Methodological Challenges in Long-term Trials for Bipolar Depression

Joseph R. Calabrese, MD
Bipolar Disorders Research Chair
Bipolar Disorders Research Center
Case Western Reserve School of Medicine
Joseph.calabrese@UHhospitals.org

Disclosure / Conflict of Interest

Joseph R. Calabrese

- Grant Support
  - Abbott, AstraZeneca, Bristol-Myers Squibb, GlaxoSmithKline, Janssen, Lilly, Pfizer, Wyeth
- Consultant/Advisory Boards/Speaker
  - Abbott (>10k), AstraZeneca (>10k), Bristol-Myers Squibb, Cephalon, GlaxoSmithKline (>10k), Janssen, Pfizer, Repligen, Servier, Solvay, Wyeth
- Speakers Bureau
  - Speaker trainer only (AstraZeneca & GlaxoSmithKline)
- Equity Interests **- none**
- Federal
  - P20 MH-66054 (Calabrese) – Developing Interventions and Services Research Center
  - DoD PE 611102 (Calabrese) Risk and Resilience Factors
  - Health Resources & Services Admin 1 C76 HF 00502-01 (Calabrese) – Bipolar Disorder Accompanied by Substance Use Disorders
- *Off label uses of medications will be discussed.*

Editorial

Bipolar Drug Development: Are We Getting Closer to the Real World?

JOSEPH R. CALABRESE, M.D.
DAVID E. KEMP, M.D.

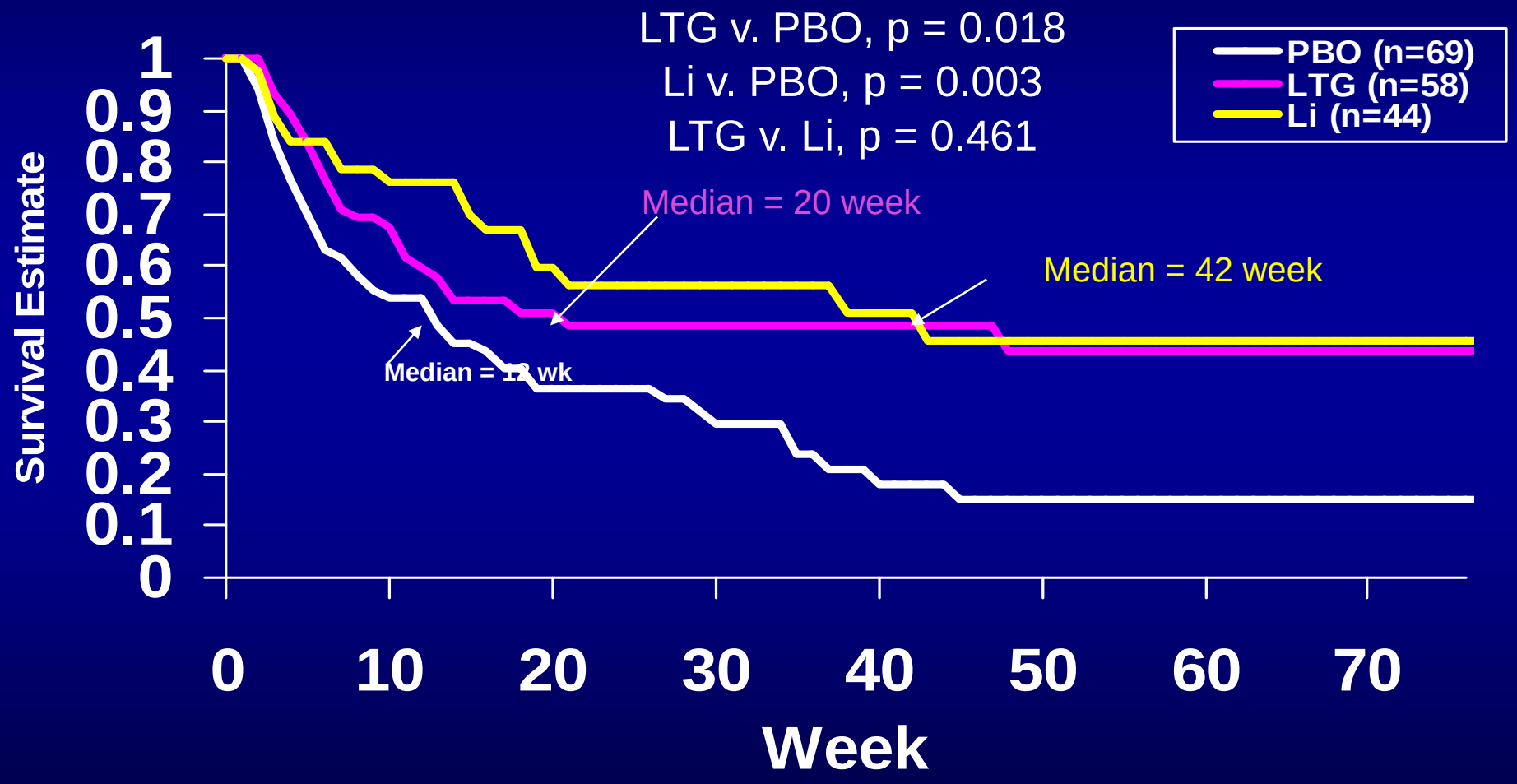
"The past 20 years of industry-sponsored treatment trials…has brought …few new treatments for bipolar depression, where our patients live most of their symptomatic lives."

Calabrese JR, Kemp DE. American Journal of Psychiatry. 2008 October;165(10): 1234-36.

Untested Assumption
‘Avoid the phase of illness with undocumented acute efficacy’

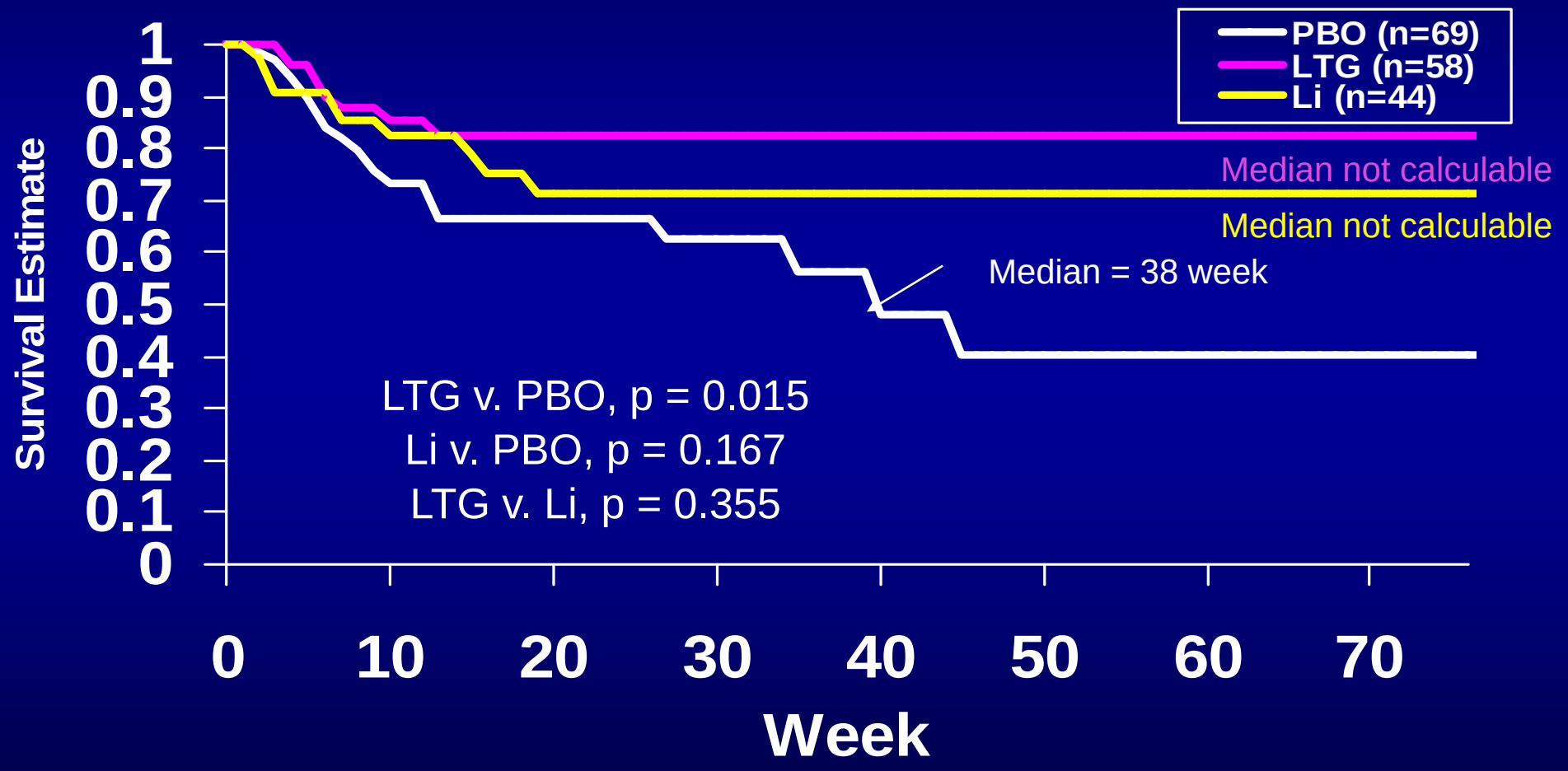
- Recently depressed patients should be excluded from relapse/recurrence prevention studies.
- We have avoided studying the phase of illness where our patients live the majority of their symptomatic lives.

Maintenance Study in Recently Manic Patients - Time to Intervention for a Mood Episode *-



Bowden C, Calabrese J, Sachs G, et al. Arch Gen Psychiatry. 2003:60:392-400. *Primary outcome.

Despite Lack of Antimanic Efficacy, Lamotrigine, but not lithium, delayed time to intervention for depression in the recently manic



Bowden C, Calabrese J, Sachs G, et al. Arch Gen Psychiatry. 2003:60:392-400.

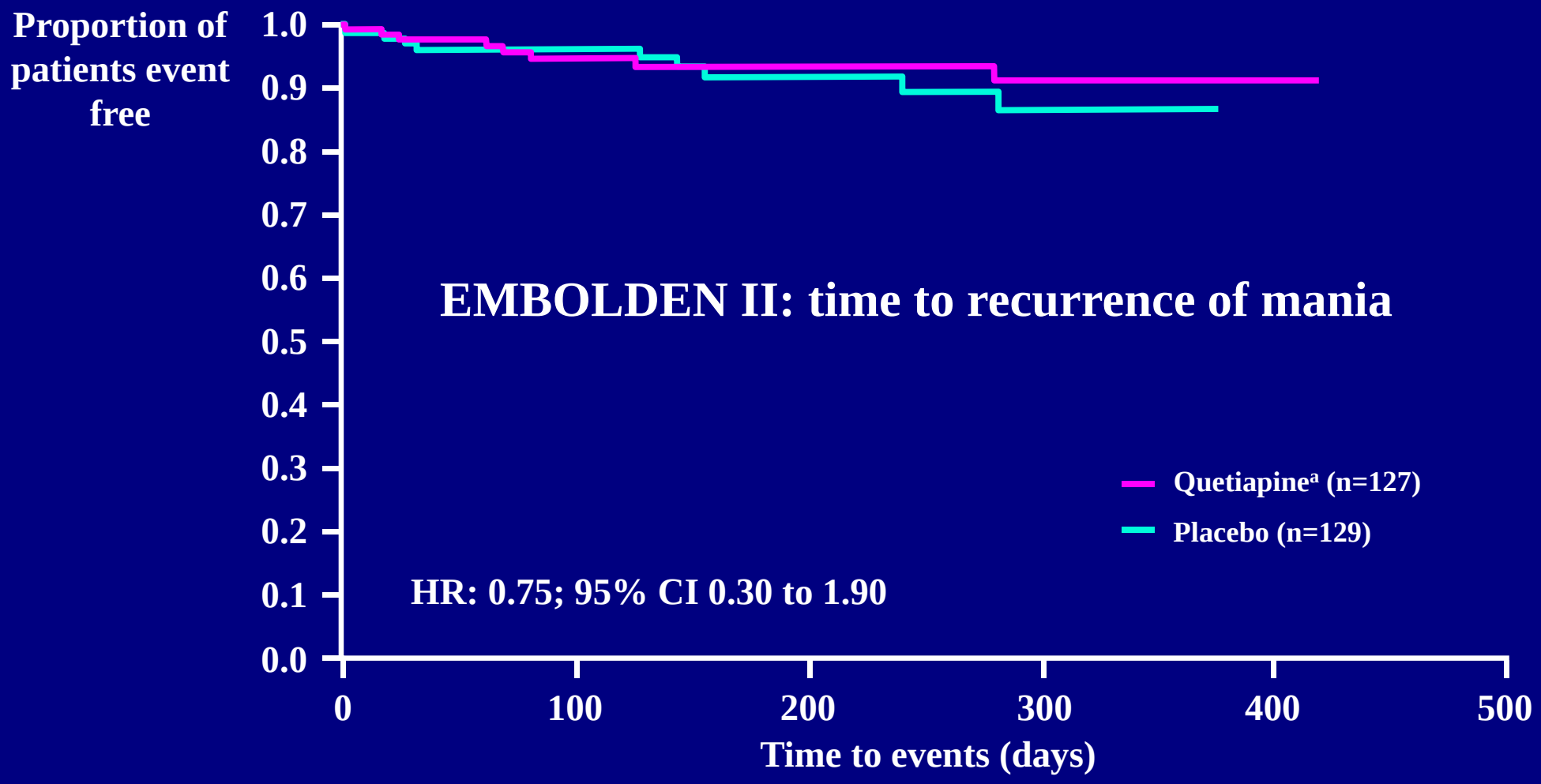
‘The Not So Obvious’

- The exclusion of recently depressed patients from relapse prevention studies:
  - Inflates the observed experimental effect size if the drug stabilizes mood ‘from above baseline’

    and

  - Deflates the effect size if the drug stabilizes mood ‘from below baseline’

The Opposite: Example deflated effect size in mania where recently manic patients were excluded



ITT population

[A] Combined group of patients randomised to 300 or 600 mg

Pivotal Maintenance Studies with Index Episodes Depression

	Prien Phase III Study [1]	LTG Study [2]
Design	Li vs. Imi vs. Pbo	Ltg vs. Li vs. Pbo
N	18 vs. 13 vs. 13	221 vs. 121 vs. 121
Duration (yrs)	2	1.5
Sites	18	79
Recent hospitalization	required	not required
Primary Outcome	% relapse rates	time to treatment
Lithium Levels mEq/L	0.5 - 1.4	0.8 - 1.1

[1] Prien et al 1973 and 1974. [2] Calabrese et al 2003. (only BP I included in both)

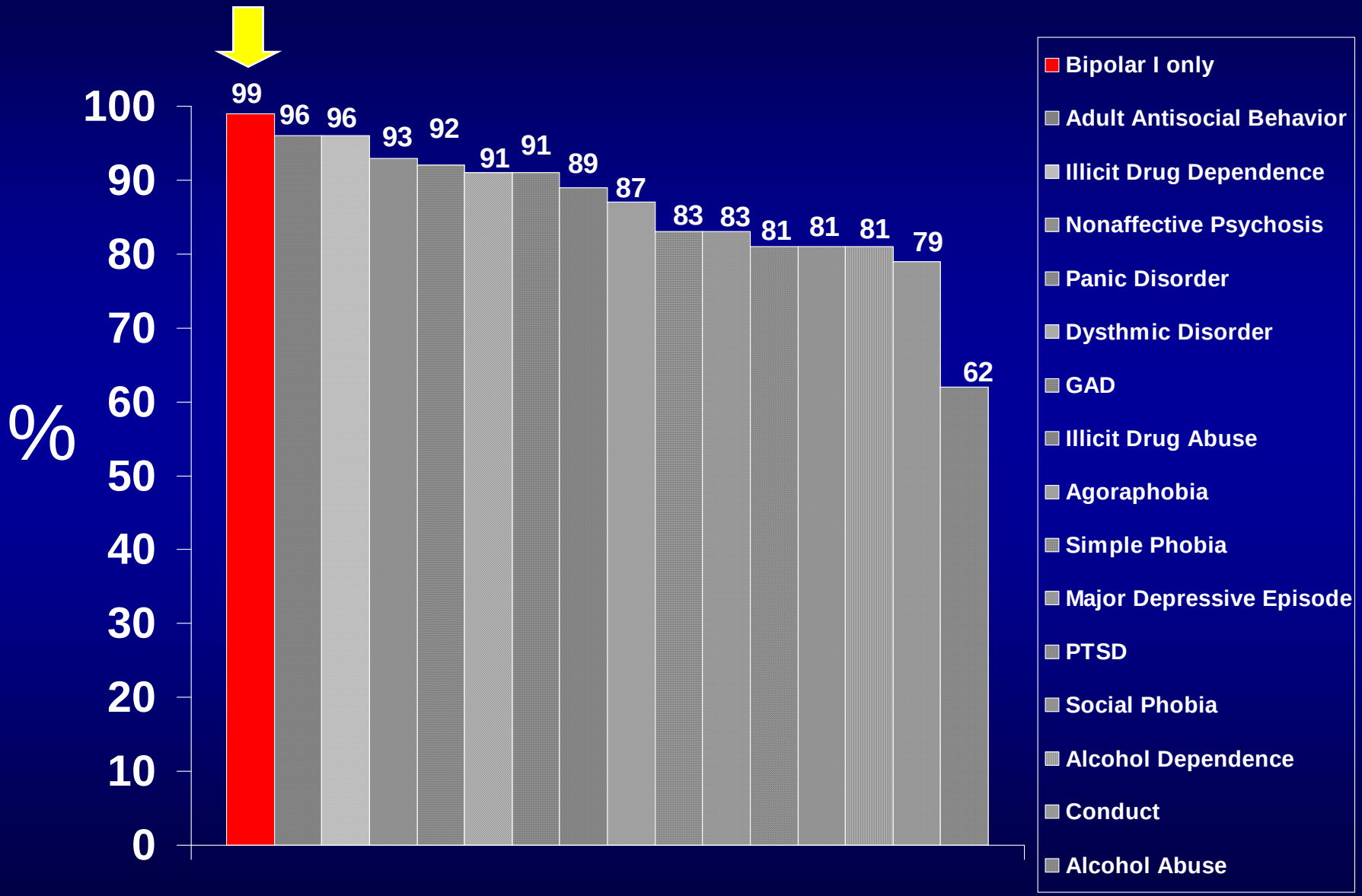
Recently Depressed Patients Included in All 5 Bipolar Recurrence Prevention Quetiapine trials (n = 3,339)

- 5 long-term bipolar I disorder recurrence prevention trials
  - 2 Continuation Phases from EMBOLDEN I and II (BP I or II)
  - 2 Adjunctive Studies with Lithium or Valproate (BP I only)
  - 1 Monotherapy Study: quetiapine vs. lithium vs. Pbo (BP I only)

Why do old and new drugs not perform as well as their data suggest they should?

- Estimates of true population therapeutic effect sizes are being artificially inflated.
  - **Old lithium studies** also inflated effects sizes by excluding patients {rapid cyclers, mixed states, type II, and substance abusers}.
  - **Recent studies** inflate effect size by: 1) enriched relapse prevention designs, 2) excluding the recently depressed, 3) substance abusers, and the 4) highly anxious.

Why is it so difficult to design and conduct bipolar studies?



Kessler and Walters. In: Tsuang and Tohen, eds. *Textbook of Psychiatric Epidemiology.* 2nd ed. 2002

Management of Heterogeneity in Bipolar Treatment Trials

- Substance use and anxiety disorders are the 2 most commonly comorbidities and are completely or partially excluded from most clinical trials (1-3).
- Substance use and anxiety comorbidity predicts non-response (4-10) and lack of tolerability to agents that delay future episdoes (8,11-13).
- As a result, variance decreases and assay sensitivity improves, but at the expense of generalizability.

1. Regier et al., JAMA 1990; 2. Merikangas et al., AGP 2007; 3. Gao et al., JAD 2008. 4. Baethge et al., AJP 2005; 5. Salloum et al., Bipolar Disord 2002; 6. Kolodziej et al., JAD 2008; 7. Frank et al., AGP 2002; 8. Feske et al., 2000; 9. Simon et al., AJP 2004; 10. Weiss et al., 2005; 11. Goldberg and Whiteside, JCP 2002; 12. Manwani et al., JCP 2006; 13. Gao et al., APA 2008).

Recent Bipolar Depression Studies Excluded Substance Use (1-12 months) and Anxiety Disorders (variably)

	Alcohol abuse	Alcohol dependence	Drug abuse	Drug dependence	Other Axis I Disorders
Olanzapine BPI depression (1)	----	3 months	----	3 months	Not primary focus
Quetiapine BPI and BPII depression (2,3)	12 months	12 months	12 months	12 months	Not primary focus in 6 months
Aripiprazole BPI depression (4)	3 months	6 months	3 months	6 months	OCD in 3 months
Lamotrigine BPI and BPII depression (5,6,7)	1 months	12 months	1 months	12 months	PD, OCD, SP in 12 months

1. Tohen et al., AGP 2003; 2. Calabrese et al., AJP 2005; 3. Thase et al., J Clin Psychopharmacol 2007; 4. Thase et al., J Clin Psychopharnacol 2008; 5. Calabrese et al., JCP 1999; 6. Calabrese et al., Bipolar Disord 2008. Geddes et al. BJP 2009.

Prevalence of Bipolar I and II disorders

	1 month	6 Month	12 Months	Lifetime
BPI BPII ECA (1)	0.4 0.2	0.6 0.3	0.6 na	0.8 0.5
BPI BPII NCS-R (2)	n/a n/a	n/a n/a	**0.6** **0.8**	**1.0** **1.1**

ECA, Epidemiologic Catchment Area study; NCS, National Comorbidity Survey; NCS-R. National Comorbidity Survey-Replication

1. Regier et al., JAMA 1990; 2. Merikagas et al., AGP 2007;

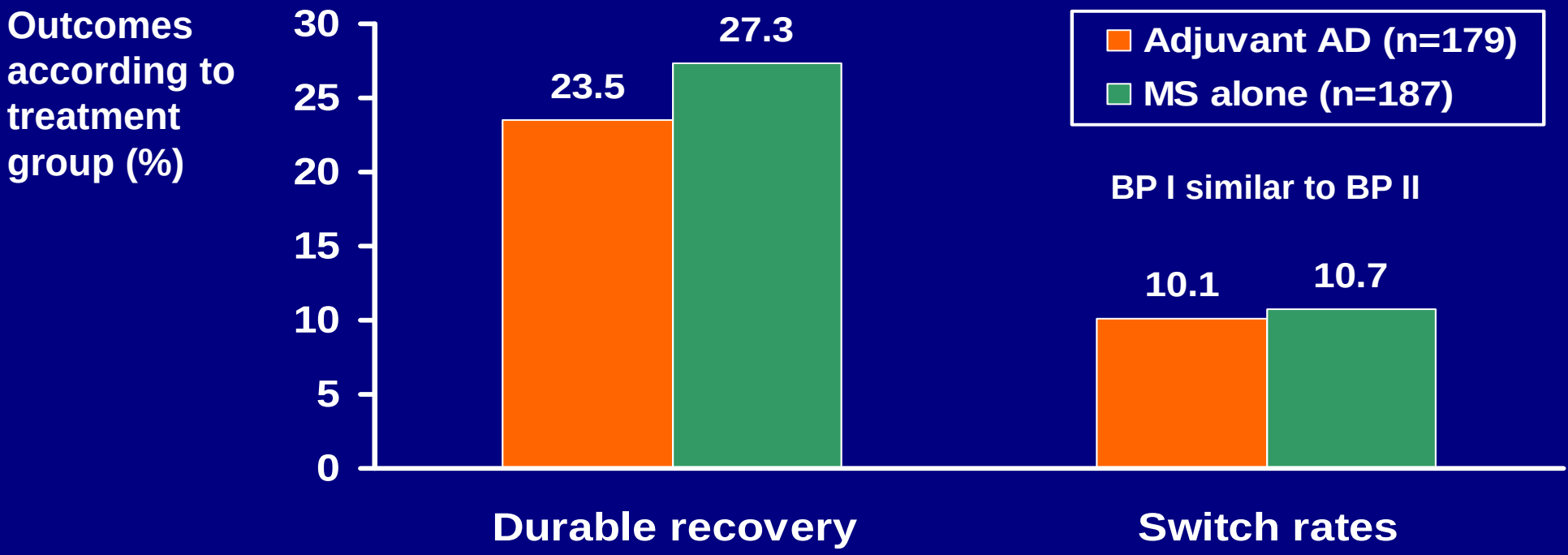
Quantifying the Impact on Generalizability

- Data from recent bipolar depression studies may not be generalizable to **16 -62%** of patients because of the exclusion of substance use and/or anxiety disorders.
- Worldwide, based on 12-month prevalence data, the health care needs of **14.7 - 56.7** million people are being ignored (22 - 85 million based on lifetime data).
- Within the US, the health care needs of **0.68 - 2.62** million people are being ignored based on 12-month prevalence data (1.0 - 3.9 million based on lifetime data).

Gao, Keming et al NCDEU presentation. 2007.

Study failure or success?

‘6 Month Randomized BP Depression Trial’



Durable recovery = 8 wks euthymia (no more than 2 depressive or 2 manic symptoms)
Switch = DSM criteria for hypomania or mania or required treatment. Mood stabiliser (MS) = lithium, carbamazepine, valproate or any approved atypical. Adjunctive antidepressant (AD) = bupropion or paroxetine.
Sachs G, Nierenberg A, Calabrese J, et al. New Engl J Med. 2007;356:1711-22.

The Challenge

- Changes in methodology need to be tested for feasibility before large-scale implementation, but how?
- There exists a need to move towards improved generalizability by employing fewer exclusion criteria.
- The additional heterogeneity should be and has been incremental.

Forcing the Issue of Sample Size and Heterogeneity (1)

- Drug development efforts for heart disease and diabetes manage heterogeneity by increasing sample size.
- These drug development efforts employ sample sizes that frequently exceed those employed in psychotropic drug development.

Forcing the Issue of Sample Size and Heterogeneity (2)

- Revenue generated from psychotropic drugs appear to be as high or higher.
- Should Congress be asked to change FDA guidelines to require fewer exclusion criteria, and to employ sample sizes capable of better generalizability?

Most Maintenance Methodology Has Employed Relapse Prevention Designs

- Stabilization (either open or blinded) followed by placebo-controlled discontinuation
- Randomized subjects have both tolerated and responded to the new treatment during the open stabilization phase, but not the active comparator.

Criticisms of Enriched Designs

- Outcome is biased in favor of the experimental agent and against the active comparator.
- The active comparator is more likely to exhibit more side effects.
- The active comparator is likely to appear less efficacious.

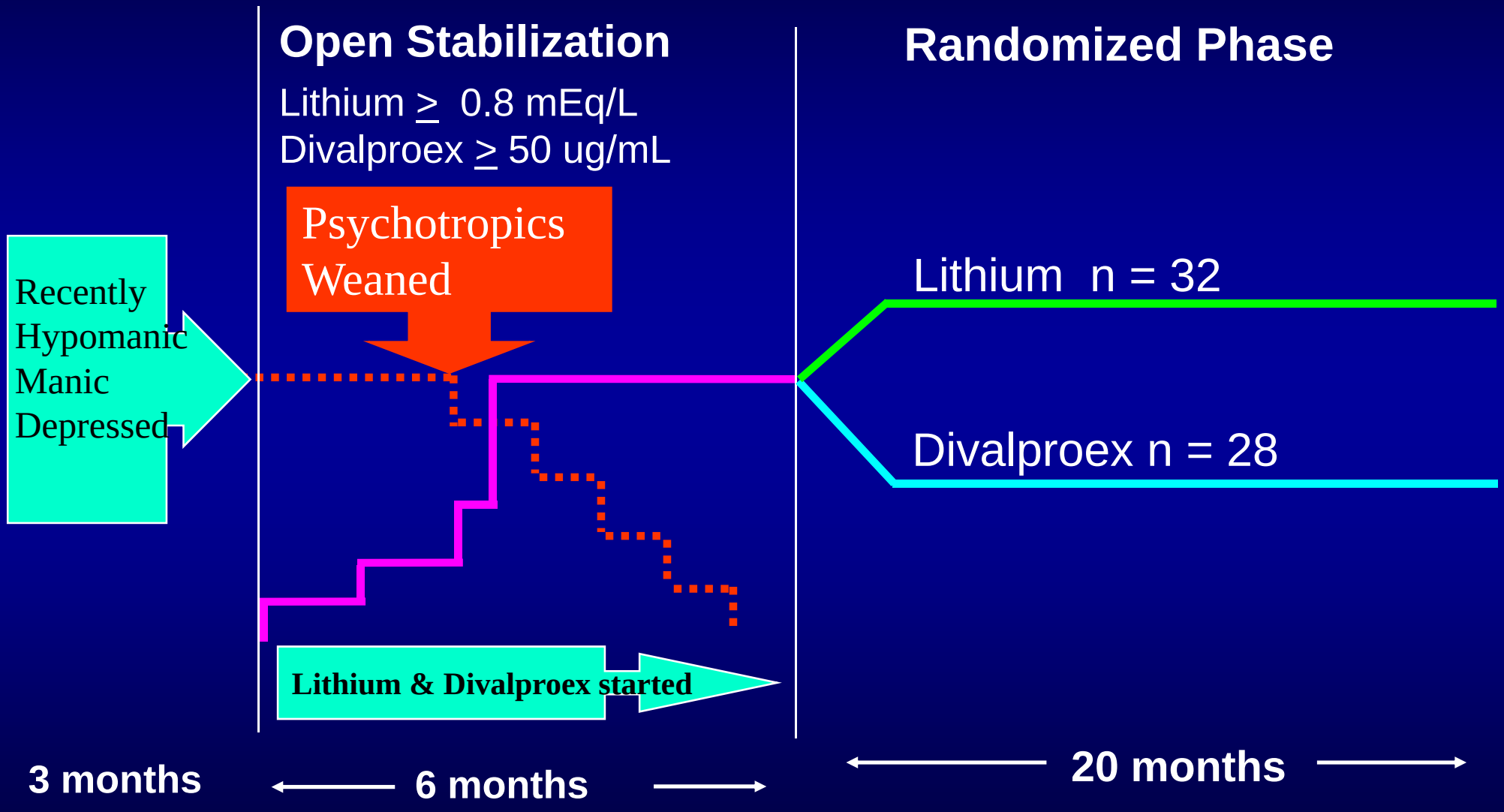
BP I Lithium Maintenance Studies

Enriched discontinuation	Baastrup et al., 1970 Hullen et al., 1972 Melia, 1972 Bowden et al., 2003. Calabrese et al., 2003.
Crossovers	Cundall et al., 1971
Non-enriched random assignment to parallel groups from a symptomatic state	Coppen et al., 1971 Stallone et al 1973 Prien et al., 1973 Bowden et al., 2000

Why Are Enriched Designs Used?

- Because minimally enriched designs have less statistical power and don’t separate from placebo.
- **The fix…**
  - Stabilization on combination therapy, both the experimental agent and the active comparator, which improves [1)] generalizability and [2)] randomization rates.

Li vs. Dvpx in Rapid Cycling Bipolar



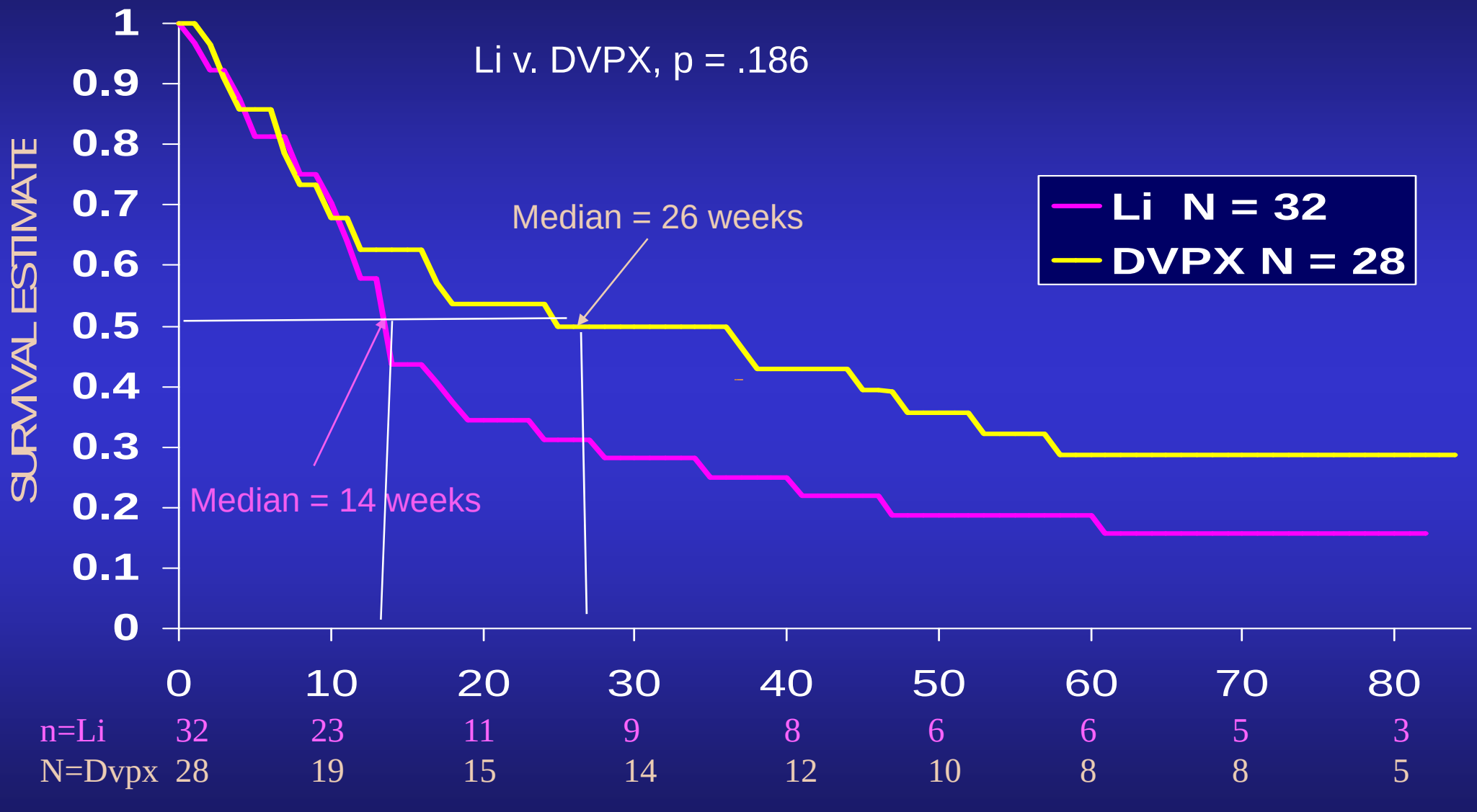
Calabrese et al., AJP 2005. R01 MH-50165

*stratified type I and II

Other Problems with Highly Enrich Designs

- Relatively little information can be obtained
  - during the first few weeks or months after randomization due to early hazards.
  - during the last 3-6 months of the blinded phase because small samples sizes.

A Typical Survival Curve Using Kaplan Meier Methodology

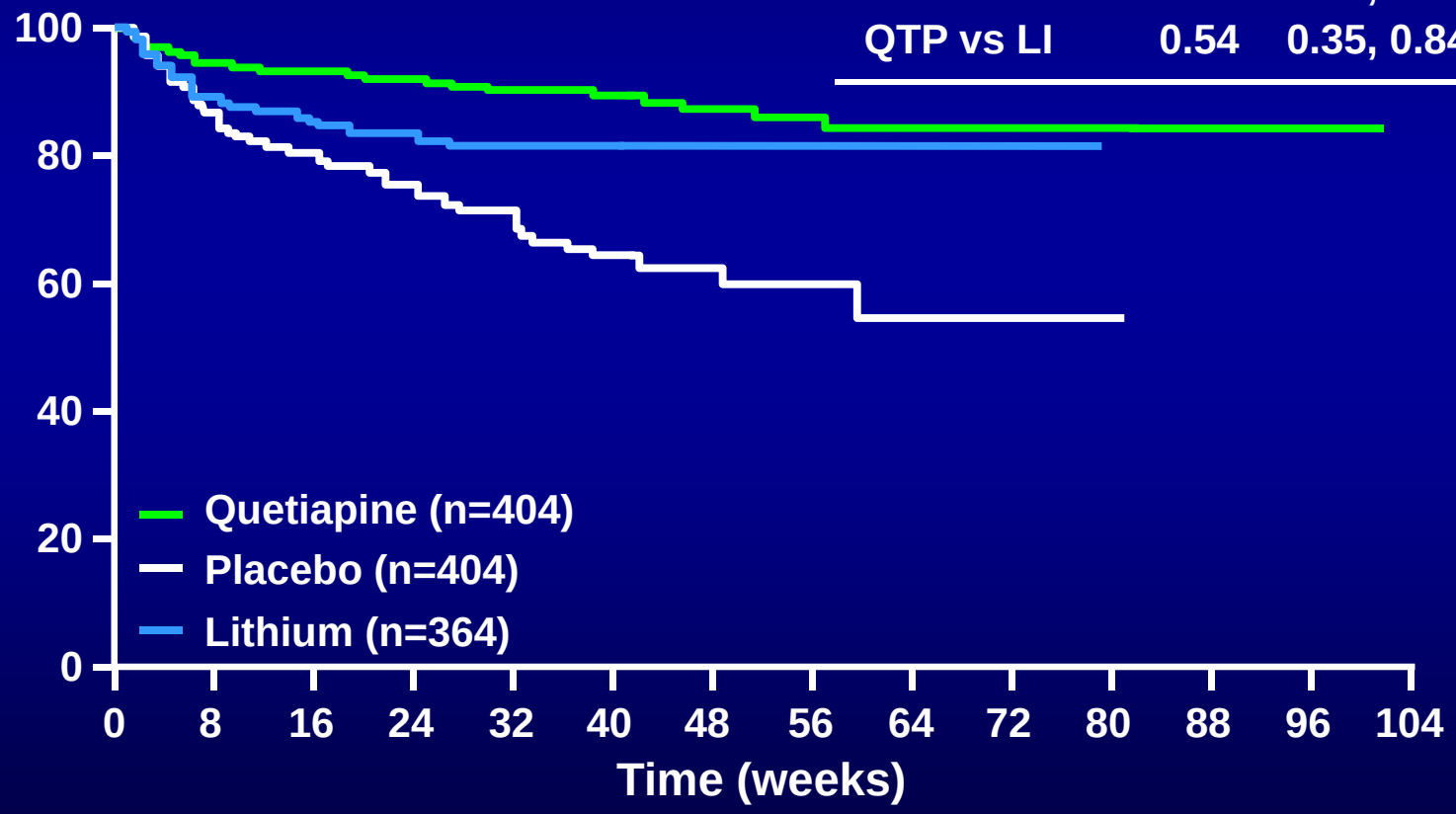


Calabrese et al. AJP. 2005.

Quetiapine monotherapy in Bipolar I (n=1,172): Time to Recurrence of Depressed Event

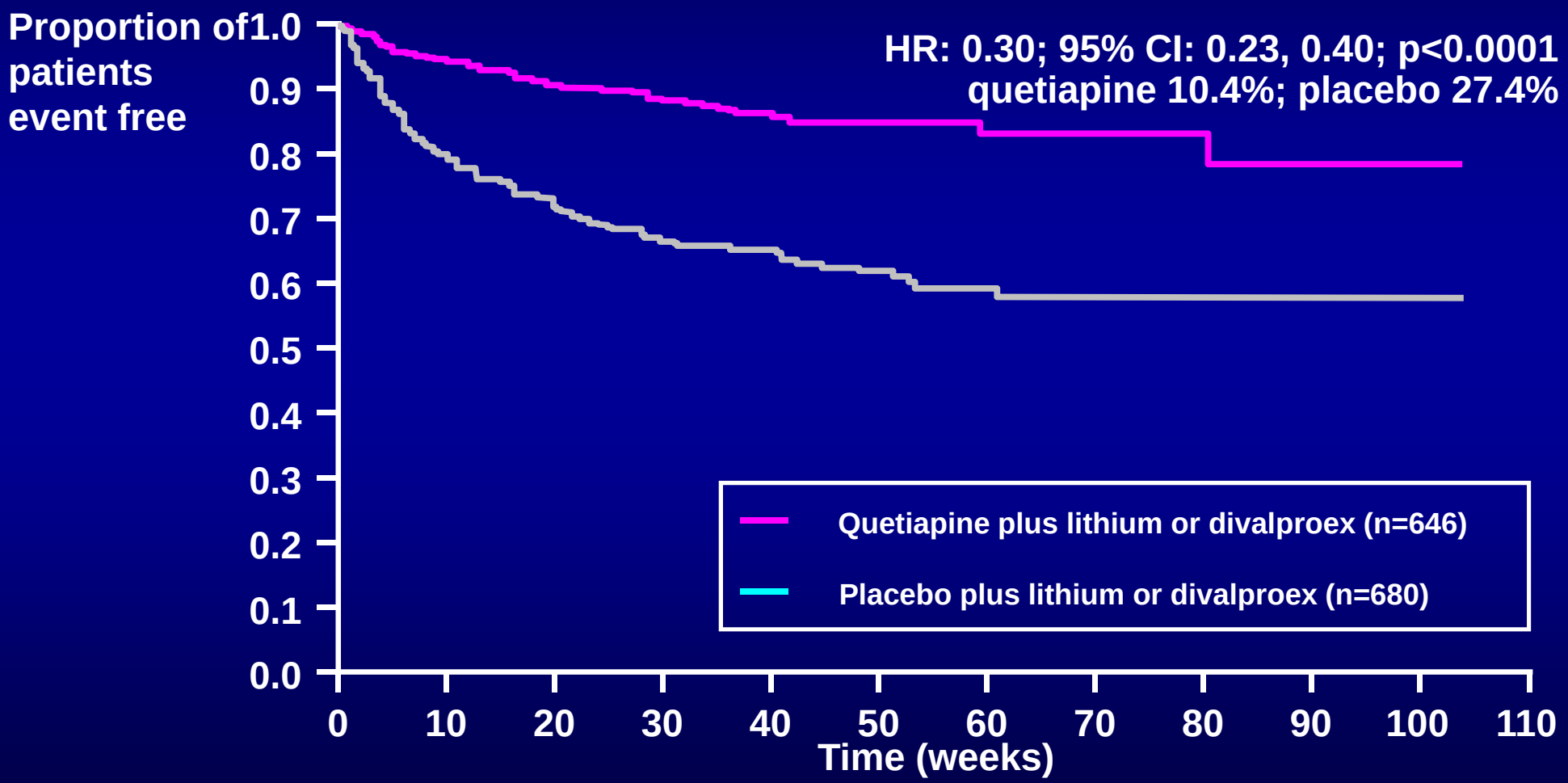
- No Patient Left Free of Depressive Episodes By End of Study -

	HR	95% CI	p-value
QTP vs PLA	0.30	0.20, 0.44	<0.001
LI vs PLA	0.59	0.42, 0.84	<0.01
QTP vs LI	0.54	0.35, 0.84	<0.01



ITT population

Adjunctive Quetiapine Prevention in BP I: Time to Depression
- Better Study Retention Due to Adjunctive Study Design -



ITT population

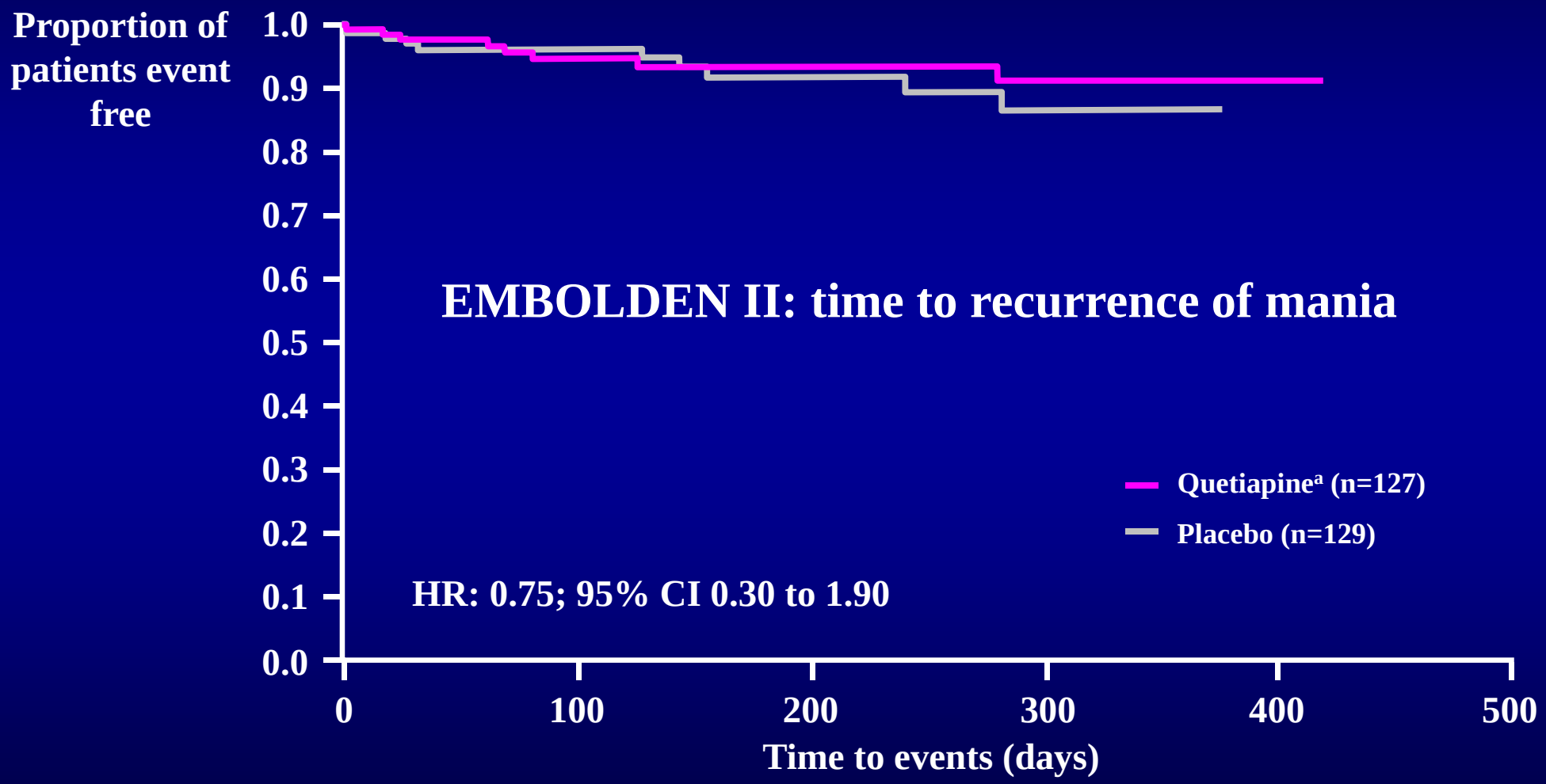
Mean (SD) dur. of exposure during rand. phase: quetiapine = 213 (184) days; placebo = 152 (163) days

Index polarity predicts the polarity of relapse on placebo during maintenance treatment

Index episode	Depression Relapse	Mania Relapse
Recently Depressed 2.4 : 1	39% 	16% 
Recently Manic or Hypomanic 1.4 : 1	30% 	41% 

Calabrese JR, Vieta E, El-Mallakh R, Findling RL, Youngstrom EA, Elhaj O, Gajwani P, Pies R: Mood state at study entry as predictor of risk of relapse and spectrum of efficacy in bipolar maintenance studies. Biol Psychiatry. 2004 Dec 15;56(12):957-73.

No Effect in Recurrence of Mania Because Recently Manic Patients Were Excluded from Study Entry



ITT population

[A] Combined group of patients randomised to 300 or 600 mg

Efficacy, Safety, and Tolerability

- Is that all there is? -

- The clinical value of ‘Predictability’
- Clinicians also want to know “what works, where, and when”.

Calabrese et al. ISBD Delhi, India January 27, 2008.
Stassen et al. 1993, 1996, 1999, & 2007. Szegedi et al. In Press.

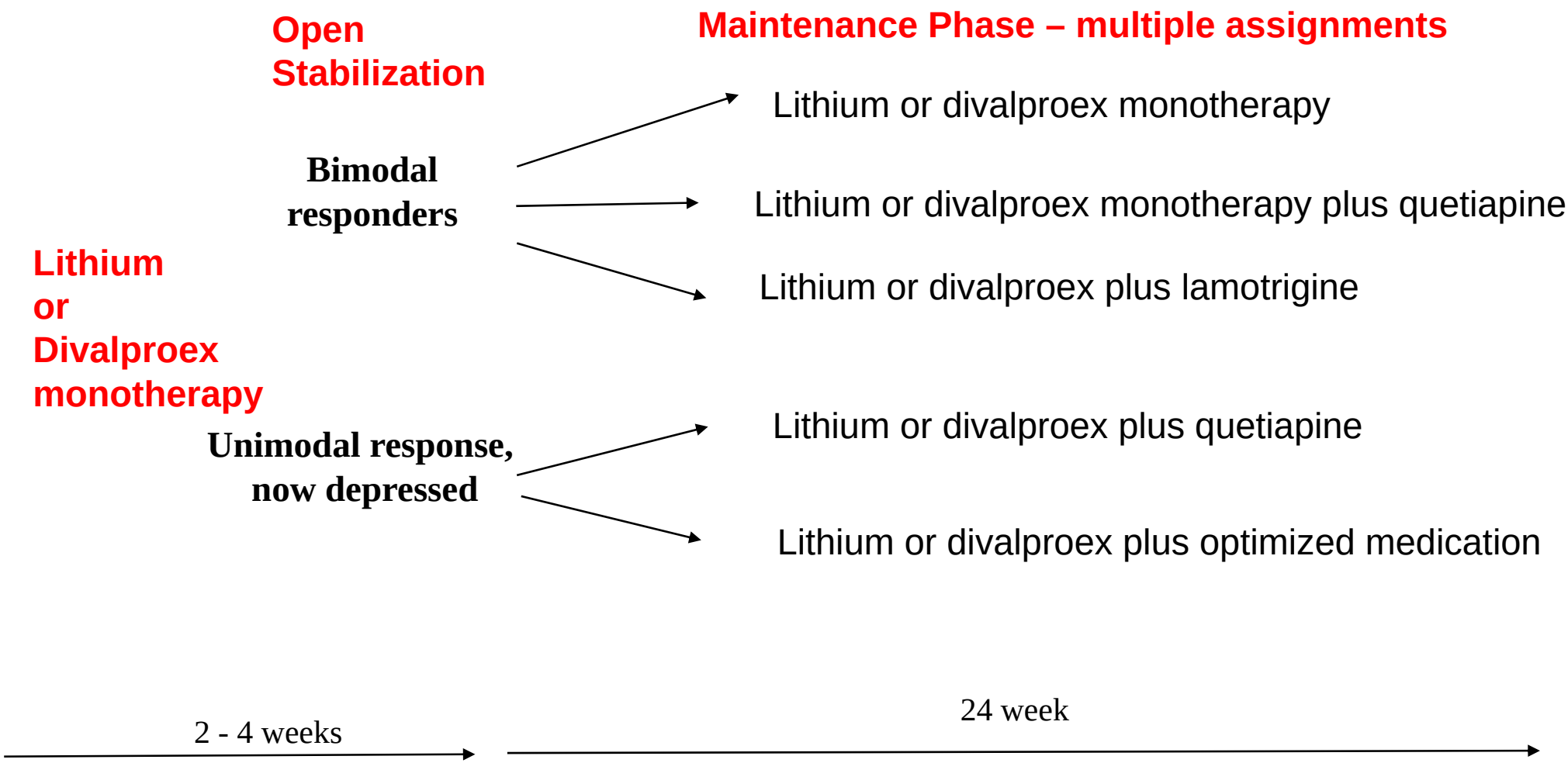
Potential Use of Adaptive Designs in Bipolar

- Highly recurrent illness
- Complex algorithms have been published, but have never been empirically evaluated.
- Adaptive Treatment Strategies are individually tailored treatments, with treatment type and dosage individualized for patient need such that clinical practice is mimicked.

Murphy SA et al. Drug Alcohol Depend. 2007 May Suppl 2:S24-30.

Sequential Multiple Assignment Randomization

- open stabilization followed by relapse/recurrence prevention -



Singh V, Thompson P, Martinez C, Bowden C, Calabrese J, Mintz J. P30 under review.

Conclusions

- The field has made incremental movement forward on the continuum of improved generalizability.
- Patients with BP type II, mixed states, and rapid cycling are now being included in maintenance studies.
- However, recently depressed patients are still be excluded from maintenance studies, which artificially inflates maintenance effect size.