute phase responders continue to have sustained benefit. Those patients with greater comorbidity, greater course fluctuations, or both are those most likely to display early relapse or recurrence. Narrowly defined or restricted populations that respond to acute phase treatment and then enter a longer term “maintenance” trial are less likely to have sufficient symptomatology or disability within a reasonable period of time to identify a clinically significant relapse or recurrence. This may have been a reason for the failure to distinguish valproate or lithium from placebo in the prevention of mania (Bowden et al 2000). To summarize, the need to generalize results from efficacy trials to clinical practice argue for broader inclusion and fewer exclusion criteria. The following lists the agreed-upon “rules” at the NIMH-sponsored 1994 conference on bipolar illness:

1. Broaden inclusion criteria to include rapid cycling and other variants of bipolar disorder.
2. Inclusion of comorbidities.
3. Inclusion of bipolar I, bipolar II, and bipolar NOS disorders.
4. Develop a widely acceptable standardized outcome assessment rating package.
5. Choose an acceptable longitudinal rating tool.
6. Broaden the list of acceptable designs (e.g., crossover) that are clinician friendly.
7. Consider conducting randomized trials in the clinical setting in an open fashion.
8. Adopt more liberal criteria for entry into the randomization phase, so that the most ill patients are not excluded.
9. Recognize the unique difficulties of studies in bipolar patients and fund studies that answer some, if not all, questions raised.
10. Develop new methods to study comparative complex combination therapies.

Anecdotal reports from multiple applicants, even recently, suggest that NIMH review committees should be familiarized with these NIMH rules, promulgated over 6 years ago, to include a major emphasis on inclusion criteria, as well as other recommendations that could lead to reviews of grant applications for the treatment studies of bipolar illness that are consistent with these guidelines.
Assessment of Outcome

It appears to be insufficient to borrow methodologies from efficacy studies in major depressive disorder to define the outcome of patients with bipolar disorder because of the diversity of symptomatology in the latter illness and the presence of mood instability (especially rapid cycling). For example, in major depressive disorder, the standard frequency of measurement (e.g., every week or two) and the typical time frame assessed (e.g., the past 7 days) may be inadequate or even misleading for patients with bipolar disorder. More frequent assessments or observations that cover more extended time periods, provide a more thorough method by which to evaluate treatment benefits. Therefore, alternative strategies to measure outcome in efficacy trials of patients with bipolar disorder are recommended.

Another impediment is a lack of a consensus regarding the appropriate outcome measures for bipolar disorder. Although there are a number of itemized symptom rating scales available (American Psychiatric Association Task Force for the Handbook of Psychiatric Measures 2000), they do not capture the qualitative and quantitative changes that rapidly occur or the level of disability associated with bipolar disorder. For example, when patients present in a mixed episode, is it necessary to complete both mania and depression rating scales? Should this also occur when patients are in a single phase of the illness? Is severity the sum of both ratings? Additionally, how can ratings of both mania and depression be compared?

Recent innovations using a life charting method (LCM; Denicoff et al 1997b; Leverich and Post 1996, 1998), and the recent redesign of the Clinical Global Impression (CGI; Guy 1976) for bipolar disorder (CGI-BP; Spearing et al 1997) provide two methods by which to gauge the type (manic, hypomanic, mixed, or depressed), severity, and associated functional disability in bipolar disorder over time periods that range from 1 day to weeks, or even months to years. These methods also can evaluate acute phase response, as well as prophylactic efficacy. The LCM summarizes the severity and functional effect of the illness without diagnostic bias as to what constitutes an “episode” of mania or depression.

The LCM captures mood state fluctuations and the degree of symptom-driven dysfunction either retrospectively (e.g., on a monthly basis) or prospectively (e.g., on a daily basis). The prospective LCM has been found to be a relatively reliable and valid method when sampling daily symptomatology over the previous 2 to 4 weeks (Denicoff et al 1997b). Specifically, the LCM correlated with the HRS-D (at 7 days), the YMRS (at 2 days), and the Global Assessment Scale (Endicott et al 1976, at 1 month).

Alternative Study Designs

Another major obstacle encountered when studying bipolar disorder is treatment trial design. The standard parallel group, randomized, double-blind, placebo-controlled trial, when used in patients with bipolar disorder who are often severely ill, can lead to several problems, including the following: 1) patients or their treating physicians decline studies because of concerns about receiving placebo; 2) investigators or institutions limit eligible patient populations to the mildly to moderately ill to increase patient safety, and 3) investigators, patients, and their families worry about untoward outcomes for an increased number of patients as some are receiving placebo. In the case of bipolar disorder, safety concerns (for self or others) can present abruptly and, although uncommon, are not rare.

An analogous set of problems has been encountered in the research on epilepsy. The neurology literature recommends several alternative study designs for evaluating the safety, efficacy, and tolerability of new medications, all of which appear well suited to treatment studies of patients with bipolar disorder (Pledger and Sahlroot 1993). Specifically, the risk of stopping partially effective mood stabilizing, antimanic, or antipsychotic agents can be as profound as the risk of stopping antiepileptic agents in some patients, especially given suggestions of “kindling-like” phenomena apparently present in both disorders. Thus, “add-on” designs similar to those used in epilepsy research may have particular utility in acute phase treatment studies in a large proportion of persons with bipolar disorder.

Add-On Placebo-Controlled Trials

In the treatment of epilepsy, as is case for bipolar disorder, multiple medications often are used to achieve complete symptom control, although monotherapy often is initially attempted. In an add-on placebo-controlled study, patients who remain symptomatic, while continuing to receive their current medication(s) at fixed dose(s), are randomized, double-blind, to either placebo or the new treatment (Handforth et al 1998; Harden et al 1999; Mawer et al 1999). This design has the advantage of being inclusive with regard to eligible patients, except that patients with bipolar disorder who respond fully to monotherapy will not be eligible for this design. This add-on design requires that some knowledge of drug interactions is available, but it also allows for the prospective evaluation of drug interactions in more representative samples.

This add-on design also benefits patient safety, a major concern, especially with the evaluation of placebo against new treatments that often have, at the particular time of study, no demonstrated efficacy. Patients entering an add-on study are not disadvantaged over their baseline
state because all patients continue to receive the medication(s) that might have been partially helpful. A subset of them have a potential added advantage (or disadvantage) of exposure to the new medication.

A variation of this add-on design (Sachdeo et al. 1997) incorporates a baseline period to determine symptom severity and consistency and, consequently, whether patients have sufficient symptomatology to enter the trial. Once inclusion and exclusion criteria are met, the experimental medication is given open label for a short period of time (e.g., 1 week) in addition to the patient’s prestudy medication(s) at fixed doses. During the open-label period, a lower than presumably therapeutic dose of the new medication is used to probe for drug interactions and tolerability (i.e., to identify drug allergies, drug intolerance, and negative drug interactions). Thus, patients who cannot tolerate modest doses of the new medication being added to their ongoing medication(s) are not randomized.

At the end of the open-label period, medications are masked. What follows is a specified period (e.g., 5 weeks) in which patients are randomized to receive either a placebo or some experimental medication used at a low dose during the open-label period, or a higher, presumably effective, dose of the new medication. As an alternative, patients may not be assigned to placebo, but rather continue on the lower dose of the experimental medication originally prescribed. During this randomized phase, prestudy medications are continued in a stable regimen.

Following the randomization phase, the prestudy medication(s) may be discontinued to compare monotherapy with the new medication to prestudy medications combined with placebo (or low dose experimental medication; Figure 1).

Thus, for bipolar disorder, a prospective baseline evaluation (e.g., 1 to 2 weeks) for patients on an established regimen is followed by open-label exposure to the new drug at low doses to determine tolerability, adverse events, and drug interactions. Thereafter, while prestudy medications continue, a double-blind randomization to placebo (or to a low dose of the new medication) or to the new medication at therapeutic doses follows. This phase evaluates the utility of the new medication as an adjunct to whatever baseline treatment(s) patients are receiving. After the randomization phase, a slow discontinuation of the original regimen could follow, with random assignment while the new medication is continued at therapeutic doses, to evaluate efficacy as a monotherapy.

Some have argued that if evidence of efficacy is established in add-on studies such as those described above, given that such studies target patients with medication-resistant conditions (i.e., patients already taking one or more antiepileptic drugs but who continue to have symptomatology), then separate trials in more amenable (i.e., less treatment-resistant) patients are unnecessary because the study medication has already been proven effective in a more complex, difficult-to-treat population (Brodie 1996; Brodie and Dichter 1996; Brodie et al. 1995). What cannot be deciphered with certainty from this type of add-on design is whether an effective add-on agent is an effective monotherapy.

This add-on or “discontinuation” design can also enhance patient safety. A survival analysis can be used in conjunction with the setting of stringent subject nonresponse or worsening criteria (such as doubling of the average monthly symptomatology, doubling of the highest 2-day symptomatology, or the development of any psychotic symptoms) to remove subjects. Then, survival analysis captures the comparative benefits (or lack thereof) of each masked treatment. Those who exit the
study without meeting those criteria (i.e., administrative or unexplained withdrawals) provide censored data. They do not contribute to the survival curve beyond the time they were in the study. This survival analysis compares the benefits of the new treatment while protecting patient safety to an optimal degree. It also allows much easier initial patient recruitment. It also allows, given randomization, an assessment of the spontaneous waxing and waning of symptoms and disability over more prolonged time periods than the traditional 4- to 6-week trials that require weekly symptom assessments as commonly used in trials of major depressive disorder.

Alternatively, the add-on of one or more medications in a randomized trial can be systematically tested for efficacy in individual patients using an off-on-off-on (B-A-B-A) design to avoid exposure to placebo altogether for nonresponders (Ballanger and Post 1978; McDermut et al 1995). The duration of the trial can also be individualized based on the patient’s presenting baseline cycle frequency, thereby determining a set threshold for the degree of acceptable improvement on and off drug (McDermut et al 1995; Post et al 1998b).

Defining Where a New Treatment Fits in Clinical Practice

Brodie (1996) highlighted the need to conduct studies (during or following regulatory approval) to determine where new medications fit in our overall therapeutic armamentarium. Because new medications cost substantially more than generic agents, it is important to determine where these newer medications best fit in terms of value for the health care dollar, as well as, whenever feasible, to identify patients who preferentially respond to either monotherapy or combination treatment with the new agent.

This issue can be addressed with the add-on or discontinuation design noted above, beginning with patients who are already receiving mood stabilizing agent(s) but who have sufficient symptomatology to warrant additional pharmacotherapeutic interventions (Figure 2). In this case, the add-on therapy could be provided initially in a low dose to assay for tolerability under open label. Then, under double-blind conditions, doses of each of the two candidate monotherapies are increased during the conversion period. Thereafter, prior mood stabilizers are discontinued. Thus, one would have a comparison of two agents as adjunctive therapies and then as monotherapies in these patients.

Alternatively, one could simply leave patients on their pretreatment medications and add (double-blind) either an additional standard medication (A), or the experimental medication to the mix or one of two new medications (X and Y) with a crossover to assess individual responsivity. This design studies the comparative value of agent X versus agent Y as adjuncts to treatment in those already receiving one or more insufficiently beneficial mood stabilizers.

Following these observations to test for comparability of efficacy between the two candidate adjunctive agents,
baseline medication(s) could be discontinued to assay for comparative efficacy of the two agents as monotherapies. The individual added benefit from the combination compared to monotherapy could be further evaluated in the off-on-off-on design (McDermut et al 1995).

Another way to address the issue of where a medication fits in the treatment plan is to design medication treatment algorithms beginning with monotherapy, but adding new medications in a systematic way to a consistent prestudy medication regimen (Figure 3). The key decision points at which each new medication is added are cued by an evaluation of current symptomatology (manic, mixed, depressed, or mood reactivity). Such algorithms have been proposed (Bauer et al 1999; Calabrese and Woyshville 1995b; Dennehy and Suppes 1999; Nolen et al 1998; Post et al 1996b; Sachs 1996; Suppes et al 1998). To our knowledge, no comparative trials of one algorithm in contrast to another have been undertaken in bipolar disorder; however, the simultaneous use of multiple mood stabilizers is becoming common practice (Freeman and Stoll 1998; Solomon et al 1996), in part because of increasing evidence that full symptomatic remission, rather than simple symptom reduction, is associated with less disability and a better prognosis (Fava et al 1998a, 1998b; Judd et al 1997, 1998; Miller et al 1998; Paykel et al 1999).

Finally, there are a plethora of practical issues, not the least of which are the timely recruitment of large numbers of representative patients with bipolar disorder, reluctance of practitioners to “give up” patients and associated income for research studies, and the overall reduction in length of hospital stays. One solution to these and other practical complexities is to engage practitioners to conduct the research and properly reimburse them, similar to what is currently being planned in the NIMH-sponsored Sequenced Treatment Enhancement Program-Bipolar Disorder (University of Pittsburgh, Epidemiology Data Center 2000b) and Sequenced Treatment Alternatives to Relieve Depression (University of Pittsburgh, Epidemiology Data Center 2000a) trials. Similarly, subjects in the protocol can be followed in and out of different care settings (inpatient, day hospital, intensive outpatient, and outpatient clinic), which in turn is more representative of actual practice. Nonetheless, post hoc statistical adjustments may be needed in such situations because other nonprotocol treatments are often provided to differing degrees in diverse care settings, and adherence to protocol procedures may also depend on the intensity and frequency of staff–patient interactions that vary across these settings.

**Other Issues**

This article has focused on studies of acute phase treatment; however, additional major design, feasibility, patient sample, ethical, outcome measurement and other issues pertain to longer term maintenance trials. This topic deserves a detailed discussion beyond the scope of this article. Designs and methods developed in longer term chronic disease management (often effectiveness) trials (e.g., in cancer or heart disease) may be especially useful when approaching these issues. We plan a subsequent paper on the problems encountered in long-term studies of bipolar disorder.

**Conclusions**

We have reviewed a number of impediments to developing and evaluating new treatments for bipolar disorder, and we have suggested some potential solutions to these problems. Some of these solutions may also be applicable to studies of patients with other severe and persistent mental illnesses. First, broadening the inclusion and narrowing the exclusion criteria in clinical trials involving patients with bipolar disorder both increases feasibility and decreases cost, while protecting sufficient internal validity and increasing external validity compared with more narrowly defined populations.

Second, using detailed, continuous, longitudinal outcome assessments conducted over longer periods of time and using broader outcome domains (e.g., both symptoms and functional status) avoids the difficulty of arbitrarily brief time periods for outcome evaluation. The LCM is one readily available method of assessment that can be used until a consensus is reached regarding standard outcome measures to be recommended for all efficacy trials involving patients with bipolar disorder.

Third, by adopting some of the study designs from efficacy trials of patients with epilepsy, it is possible to study treatment-resistant (e.g., partial or nonresponders to standard pharmacotherapy) patients who are in need of additional medication by using an add-on protocol. Such designs retain an optimal degree of patient safety while determining whether the new medication offers additional therapeutic value beyond current monotherapies or treatment combinations. If these add-on studies establish both safety and efficacy, they can then be followed by a subsequent discontinuation of prestudy medications to test the new agent as a monotherapy. Taken together, these approaches address the question of whether new medications are effective adjuncts and whether they are effective monotherapies. If both questions are answered in the affirmative, then acute phase efficacy studies with a placebo comparison in treatment-resistant, medication-free patients may not be needed. Finally, the diligent introduction of medications, their discontinuation, or both can help to identify the clinical place of new medications (either as monotherapy or in combination) in the overall management of patients with bipolar disorder.
References


Preparation of this report was supported in part by the Theodore and Vada Stanley Foundation, National Institute of Mental Health (Grant No. MH-53899), the Betty Jo Hay Distinguished Chair in Mental Health, the Rosewood Corporation Chair in Biomedical Science, and the Sarah M. and Charles E. Seay Center for Basic and Applied Research in Psychiatry. The authors appreciate the secretarial support of David Savage and Fast Word, Inc., Dallas. We are also grateful to Kenneth Z. Altshuler, M.D., Stanton Sharp Chair and Professor, Department of Psychiatry, UT Southwestern Medical Center, for his administrative support.

Aspects of this work were presented at the conference “Bipolar Disorder: From Preclinical to Clinical, Facing the New Millennium,” January 19–21, 2000, Scottsdale, Arizona. The conference was sponsored by the Society of Biological Psychiatry through an unrestricted educational grant provided by Eli Lilly and Company.


