Bipolar Depression: Molecular Neurobiology, Clinical Diagnosis and Pharmacotherapy
Edited by C.A. Zarate and H.K. Manji

Pharmacological treatment of the maintenance phase of bipolar depression: focus on relapse prevention studies and the impact of design on generalizability

Keming Gao, David E. Kemp and Joseph R. Calabrese

Department of Psychiatry, Bipolar Disorder Research Center in Mood Disorders Program, University Hospitals Case Medical Center, Case Western Reserve University School of Medicine, 10524 Euclid Avenue, 12th Floor, Cleveland, Ohio 44106, USA

Abstract

The goal of pharmacological treatment of bipolar disorder is to prevent future occurrences of mood episodes. To achieve this goal, medications must demonstrate efficacy in the prevention of both manic/hypomanic and depressive relapses/recurrences. Currently, the efficacy of most pharmacological agents in the maintenance treatment of bipolar disorder has been studied using relapse prevention designs, in which only patients who tolerate and respond to a studied drug(s) in the acute phase (mania or depression) can enter the maintenance phase. Subsequently, the results from relapse prevention studies are not generalizable, not only because of the design, but also because of different index mood episodes. So far, however, only lithium and lamotrigine, and to some extent divalproex, have been investigated in both manic and depressive index episodes, while olanzapine and aripiprazole have been evaluated in manic index episodes. To facilitate the application of currently available data, this chapter will systematically examine randomized, blinded, controlled maintenance studies enrolling ≥100 patients and lasting ≥6 months.

Introduction

Symptomatic patients with Bipolar I disorder experience depressive symptoms three to four times more commonly than manic symptoms [1], while symptomatic patients with Bipolar II disorder experience depressive symptoms approximately 39 times more commonly than hypomanic symptoms [2]. In a prior report of 593 subjects who screened positive for bipolar disorder, moderate to extreme impairment in work, social life, and family interactions occurred significantly more often from depressive than manic symptoms [3]. In addition, significant psychosocial impairment during illness-free periods was more strongly predicted by the number of past depressive episodes than past manias [4, 5]. Moreover, there is a higher risk for suicide in bipolar disorder than in other major psychiatric disorders, particularly during depressive episodes [6–8].

From these findings, it is evident that the prevention of acute major depressive episodes is critical in order to reduce the morbidity and mortality associated with bipolar disorder. This necessarily involves the discovery and implementation of more effective treatments during the maintenance phase. Disappointingly, the prevention of bipolar depression during maintenance treatment is less well studied compared with the prevention of mania. More importantly, most maintenance studies have used relapse prevention designs instead of prophylaxis designs. "In the prophylaxis design, any patient who is euthymic, regardless of how that person got well, is eligible to be randomized to drug *versus* placebo or a comparator. In the relapse prevention design, only those patients who respond acutely to the drug being studied are eligible to enter the maintenance phase, when they are randomized to remain on the drug or be switched (usually abruptly) to placebo and/or an active comparator…Thus results from a prophylactic design are generalizable, whereas those from a relapse prevention trial are not." (p. 709) [6]. Evidence suggests the index mood episode often predicts the polarity of future relapse at different times after remission [9], but it also may predict different responses to treatment (see below) [10]. Likely, results from patients initially presenting with mania may not be applicable to those initially presenting with depression or *vice versa*. Even results from studies with the same index mood polarity may not be generalizable because of differences in inclusion and exclusion criteria or other variations in study designs. Therefore, an understanding of these differences is important when considering generalizability to clinical practice. To achieve this goal, we will provide an overview of the various trial methodologies and summarize key efficacy findings from existing bipolar maintenance studies.

Methodology in bipolar maintenance research

Although there is little consensus on the methodology of bipolar maintenance studies, the methods that have evolved the most include study enrollment, study design, outcome measures, and statistical analysis [6, 11]. These components can directly affect interpretation of the results and will be reviewed independently.

Study enrollment

Due to different inclusion and exclusion criteria, selection biases stemming from study enrollment are unavoidable. Most early maintenance studies published between 1970 and 1976 tended to evaluate small cohorts of patients, typically from 5 to 40 patients per arm. Most studies only enrolled patients who were hospitalized. At the time of these early studies, concepts such as Bipolar II disorder, secondary bipolar disorder, and rapid cycling were not yet

introduced into practice; only mood-congruent psychotic features were allowed. Furthermore, the use of hospitalization to confirm a diagnosis of mania was commonly employed. Consequently, patients with mania enrolled into the early studies tended to have classic euphoric mania. After redefining the criteria for the diagnosis of mania in DSM-III in 1980 and permitting outpatients into the study, more recent maintenance studies probably included less-impaired patients who were more likely to respond to placebo treatment [11].

Study design (crossover versus *parallel)*

Crossover designs were used in some early lithium maintenance studies. In a typical crossover study, each patient's response under treatment A is compared with his or her response under treatment B. The advantage of this design is the increased homogeneity of the study population. The primary disadvantage associated with the use of a crossover design is the increased risk for false-positive results due to abrupt discontinuation of the initial treatment [11].

In a parallel study design, all patients are randomized to different treatments at the beginning of the study. The advantage of the parallel design is that it is less dependent on the crossover design's assumption that each individual patient has a similar disease process prior to and during a treatment, and generally produces a lower dropout rate because each patient is exposed to only one treatment. The primary disadvantage of the parallel design is the disturbance caused by spontaneous remissions or erratic, short-lived fluctuations in mood states. Most maintenance studies have used an open label stabilization phase in which all patients receive active drug(s), followed by random assignment to different treatment arms in a parallel fashion. The 'enriched' nature of the sample tends to increase homogeneity, similar to the crossover design. Like the crossover scheme, the parallel discontinuation design used in most recent maintenance studies might also inflate the response of the studied drugs.

Outcome measures

Most early lithium studies did not measure mood severity with standardized rating scales, such as the Hamilton Rating Scale of Depression (HAM-D) or the Young Mania Rating Scale (YMRS). Typically, these trials evaluated efficacy through general indices of outcome during the study period. Such outcomes included the number of manic and depressive episodes, the probability of relapsing into a manic or depressive episode, experiencing a relapse severe enough for hospitalization, or experiencing a relapse severe enough for a pharmacological intervention. More recent maintenance studies have used standardized rating scales to quantify the minimum severity for an index episode or DSM-IV criteria for a mood episode. The advantage of using rating scales

is that they can detect rather minor changes in illness severity. However, they are limited by their cross-sectional assessment of a period of 7 days prior to the completion of the rating scale. Some recent studies have used the time to intervention as a primary outcome measure, but each individual study has its own criteria for intervention [11].

Statistical analyses

Early maintenance studies used responder analyses with little or no distinction between primary and secondary outcome measures. The proportion of patients who experienced a relapse or recurrence was commonly compared [12–14] with the exception of one study that used additional Kaplan-Meier life table analysis [15]. This approach is inadequate for a maintenance study because it does not consider the length of time that patients remain well before relapsing. A further drawback of this method is that the number of patients withdrawing prematurely without experiencing a relapse is either ignored or analyzed incorrectly.

The use of survival analysis has become the standard method of examining data from bipolar disorder maintenance studies. Relapse prevention trials typically assess the length of time after open stabilization, starting at the point of randomization and ending when relapse or recurrence has occurred. These survival data are commonly depicted with a Kaplan-Meier curve, which can also be used to demonstrate median survival (the time at which 50% of patients have relapsed/recurred). There are several methods available to conduct significance testing on time-to-event data, including log rank, Cox proportional hazards, and the Wilcoxon two-sample test. The log rank test is the most commonly used method and was used in the lamotrigine/Lamictal® [16–18], aripiprazole/Abilify® [19], and olanzapine/Zyprexa® maintenance studies [20]. Cox regression testing was used in the other olanzapine maintenance studies, presumably because of its flexibility, which permits the use of multiple predictors and covariates [21, 22]. In the divalproex/Depakote® maintenance study, life-table methods were constructed to compare survival curves using the Wilcoxon test because of this method's sensitivity to early group differences [23]. However, the primary problems with survival analysis techniques derive from sample size and drop-outs. As the sample size decreases over time due to drop-outs, a survival analysis is most valid for the earlier portion of the curves, where there are a larger number of patients in the study [6].

The index mood episode

In relapse prevention studies, an index mood episode (a presenting mood episode) is commonly used to start a trial of a study drug, although each study has its own severity requirement (Tab. 1). The importance of the index mood

Table 1. Summary of the characteristics of maintenance studies during open-label phase

Open label enrichment	Study index severity	Open-label treatment (s)	Enrollment		Stabilization Criteria	Randomization scheme	Randomized treatments
			Entry (N)	Completion N (%)			
Agents studied in both mania and depression index episodes							
Lithium or Imipramine [12][¶]							
	Bipolar depression or Unipolar depression Hospitalized	Antidepressant or ECT first, then Lithium (0.5–1.4 mEq/L) or Imi (50–200 mg/day)	122	–	After remission	Abrupt discontinuation Parallel assignment	Li Imi PBO
Lithium [13][¶]							
	Acute mania Hospitalized	Li (0.5–1.4 mEq/L)	205	–	After remission	Abrupt discontinuation Parallel assignment	Li PBO
Lithium and Imipramine [15][*]							
	Acute mania, bipolar depression or unipolar depression RSDMS ≥ 7, GAS ≤ 60 Inpatients (42%) and outpatients	Treatment of choice first, then Imi (75–150 mg/day) plus Li (0.6–0.9 mEq/L) for 2 months	216	117 (54.2)	RSDMS < 7, GAS > 60 for 2 months	Abrupt discontinuation Parallel assignment	Li + Imi Li Imi
Lamotrigine [16]							
	Bipolar I depression A current DSM-IV MDE or MDE within 6 months but symptomatic at enrollment	LTG (100–200 mg/day) adjunctive or monotherapy for 8–16 weeks	966	480 (49.7)	CGI-S ≤ 3 for ≥4 weeks	Abrupt discontinuation Parallel assignment lithium with 3 weeks[§]	LTG Li PBO

(Continued on next page)

Table 1. (Continued)

Open label enrichment	Study index severity	Open-label treatment (s)	Enrollment		Stabilization Criteria	Randomization scheme	Randomized treatments
			Entry (N)	Completion N (%)			
Lamotrigine [17]							
	Bipolar I disorder Current mania/hypomania or mania/hypomania with 60 days but symptomatic at enrollment	LTG (100–200 mg/day) adjunctive or monotherapy for 8–16 weeks	349	175 (50.1)	CGI-S ≤ 3 for ≥4 weeks	Abrupt discontinuation Parallel assignment lithium with 3 weeks[§]	LTG Li PBO
Lithium and Divalproex [18]							
	Bipolar I or II disorder Rapid cycling in last 12 months Mania, hypomania, mixed state within last 3 months	Li (≥ 0.8 mEq/L) Val (≥ 50 μg/ml) for 12–20 weeks	254	60 (23.6)	HAM-D-24 ≤ 20, YMRS ≤ 12, GAS ≥ 51 for ≥4 weeks	Gradual discontinuation for average 6 weeks	Li + PBO Val + PBO
Agents only studied in Mania Index Episode							
Aripiprazole [19]							
	Bipolar I disorder, mania or mixed YMRS ≥ 20	ARIP (15 or 30 mg/day) for 6–18 weeks	567	206 (36.3)	YMRS ≤ 10, MADRS ≤ 13 for ≥6 weeks	Abrupt discontinuation Parallel assignment	ARIP PBO
Lithium and Olanzapine [20]							
	Bipolar I disorder, mania or mixed YMRS ≥ 20 Inpatients and outpatients	OLZ (5–20 mg/day) Li (0.6–1.2 mEq/L) for 6–12 weeks	543	431 (79.4)	YMRS ≤ 12, HAM-D-21 ≤ 8	Gradual discontinuation for 4 weeks	OLZ + PBO Li + PBO

(Continued on next page)

Table 1. (Continued)

Open label enrichment	Study index severity	Open-label treatment (s)	Enrollment Entry (N)	Enrollment Completion N (%)	Stabilization Criteria	Randomization scheme	Randomized treatments
Olanzapine [21]							
	Bipolar I disorder, mania or mixed YMRS ≥ 20	OLZ (5–20 mg/day) for 6–12 weeks	731	361 (49.4)	YMRS ≤ 12, HAM-D-21 ≤ 8 2 consecutive weekly visit	Abrupt discontinuation Parallel assignment	OLZ PBO
Lithium or Valproate and Olanzapine [22][**]							
	Bipolar I disorder, mania or mixed YMRS ≥ 16 Inadequate response to Li or Val for 2 weeks	Li (0.6–1.2 mEq/L) Val (50–125 µg/ml) OLZ (5–20 mg/day) for 6 weeks	229	99 (43)	YMRS ≤ 12, HAM-D-21 ≤ 8 Syndromic remission[***]	Abrupt discontinuation Parallel assignment	Li/Val + OLZ Li/Val + PBO
Divalproex, Lithium or both, others, or none [23]							
	Bipolar I disorder, manic or partially recovered from mania or euthymic after mania	Val, Lithium, or both, others or none for ≤ 3 months	571	372 (65.1)	MAR ≤ 11, DSS ≤ 13, GAS ≥ 60 for ≥6 days	Gradual discontinuation Parallel assignment for 2 weeks	Val Li PBO

[¶] Relapsed patients were allowed to continue the study; [§] For those taking lithium during open-label phase, the dosage was tapered over at least 3 weeks and discontinued a minimum of 1 week prior to randomization; [#] Research Diagnostic Criteria for primary major depressive disorder or manic disorder; [**] Olanzapine was a part of blinded acute treatment for mania; [***] Syndromic remission was defined as 1) DSM-IV 'A' criteria for current manic episode no worse than mild, 'B' criteria no worse than mild, and no more than 2 'B' criteria that were mild; 2) All DSM-IV 'A' criteria for current major depressive episode no worse than mild, and no more than 3 'A' criteria mild. Abbreviations: ARIP, aripiprazole, CGI-S, Clinical Global Impression – Severity scale; DSS, Depression Severity scale; GAS, Global Assessment scale; HAM-D-21, Hamilton Depression Rating scale 21 items; HAM-D-24, Hamilton Depression Rating scale 24 items; Imi, imipramine; Li, lithium; LTG, lamotrigine; MADRS, Montgomery – Asberg Depression Rating Scale; MAR, Mania Rating Scale; MDE, major depressive episode; n, number of patients; OLZ, olanzapine; PBO, placebo; RSDMS, Raskin Severity of Depression and Mania Scale; Val, valproate or divalproex; YMRS, Young Mania Rating Scale.

episode is that not only may it bring patients into acute treatment; it may also predict the mood polarity of future relapses, as well as treatment response. After reviewing 11 published articles, Calabrese and colleagues concluded that patients presenting with depression for a maintenance study tended to relapse into depression, and that those presenting with mania tended to relapse into a manic, hypomanic, or mixed episode with a ratio of about 2:1 to 3:1; this was particularly true during the first few months after randomization [24]. However, the polarity of a new mood episode differed depending on the time after randomization from an index mood episode. Relapses within the first 90 days were more likely to be of the same polarity as the index mood episode, while relapses after 180 days were more likely to be of the opposite polarity as the index mood episode [9]. Furthermore, in a *post hoc* analysis of a bipolar maintenance study [10], Shapiro and colleagues found that lithium monotherapy or the combination of lithium and imipramine were superior to imipramine alone in prevention of relapses in patients presenting with mania, but not in those presenting with depression.

Clearly, any drug must be studied with different index mood episodes, and the study must last long enough to show a spectrum of efficacy in the prevention of early and later mood relapses. However, due to the nature of bipolar disorder, no consensus has been achieved for the ideal duration of a bipolar maintenance study. Investigators have purported that 'pure' maintenance efficacy of maintenance studies cannot be established if the study duration is shorter than 6 months [6, 25]. Therefore, in this chapter, randomized, blinded, controlled trials with an index episode of mania or depression enrolling ≥100 patients and lasting more than 6 months will be systematically reviewed.

Agents studied in both manic and depressive mood episodes

So far, only lithium and lamotrigine have been studied in patients presenting with a manic or depressive episode. In most early studies, patients were 'enriched' with lithium monotherapy or a lithium + imipramine/Tofranil® combination before they were randomized. In recent studies, patients have been 'enriched' with lamotrigine (Tabs 1 and 2).

Lithium or lithium + imipramine (mildly enriched)

Lithium's observed acute antimanic efficacy led to mania being studied as the index episode for a number of early maintenance studies. In response to criticisms raised over the use of crossover designs in some early lithium studies and the skepticism relating to its maintenance efficacy, Prien and colleagues (1973) published two randomized, single-blinded (rater-blinded) prospective maintenance studies [12, 13]. In these studies, hospitalized manic patients were stabilized with lithium (*lithium-enrichment*) and hospitalized depressed

Table 2. Summary of maintenance studies during randomized, blinded, controlled phase

Enrichment Index mood episode Randomization criteria	Randomized treatments	Duration	Completion (%)	Primary outcome measure	Median time to intervention/relapse (Active arm *versus* PBO for PBO-controlled trials) or (Active *versus* active for active comparator trials)			Rates of relapse/recurrence (%) (Active arm *versus* PBO for PBO-con trolled trials) or (Active *versus* active for active comparator trials)		
					Any mood	Mania	Depression	Any mood	Mania	Depression
Agents studied in both mania and depression index episodes										
Lithium or Imipramine [12]										
Depression	Li (0.5–1.4 mEq/L), n = 18	24 months	45*	Occurrence of affective	—	—	—	†28¥	11	22¥
After remission	Imi (50–200 mg/day), n = 13			episodes (hospitalization				†77	54	31
	PBO, n = 13			or supplementary drug)	—	—	—	†77	38	62
Lithium [13]										
Mania	Li (0.5–1.4 mEq/L), n = 101	24 months	59*	Frequency and severity of	—	—	—	42¥	32¥	16
After remission	PBO, n = 104			relapse				80	68	26
				(hospitalization-"severe" relapse, supplementary drug-"moderate")	—	—	—			
Lithium and Imipramine [15]**										
Depression or mania	Li (0.45–1.1 mEq/L) +	30 months	17***	Occurrence of affective	—	—	—	—	28	22
RSDMS < 7,	Imi (75–150 mg/day), n = 36			episodes						
GAS > 60	Li (0.45–1.1 mEq/L), n = 42			(RDC Criteria of MDE or	—	—	—	—	26	29
for 2 months	Imi (75–150 mg/day), n = 36			ME and GAS ≤ 60)	—	—	—	—	53	28

(Continued on next page

Table 2. (Continued)

Enrichment Index mood episode Randomization criteria	Randomized treatments	Duration	Completion (%)	Primary outcome measure	Median time to intervention/relapse			Rates of relapse/recurrence (%)		
					Any mood	Mania	Depression	Any mood	Mania	Depression
Lamotrigine [16]										
Depression CGI-S ≤ 3 for at least 4 weeks	LTG (50, 200, 400 mg/day), n = 221	76 weeks	17	Time to intervention (additional pharmacotherapy or electroconvulsive therapy)	200 days, p = 0.029	p = ns	Longer, p = 0.047	50	16	35
	Li (0.8–1.1 mEq/L), n = 121		17		170 days, p = 0.029	Longer, p = 0.026	p = ns	47	8	38
	PBO, n = 121		10		93 days			56	16	40
Lamotrigine [17]										
Mania CGI-S ≤ 3 for at least 4 weeks	LTG (100–400 mg/day), n = 59	76 weeks	5	Time to intervention (additional pharmaco therapy or electroconvulsive therapy)	141 days, p = 0.02	NE, p = ns	NE, P = 0.02	48	35	14
	Li (0.8–1.1 mEq/L), n = 46		2		292 days, p = 0.003	NE, p = 0.006	NE, p = ns	41	18	23
	PBO, n = 70		0		85 days	203 days	269 days	71	41	30
Lithium and divalproex [18]										
Mania/hypomania in 3 months	Li (0.92 mEq/L), n = 32	20 months	16	Time to treatment for a mood episode	18 weeks, p = ns	NE, p = ns	NE, p = ns	56	22	34
HAM-D-24 ≤ 20, YMRS ≤ 12, GAS ≥ 51 for ≥ 4 weeks	Val (77 μg/ml), n = 28		29	(emerging symptoms of a relapse judging by Investigator or a full relapse)	45 weeks	NE	36 weeks	50	22	29

(Continued on next page)

Table 2. (Continued)

Enrichment Index mood episode Randomization criteria	Randomized treatments	Duration	Completion (%)	Primary outcome measure	Median time to intervention/relapse			Rates of relapse/recurrence (%)		
					Any mood	Mania	Depression	Any mood	Mania	Depression
Agents studied only in mania index episode										
Aripiprazole [19]										
Mania YMRS ≤ 10,	ARIP (15 or 30 mg/day), n = 78	100 weeks	18	Time to relapse for a mood episode	Longer, p = 0.011	Longer, p = 0.005	p = ns	33[¥]	12[¥]	14
MADRS ≤ 13 for ≥ 6 weeks	PBO, n = 83		19	(Manic, depressed, or mixed, discontinuation due to lack of efficacy)				52	23	16
Lithium and Olanzapine [20]										
Mania YMRS ≤ 12,	OLZ (13.5 ± 4 mfg/day), n = 217	52 weeks	47	Symptomatic mood episode recurrence	Longer, p = 0.07	—	—	30	14[¥]	16
HAM-D-21 ≤ 8	Li (0.697 ± 0.14 mEq/L), n = 214		33	(YMRS ≥ 15, HAM-D-21 ≥ 15)		—	—	39	23	11
Olanzapine [21]										
Mania YMRS ≤ 12,	OLZ (5–20 mg/day), n = 225	48 weeks	21	Time to symptomatic relapse	174 days, p = 0.001	[£]NE, p < 0.001	[£]49 days, p < 0.001	47[¥]	12[¥]	30
HAM-D-21 ≤ 8 for 2 consecutive weekly visit	PBO, n = 136		7	(YMRS ≥ 15, HAM-D-21 ≥ 15 or hospitalization for any mood)	22 days	[£]26 days	[£]18 days	80	32	39
Lithium or Valproate and Olanzapine [22]										
Mania Syndromic remission	Li (0.76 mEq/L) or Val (67.8 µg/ml) + OLZ (8.6 mg/day), n = 51	18 months	31	Syndromic relapse (DSM-IV criteria for a manic, mixed or depressive episode) & symptomatic relapse (YMRS ≥ 15, HAM-D-21 ≥ 15)	[§]163 days, p = ns	172 days, p = ns	163 days, p = 0.071	37	20	23
	Li (0.74 mEq/L) or Val (66.3 µg/ml) + PBO, n = 48		10		42 days	59 days	55 days	55	29	40

(Continued on next page)

Table 2. (Continued)

Enrichment Index mood episode Randomization criteria	Randomized treatments	Duration	Completion (%)	Primary outcome measure	Median time to intervention/relapse			Rates of relapse/recurrence (%)		
					Any mood	Mania	Depression	Any mood	Mania	Depression
Divalproex, lithium or both, others, or none [23]										
Mania, partially recovered from	Val (71–125 μg/L), n = 187	12 months	[#]38	Time to any mood episode	275 days,[£] p = ns	> 365 days,[£] p = ns	126 days, p = ns	—	—	27[¶]
mania, euthymic after mania	Li (0.8–1.2 mEq/L), n = 91		[#]24	(MRS ≥ 16 or requiring hospitalization) or	189 days, p = ns	[£]293 days, p = ns	[£]81 days, p = ns	—	—	28[¶]
MAR ≤ 11, DSS ≤ 13, GAS ≥ 60 for ≥6 days	PBO, n = 94		[#]25	(antidepressant use or discontinuation because of depressive symptoms)	173 days	[£]189 days	[£]101 days	—	—	26[¶]

[*] Relapsed patients were allowed to continue the study; [†] Data from a reanalysis of the original study [14]; [¥] Significant difference (P < 0.05 to <0.001); [**] Research Diagnostic Criteria for primary major depressive disorder or manic disorder; [***] Excluding those terminated while in good clinical state during year 1 (13%), terminated while in good clinical state during year 2 (3%), and remained well for the study duration (22%); [£] Estimated 25% time to relapse; [§] there was also no significant difference between combination therapy and monotherapy in delaying the time to syndromic relapse, 94 days *versus* 41 days, p = 0.742; [#] Patients receiving antidepressants for depression were allowed to continue the study; [¶] Data from a *post hoc* analysis of the original study [26].

Abbreviations: ARIP, aripiprazole, CGI-S, Clinical Global Impression – Severity scale; DSS, Depression Severity scale; GAS, Global Assessment scale; HAM-D-21, Hamilton Depression Rating scale 21 items; HAM-D-24, Hamilton Depression Rating scale 24 items; Imi, imipramine; Li, lithium; LTG, lamotrigine; MADRS, Montgomery – Asberg Depression Rating Scale; MAR, Mania Rating Scale; MDE, major depressive episode, ME, mood episode; n, number of patients; NE, not evaluable; ns, no significance; OLZ, olanzapine; PBO, placebo; RSDMS, Raskin Severity of Depression and Mania Scale; Val, valproate or divalproex; YMRS, Young Mania Rating Scale.

(bipolar or unipolar) patients were stabilized with lithium-imipramine (*lithium-imipramine-enrichment)* after the remission of an acute episode, and prior to discharge from the hospital (Tab. 1). After stabilization, depressed patients were randomized to receive lithium, imipramine, or placebo [12]. In the primary outcome analysis, bipolar patients treated with placebo or imipramine experienced more episodes than patients receiving lithium. However, the differences were statistically significant during months 5 to 24, but not at the 4 month endpoint. The superiority of lithium in preventing any mood episode relapse compared to imipramine and placebo was further demonstrated when the 24-month data were analyzed as a whole [14]. The difference between the lithium and the imipramine groups was due almost entirely to the higher incidence of manic episodes in the imipramine group. However, the difference between the lithium and placebo groups was due to both manic and depressive episodes (Tab. 2).

In the mania study [13], patients were randomized to receive lithium or placebo after stabilization. 70 of 104 patients on placebo had at least one severe relapse compared to only 31 of 101 on lithium ($p < 0.001$), but no significant difference was found in the rate of moderate relapses between the two groups. Analysis of first or second year results yielded similar findings. Despite the fact that stabilizations were carried out after remissions (mild enrichments) of the acute episodes for both studies, these results may not be generalizable to modern clinical practice due to the unstructured diagnostic criteria for bipolar disorder. Furthermore, as with other early prospective studies, the abrupt discontinuation of lithium in the placebo or imipramine group might have inflated the effect of lithium.

In order to provide more definitive data regarding the long-term preventive treatment of recurrent affective illness, Prien and colleagues (1984) published a double-blinded study of lithium, imipramine, or both in a mixed group of patients with mania, bipolar depression, or unipolar depression [15]. In contrast to the prior studies that enrolled only hospitalized patients, this study permitted the inclusion of outpatients. In addition, diagnoses were based on the Research Diagnostic Criteria (RDC), which were used to define major depressive or manic episodes. Subjects were also required to have a Global Assessment Scale (GAS) score of ≤60, but hospitalization was not required (Tabs 1 and 2).

The primary analysis showed no difference in the rates of depressive recurrence among the three groups (Tab. 2). However, there was a significantly higher rate of manic recurrence in the imipramine monotherapy group than in the lithium monotherapy group (Tab. 2). The Kaplan-Meier life-table analysis showed that lithium alone and lithium in combination with imipramine were superior to imipramine monotherapy in delaying the recurrences of mood episodes. When analyzed by the study index mood episode, patients with a manic or mixed episode responded much better to lithium alone or in combination with imipramine as compared with imipramine alone, with corresponding success rates of 53%, 47%, and 8% respectively. By contrast, there were

no significant differences among the three treatments for patients with a depressive index episode, with corresponding success rates of 22% for lithium, 18% for the combination, and 9% for imipramine alone.

Allowing outpatients into the study and using less restricted treatment during the open-label phase made the results of this study more generalizable. However, like many other early studies of lithium, the time to relapse into mania or depression was not used as a primary outcome. Again, the abrupt discontinuation of lithium in the imipramine alone group might have increased the risk of relapse for this group.

In a *post hoc* analysis of this study that used Kaplan-Meier survival analysis (product-limit method) and a Cox regression model [10], Shapiro and colleagues not only replicated the initial findings, but also determined that patients with a manic index episode who took imipramine were almost 11 times more likely to have a recurrence than subjects taking lithium, and five times more likely than those taking the combination. More importantly, among patients with a depressive index episode, the treatments differed significantly. The combination was significantly superior to imipramine, but it failed to reach statistical superiority to lithium alone. Patients with a depressive index episode taking imipramine were three times more likely to suffer a recurrence than those taking the combination. During a 24 month period, the estimated median time in remission for those with a manic index episode was not calculable for the lithium alone group because fewer than half of patients relapsed; time in remission was 14.8 months for those patients taking the combination, and 3.1 months for those taking imipramine alone. The median time in remission for those with a depressive index episode was 3.4 months for lithium, 7.6 months for combination, and 4.8 months for imipramine, respectively. This *post hoc* analysis highlighted the limitations of statistical analyses performed in previous studies.

Lamotrigine (moderately enriched)

The efficacy of lamotrigine in the maintenance treatment of bipolar disorder was compared with lithium and placebo among patients with both mania and depression index episodes [16, 17]. The design of these two studies was quite similar except for the study index mood episode (Tabs 1 and 2). Patients were 'enriched' by responding to open treatment with lamotrigine adjunctively or as monotherapy prior to randomization. In the depression study [16], the median times to treatment intervention were 93 days for placebo, 170 days for lithium, and 200 days for lamotrigine (combined 200 mg/day and 400 mg/day). Both lithium and lamotrigine were superior to placebo in prolonging the time to intervention for any mood episode (Tab. 2). Lithium and lamotrigine did not differ from each other on this measure. The median times to depression intervention were similar between lithium and placebo, but significantly longer with lamotrigine as compared with placebo. On the other hand, the median

time to mania intervention was significantly longer in lithium- than placebo-treated subjects. Once again there was no significant difference between lithium and lamotrigine. The rates of depression relapse were lower in the lamotrigine group, but rates of manic relapse were lower in the lithium group (Tab. 2).

In the mania study [17], both lamotrigine and lithium were significantly superior to placebo on the median time to intervention for any mood episode (141 days for lamotrigine, 292 days for lithium, and 85 days for placebo). Lamotrigine and lithium did not differ from each other on this measure. Lamotrigine, but not lithium, was superior to placebo at prolonging the time to a depressive episode. Lamotrigine and lithium did not differ on this measure. In contrast, lithium, but not lamotrigine, was superior to placebo at prolonging time to a manic/hypomanic/mixed episode. A trend favored lithium over lamotrigine on this parameter ($p = 0.09$). There were fewer incidences of depressive relapse in the lamotrigine group, but fewer incidences of manic relapse in the lithium group (Tab. 2).

These trials represent the first studies ever conducted in a Bipolar I population in which lithium differentiated from placebo using DSM-IV criteria and modern survival analytic methods, providing some of the strongest evidence available for the efficacy of lithium in the maintenance treatment of bipolar disorder. By using time to intervention for an emerging mood episode as the primary outcome measure, the threshold for detecting a treatment 'failure' was essentially lowered, improving the overall sensitivity for mood worsening [16]. Both studies also slowly tapered lithium for those subjects who received lithium during the open-label phase, rather than abruptly discontinuing the dose. However, these studies have several methodological limitations. In the depression study, comparisons between lithium and lamotrigine are problematic because of the unbalanced design and because the *a priori* primary efficacy analysis combined lamotrigine 200 mg/day and 400 mg/day. As with other studies, in both studies patients with co-morbid anxiety disorders (except for generalized anxiety disorder), substance use disorders, or those who were currently suicidal were excluded. In addition, both studies employed an 'enriched' double-blind discontinuation design, and only about half of patients entering the open-label phase were randomized with completion rates of ≤20% (Tabs 1 and 2).

Lithium and divalproex (moderately enriched)

The efficacy of lithium in the maintenance treatment of bipolar disorder was also compared with divalproex in a lithium-divalproex 'enriched' group with rapid cycling Bipolar I or II disorder. Although a history of at least one episode of hypomania, mania, or a mixed episode within 3 months of the study was required (Tab. 1), 58% of patients presented with depression, 36% with hypomania/mania/mixed state, and 7% with euthymia at the screening visit [18].

The primary analysis did not find significant differences between lithium and divalproex in the time to treatment for a mood episode, the time to premature discontinuation for any reason, the time to treatment for depression, and the time to treatment for a hypomanic/manic/mixed episode. The rates of mood episode relapse were also similar between the two groups (Tab. 2).

Several aspects of the study design were innovative. The open-label stabilization extended up to 6 months, which was longer than any of the previously conducted maintenance studies. The results from the combination of divalproex and lithium, two commonly used treatments for bipolar disorder, are likely to be clinically meaningful. The 20-month duration of the maintenance phase of this study was also longer than most recently conducted maintenance studies in bipolar disorder. More importantly, this study included 62% of patients with Bipolar II disorder that other recent maintenance studies did not include. However, the study had several limitations: 1) the sample size was modest; 2) lithium levels were kept at a minimum of 0.8 mEq/L and divalproex levels at a minimum of 50 μg/ml, which might have disadvantaged the divalproex arm; 3) because the combination of lithium and divalproex possessed better acute and continuation efficacy for mania/hypomania than depression, more patients with depressive episodes not responsive to the combination might have been excluded from the maintenance phase; and 4) this study only included patients with rapid cycling. Therefore, the results might not be applicable to other populations.

Agents only studied in manic episodes

The trend of using mania as a study index episode for maintenance trials has continued unabated in recent years (Tabs 1 and 2). In addition to the limitations of relapse prevention trials, the data from a manic index mood episode may not be applicable to patients presenting with a depressive episode [10] (although lithium and lamotrigine have been studied in both index mood episodes, showing similar results regardless of the index mood state [16, 17]). However, a discussion of these mania studies may shed light on the efficacy of these agents in preventing depressive relapses.

Aripiprazole (highly enriched)

Aripiprazole is the second atypical antipsychotic to be studied in the maintenance treatment of bipolar disorder with a manic index mood episode [19]. This study used the most stringent criteria to date to define stability before randomization, i.e., YMRS total score ≤10 and MADRS score of ≤13 maintained for at least six consecutive weeks (Tabs 1 and 2). At the end of week 26, the time to relapse was significantly longer for the aripiprazole group. A secondary analysis determined that aripiprazole was superior to placebo in delaying

manic relapse, but no significant difference was observed in the time to depressive relapse (Tab. 2). Overall, aripiprazole-treated patients had significantly fewer mood relapses and manic relapses than placebo-treated patients, but there was no significant difference in rates of depressive relapses between the two groups (Tab. 2). The double-blind phase was extended to an additional 74 weeks. At the end of 100 weeks, the results were similar. In addition to other potential limitations (Tabs 1 and 2), this study was limited by the fact that only 36% of patients completed the open-label treatment, the lowest in a non-rapid cycling population.

Olanzapine (highly enriched)

Among antipsychotics, olanzapine is the most studied antipsychotic in the maintenance treatment of bipolar disorder [20–22]. The studied population was either 'enriched' with olanzapine monotherapy [21], olanzapine combination therapy with lithium [20], or olanzapine adjunctive therapy to mood stabilizers [22]. All olanzapine maintenance studies were extensions of acute mania studies in which only those patients who tolerated and responded to the treatments were randomized (Tabs 1 and 2). In the monotherapy study [21], the estimate median time to symptomatic relapses was significantly longer in patients treated with olanzapine compared to that of placebo (Tab. 2). For relapse into mania or depression alone, the estimated 25th percentile time to relapse was significantly longer in the olanzapine group than in the placebo. The rates of symptomatic relapse into any mood were significantly lower in olanzapine-treated patients. However, olanzapine was more effective in preventing manic relapse than depressive relapse (Tab. 2). Similar results were also observed among those subjects who received a combination of olanzapine and lithium during the open-label treatment [20]. Olanzapine and lithium did not significantly differ in the proportion of patients who had a depressive recurrence. However, significantly fewer olanzapine-treated patients had the recurrence of a manic or mixed episode. In contrast, in patients who had inadequate response to mood stabilizer during the first two weeks, the efficacy of adjunctive olanzapine to mood stabilizer was not robust even in the prevention of manic relapses [22] (Tab. 2).

These three olanzapine maintenance studies possess unique features. In the monotherapy study [21], the time to symptomatic relapse was used as the primary outcome measure, which may detect patients with symptoms before clinical intervention. In addition, hazard ratios were calculated to quantify the magnitude of difference between olanzapine and placebo in the odds of relapsing into a mood episode. In the lithium-olanzapine enriched study [20], a 4-week period of discontinuation was employed to minimize the withdrawal effect associated with the discontinuation of lithium and olanzapine. Potential limitations of the three studies are summarized in Tables 1 and 2.

Lithium, divalproex, or both (mildly enriched)

In the first maintenance study employing modern methods such as DSM-III diagnostic criteria for bipolar disorder, time to intervention as a primary outcome, and survival analysis methodology [23], the efficacy of divalproex was compared with lithium and placebo in Bipolar I patients who had a recent manic episode (Tab. 1). In the primary analyses, the time to development of any mood episode did not differ significantly among the treatment groups, although a trend was observed favoring divalproex over lithium ($p = 0.06$). Some secondary outcome measures also favored divalproex over lithium (Tab. 2).

One should interpret these negative results cautiously. First, the lithium group had a larger proportion of patients who dropped out of the study due to intolerance or to non-compliance with treatment. Second, fewer patients were randomized to the lithium group than to divalproex because of a 2:1:1 ratio of assignment, which reduced the power for lithium-placebo comparisons. Another factor that might have contributed to the surprisingly good outcomes in the placebo group was that patients with milder forms of bipolar disorder were selected for this group; thus, the disease burden was milder in the randomized patients than the non-randomized. Although the treatment during the open-label phase was at the discretion of clinicians, the sample was somewhat 'enriched' with divalproex, lithium, or both.

Conclusions and clinical implications

Data from these relapse prevention trials have shown that lamotrigine and lithium are effective in preventing depressive and manic relapses, respectively, regardless of the index mood episode polarity. Both olanzapine and aripiprazole are effective in preventing manic, but not depressive relapses when patients present with mania. The efficacy of divalproex in maintenance treatment requires further study. Obviously, the generalizability of these results is limited, not only because of differences inherent to each individual trial, but also because of the universal exclusion of patients with co-morbid conditions such as substance use disorders, anxiety disorders, and those with severe suicidality. Therefore, when choosing an agent for maintenance treatment of bipolar disorder, it should always be kept in mind that most maintenance studies to date only enrolled patients with Bipolar I disorder during a manic episode, and that all these results were biased in some way because of the relapse prevention design.

Although results from a relapse prevention study are not generalizable, this design will continue to be used because of regulatory requirements for approving new drugs or indications, commercial interests in bipolar research, and the feasibility of conducting a study [11]. There will never be a perfect study even with a prophylaxis design. Future studies should not only focus on efficacy, but should also strive for improved generalizability. The generalizability of a

relapse prevention study can be increased by broadening the inclusion criteria of the study and using less 'enriched' treatments during the open-label phase.

Furthermore, the restricted inclusion criteria of most published studies to date can often exclude patients with severe illness. Because placebo response rates in less-impaired patients are higher, it is difficult to detect differences between a study drug and placebo if only less severely ill patients are enrolled into the study. However, the inclusion of the severely ill may pose legal and ethical challenges. Use of randomized add-on designs such as those employed in the epilepsy studies may help relieve this problem [11]. Gradual discontinuation of open-label drugs may reduce 'false' efficacy of the studied drugs whether it is a crossover or a parallel scheme. Because the polarities of mood episodes are different from the index mood episode at different times after randomization [9], the duration of any study – prophylaxis or relapse prevention – should last long enough (at least 12 months), and analyses should be divided into the first 6 months and afterward so that a spectrum of efficacy in the prevention of early and later relapses can be demonstrated. These early and late relapses may represent different pathological processes. In terms of analysis, the time to event survival analysis data and rate of occurrence of manic or depressive episodes should be reported; this will provide information on how fast, how often, and what kind of mood episodes occur. In addition, hazard ratios should be calculated to quantify the magnitude of difference between a drug and placebo on the odds of relapsing into a mood episode.

In conclusion, it is clear that more longitudinal maintenance studies for bipolar depression are urgently needed, especially in patients presenting with an index depressive episode. Achieving adequate methodological rigor without sacrificing overall study feasibility has become an important scientific focus.

Acknowledgements
This manuscript is supported in part by the National Institute of Mental Health P20 MH-66054 and the NARSAD.

References

1 Judd LL, Akiskal HS, Schettler PJ, Endicott J, Maser J, Solomon DA, Leon AC, Rice JA, Keller MB (2002) The long-term natural history of the weekly symptomatic status of bipolar I disorder. *Arch Gen Psychiatry* 59: 530–537
2 Judd LL, Akiskal HS, Schettler PJ, Coryell W, Endicott J, Maser JD, Solomon DA, Leon AC, Keller MB (2003) A prospective investigation of the natural history of the long-term weekly symptomatic status of bipolar II disorder. *Arch Gen Psychiatry* 60: 261–269
3 Calabrese JR, Hirschfeld RMA, Frye MA, Reed ML (2004) Impact of depressive symptoms compared with manic symptoms in bipolar disorder: Results of a U.S. community-based sample. *J Clin Psychiatry* 65: 1499–1504
4 MacQueen GM, Young LT, Joffe RT (2001) A review of psychosocial outcome in patients with bipolar disorder. *Acta Psychiatr Scand* 103: 163–170
5 MacQueen GM, Young LT, Robb JC, Marriott M, Cooke RG, Joffe RT (2000) Effect of number of episodes on wellbeing and functioning of patients with bipolar disorder. *Acta Psychiatr Scand* 101: 374–381
6 Goodwin F, Jamison K (eds): (2007) *Manic-depressive illness: bipolar disorder and recurrent depression*, 2nd edition. Oxford University Press, New York

7 Isometsä ET, Henriksson MM, Aro HM, Lönnqvist JK (1994) Suicide in bipolar disorder in Finland. *Am J Psychiatry* 151: 1020–1024
8 Dilsaver SC, Chen YW, Swann AC, Shoaib AM, Tsai-Dilsaver Y, Krajewski KJ (1997) Suicidality, panic disorder and psychosis in bipolar depression, depressive-mania and pure- mania. *Psychiatry Res* 73: 47–56
9 Calabrese JR, Goldberg JF, Ketter TA, Suppes T, Frye M, White R, DeVeaugh-Geiss A, Thompson TR (2006) Recurrence in bipolar I disorder: a *post hoc* analysis excluding relapses in two double-blind maintenance studies. *Biol Psychiatry* 59: 1061–1064
10 Shapiro DR, Quitkin FM, Fleiss JL (1989) Response to maintenance therapy in bipolar illness. Effect of index episode. *Arch Gen Psychiatry* 46: 401–405
11 Calabrese JR, Rapport DJ, Shelton MD, Kimmel SE (2001) Evolving methodologies in bipolar disorder maintenance research. *Br J Psychiatry* 178 (Suppl 41): S157–163
12 Prien RF, Klett CJ, Caffey EM Jr (1973) Lithium carbonate and imipramine in prevention of affective episodes: A comparison in recurrent affective illness. *Arch Gen Psychiatry* 29: 420–425
13 Prien RF, Caffey EM Jr, Klett CJ (1973) Prophylactic efficacy of lithium carbonate in manic-depressive illness. *Arch Gen Psychiatry* 28: 334–337
14 Prien RF, Klett CJ, Caffey EM Jr (1974) Lithium prophylaxis in recurrent affective illness. *Am J Psychiatry* 131: 198–203
15 Prien RF, Kupfer DJ, Mansky PA, Small JG, Tuason VB, Voss CB, Johnson WE (1984) Drug therapy in the prevention of recurrences in unipolar and bipolar affective disorders. *Arch Gen Psychiatry* 41: 1096–1104
16 Calabrese JR, Bowden CL, Sachs G, Yatham LN, Behnke K, Mehtonen OP, Montgomery P, Ascher J, Paska W, Earl N et al (2003) A placebo-controlled 18-month trial of lamotrigine and lithium maintenance treatment in recently depressed patients with bipolar I disorder. *J Clin Psychiatry* 64: 1013–1024
17 Bowden CL, Calabrese JR, Sachs G, Yatham LN, Asghar SA, Hompland M, Montgomery P, Earl N, Smoot TM, DeVeaugh-Geiss J (2003) A placebo-controlled 18-month trial of lamotrigine and lithium maintenance treatment in recently manic or hypomanic patients with bipolar I disorder. *Arch Gen Psychiatry* 60: 392–400
18 Calabrese JR, Shelton MD, Rapport DJ, Youngstrom EA, Jackson K, Bilali S, Ganocy SJ, Findling RL (2005) A 20-month, double-blind, maintenance trial of lithium *versus* divalproex in rapid-cycling bipolar disorder. *Am J Psychiatry* 162: 2152–2161
19 Keck PE, Calabrese JR, McIntyre RS, McQuade RD, Carson WH, Eudicone JM, Carlson BX, Marcus RN, Sanchez R; for the Aripiprazole Study Group (2007) Aripiprazole monotherapy for maintenance therapy in bipolar I disorder: a 100-week, double-blind study *versus* placebo. *J Clin Psychiatry* 68: 1480–1491
20 Tohen M, Greil W, Calabrese JR, Sachs GS, Yatham LN, Oerlinghausen BM, Koukopoulos A, Cassano GB, Grunze H, Licht RW et al (2005) Olanzapine *versus* lithium in the maintenance treatment of bipolar disorder: a 12-month, randomized, double-blind, controlled clinical trial. *Am J Psychiatry* 162: 1281–1290
21 Tohen M, Calabrese JR, Sachs GS, Banov MD, Detke HC, Risser R, Baker RW, Chou JC, Bowden CL (2006) Randomized, placebo-controlled trial of olanzapine as maintenance therapy in patients with bipolar I disorder responding to acute treatment with olanzapine. *Am J Psychiatry* 163: 247–256
22 Tohen M, Chengappa KN, Suppes T, Baker RW, Zarate CA, Bowden CL, Sachs GS, Kupfer DJ, Ghaemi SN, Feldman PD et al (2004) Relapse prevention in bipolar I disorder: 18-month comparison of olanzapine plus mood stabilizer v. mood stabilizer alone. *Br J Psychiatry* 184: 337–345
23 Bowden CL, Calabrese JR, McElroy SL, Gyulai L, Wassef A, Petty F, Pope HG Jr, Chou JC, Keck PE Jr, Rhodes LJ et al (2000) A randomized, placebo-controlled 12-month trial of divalproex and lithium in treatment of outpatients with bipolar I disorder. Divalproex Maintenance Study Group. *Arch Gen Psychiatry* 57: 481–489
24 Calabrese JR, Vieta E, El-Mallakh R, Findling RL, Youngstrom EA, Elhaj O (2004) Mood state at study entry as predictor of the polarity of relapse in bipolar disorder. *Biol Psychiatry* 56: 957–963
25 Ghaemi SN, Pardo RR, Hsu DJ (2004) Strategies for preventing the recurrence of bipolar disorder. *J Clin Psychiatry* 65 (Suppl 10): 16–23
26 Gyulai L, Bowden CL, McElroy SL, Calabrese JR, Petty F, Swann AC, Chou JC, Wassef A, Risch CS, Hirschfeld RM et al (2003) Maintenance efficacy of divalproex in the prevention of bipolar depression. *Neuropsychopharmacology* 28: 1374–1382