

Word count: 3370

Number of tables: 3

Number of figures: 2

Original investigation

Initial Severity of Bipolar I Disorder Associated with the Efficacy of Olanzapine:

Individual Participant Data Meta-analysis of Five Placebo-Controlled Studies

Myrto T. Samara*, MD; Yair Goldberg*, PhD; Stephen Z. Levine, PhD; Toshi A. Furukawa, PhD;
John R. Geddes, MD; Andrea Cipriani, PhD; John M. Davis, MD; Stefan Leucht, MD

(*: joint first authors)

Author affiliations:

Myrto T. Samara: Department of Psychiatry and Psychotherapy, Technische Universität München, Munich, Germany; 3rd Department of Psychiatry, School of Medicine Aristotle University of Thessaloniki, Thessaloniki, Greece.

Yair Goldberg: Department of Statistics, Faculty of Social Sciences, University of Haifa, Haifa, Israel.

Stephen Z. Levine: Department of Community Mental Health, Faculty of Social Welfare and Health Sciences, University of Haifa, Haifa, Israel.

Toshi A. Furukawa: Departments of Health Promotion and Human Behavior and of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan.

John R. Geddes and Andrea Cipriani: Department of Psychiatry, University of Oxford, Oxford, UK; Oxford Health NHS Foundation Trust, Warneford Hospital, Oxford, UK.

John M. Davis: Psychiatric Institute, University of Illinois at Chicago, Chicago, IL; Maryland Psychiatric Research Center, University of Maryland, Baltimore, USA.

Stefan Leucht: Department of Psychiatry and Psychotherapy, Technische Universität München, Munich, Germany.

Corresponding author:

Myrto Samara, MD.

Department of Psychiatry and Psychotherapy, Technische Universität München,

Klinikum rechts der Isar, Ismaningerstr. 22, D-81675, Munich, Germany

Tel: +49-89-41404249; +30-69-80107450, Fax: +49-89-41404888; +30-23-10321090

e-mail: samaramyrto@gmail.com

Abstract

Background: The efficacy of antipsychotics across the initial severity range in acute mania remains unclear. Therefore, we decided to examine the influence of baseline severity on the efficacy of olanzapine.

Methods: We conducted an individual participant data (IPD) meta-analysis of five double-blind, randomized controlled trials comparing olanzapine versus placebo. 939 patients with acute mania associated with bipolar I disorder were included. The relationship between baseline and change scores on the Young Mania Rating Scale (YMRS) up to three weeks for olanzapine versus placebo groups was examined.

Findings: The interaction between baseline symptom severity and treatment was statistically significant ($p = 0.013$). The greater the baseline severity, the greater the magnitude of the differences between olanzapine and placebo. The mean YMRS score difference was 2.6 points for patients with a baseline score from 20 to 25, 4.7 points for patients with a baseline YMRS score around 30, and 8 points for patients with a baseline score up to 60.

Interpretation: Benefits of an antipsychotic drug like olanzapine can be expected for the full severity spectrum of patients likely to be treated for acute mania. However, less severely ill patients seem to benefit less in terms of efficacy but still may experience the same side-effects as more severely ill patients. Thus, clinicians and patients should carefully consider the benefit/risk ratio of olanzapine and its additional, prophylactic effect against relapse in the long-term. The generalizability of these results to other antipsychotics, trial designs, and medical conditions remains to be established.

Funding: None.

Key Words: bipolar, mania, efficacy, severity, individual participant data meta-analysis,
antipsychotic drugs, olanzapine

Introduction

Antipsychotic drugs are used to treat manic or mixed episodes associated with bipolar I disorder, both in the acute phase for symptoms' control and in the long-term for relapse prevention.¹⁻⁴ Nevertheless, the efficacy of some psychotropic agents has been recently called into question. Starting with antidepressants, some studies suggested they may be less efficacious for the milder spectrum of the disorder compared to placebo⁵⁻⁹ whereas others did not¹⁰⁻¹³ leading to a considerable debate among scientists and extreme titles in public newspapers.¹⁴⁻¹⁹ The impact on patients' adherence is still unknown but, given the extensive media coverage of the topic, it cannot be negligible. In response to this longstanding debate, a recent individual participant data (IPD) meta-analysis examined the association between symptom change and levels of initial symptom severity in the antipsychotic drug treatment of schizophrenia.²⁰ That study found that antipsychotics were efficacious across the whole initial symptom range in the psychopharmacological treatment of schizophrenia, but also that a statistically significant interaction between baseline symptom severity and efficacy of antipsychotics versus placebo was present.²⁰ An IPD meta-analysis in children with autism spectrum disorder treated with risperidone delivered similar results confirming the presence of the baseline symptom severity effect.²¹

In acute mania, studies examining the efficacy of antipsychotics versus placebo across the initial severity spectrum are limited and none has detected a baseline symptom severity effect.²²⁻²⁴ Nevertheless, these studies had important limitations. The first one²² pooled participants from four randomized trials which broke the randomization process. In addition, participants were

divided in two groups as high or low baseline symptom severity based on an arbitrary cut-off and all analyses just compared efficacy results between these two groups. The other two studies were meta-analyses that used study-level aggregated data instead of IPD.^{23,24} Therefore, a number of critical issues arose such as inconsistent handling of missing data, the use of different definitions and statistical analyses across studies, and the lack of ability to explore the role of same patient level effect modifiers across studies. These shortcomings can be addressed by using an IPD meta-analysis.^{25,26}

It is also important to ascertain whether the baseline symptom severity effect is a general pattern that occurs as a rationale or it is specific to some mental disorders like schizophrenia and autism, but not others. If more mildly ill manic patients did not respond or responded less to antipsychotic drugs, this would have implications for clinical practice such as using lower doses, better tolerable drugs or even not recommending medication for them. To our knowledge, this is the first IPD meta-analysis to examine the association between initial symptom severity and efficacy of antipsychotic drugs for the treatment of manic or mixed episodes in bipolar disorder.

Methods

Included trials

We conducted a search of trials in ClinicalStudyDataRequest.com on 2 February 2016.

ClinicalStudyDataRequest.com is a secure Internet cloud-based platform that allows researchers to analyze patient-level clinical trial data provided by several pharmaceutical companies. Within this platform, we searched for randomized controlled trials (RCTs) that compared the efficacy of any antipsychotic versus placebo for the acute treatment of manic or mixed episodes associated with bipolar I disorder. For this indication we identified five double-blind RCTs that compared the efficacy of olanzapine versus placebo. No other relevant RCT was available within this platform. An a priori written protocol was submitted in www.clinicalstudydatarequest.com (Appendix S1).

All included RCTs were sponsored by Eli Lilly and have already been published elsewhere.²⁷⁻³¹

Two of these RCTs established olanzapine's efficacy in the treatment of adult patients with acute mania and were used by the US Food and Drug Administration (FDA) for the registration of the compound.^{27,29} A third RCT in adolescent patients, aged 13 to 17 years, led to the FDA approval of olanzapine for the treatment of adolescent bipolar I disorder.²⁸ Furthermore, one RCT included only patients with a mild to moderate manic or mixed episode.³⁰ All the included arms were flexible-dose arms, i.e. olanzapine 5 to 20 mg/d for adult patients and 2.5 to 20 mg/d for adolescent patients. Two studies had a third arm, one was divalproex³⁰ and one was haloperidol.³¹ By study protocol, we restricted the analysis to antipsychotics; therefore, the divalproex arm was dropped from the analysis. The haloperidol arm was also dropped because its sample size was too small (N=20) to assess any reliable association. All the data were completely anonymized before we had access to them.

Statistical analyses

IPD meta-analysis was conducted to investigate the relationship between baseline symptom severity and subsequent symptom change in the comparison of olanzapine versus placebo in the acute treatment of patients with manic or mixed episode associated with bipolar I disorder. All studies used the Young Mania Rating Scale (YMRS) to measure total severity of manic symptoms.³² The YMRS has 11 items; four items are rated on a 0–8 scale and the remaining seven items are rated on a 0–4 scale. Clear anchor-points are provided to help the clinician determine severity. The overall score can range between 0 and 60, and 12 or less is the standard symptom remission cut-off score.^{27,33,34}

To facilitate the clinical interpretation of our findings, we calculated the between-group expected effect sizes for different baseline severity scores as well as the numbers needed to treat (NNTs) for achieving a minimal important change (MIC). Generally, between-group effect sizes of 0.2, 0.5, and 0.8 roughly correspond to small, medium, and large effects respectively.³⁵ As for the MIC, it is defined as the smallest measured improvement that the patient would perceive as beneficial. But in the case of manic or mixed episodes, a validated MIC score has not yet been established. To date, only one study has attempted to identify the MIC of YMRS, but, in this study, only cross sectional baseline severity data, not longitudinal change data, were available;³⁶ therefore, the results require replication. Owing to the absence of a longitudinally validated MIC score, we decided to calculate the expected NNTs for a number of different percentage score changes in YMRS ranging from a 10% to 50% reduction. As the cut-off of 50%

YMRS score reduction is commonly used as a measure of response in studies of manic or mixed episodes,³⁷⁻⁴¹ we defined it as the primary cut-off to calculate expected NNTs.

To conduct the IPD meta-analysis, we used the approach developed by Hedeker and Gibbons⁴² and applied a 3-level mixed-effects model repeated-measures analysis (MMRM) with maximum likelihood estimation as in previous papers.^{11,20,21} The levels accounted for the data structure such that level 1 represented time, level 2 the participant, and level 3 the trial. For all the models, we included a random intercept for trial and a random linear increment for time. We modelled the absolute YMRS change (computed as baseline minus endpoint) as a function of the baseline, treatment and other covariates, depending on the model.

The following competing models with increasing complexity were tested: model 1, linear time, treatment, and the two-way linear time × treatment interaction (namely, the null model for comparison without baseline as a covariate); model 2, model 1 plus baseline YMRS score and all two-way interactions among linear time, treatment, and baseline score; model 3, model 2 plus the three-way interaction of linear time × treatment × baseline score; and model 4, model 3 plus the two-way and three-way interactions among quadratic time, treatment, and baseline score. The effect of interest was the interaction between treatment and baseline score. The models were tested unadjusted and adjusted for confounders (i.e. sex and age). The model with the smallest Bayesian Information Criterion value was chosen as the most parsimonious.⁴³ In addition Root Mean Square Error (RMSE) values were computed as another way to assess model fit.

For the most parsimonious model, we tested the assumptions of MRMM with regression diagnostics. Three-week post-baseline results based on the best-fitting models were used because that was the maximum duration for four out of the five included trials; therefore, we were able to perform an analysis that accounted for follow-up time differences across studies. We computed effect sizes and NNTs for a gradient of YMRS scores. All confidence intervals are calculated given the model selected with the lowest BIC. We used bootstrap to estimate 95% confidence intervals for both effect sizes and NNTs. The computation was based on the percentile method with 1,000 bootstrap samples. No secondary analysis was conducted.

All statistical analyses were done in R statistical software version 3.1.0⁴⁴ using the nlme package version 3.1-117 for mixed modeling and (R Foundation for Statistical Computing).⁴⁵ Bootstrap was conducted with R package boot.⁴⁶

Results

Characteristics of the Included RCTs

In total, 939 trial participants diagnosed with manic or mixed episode associated with bipolar I disorder according to DSM-IV or DSM-IV-TR were included in the analysis. Participants were randomized to olanzapine (N=552) or placebo (N=387), had a mean age of 36.01±10.6 years, and about half were male (51.6%). Important characteristics of the included trials are presented in Table 1.

Baseline Severity and Symptom Change

The best fitting MMRM model according to the BIC criterion was model 2 unadjusted for confounders (BIC=15838.90; Appendix S2). The RMSE was greatest for models without baseline, compared to those with baseline. The models with the baseline parameter had similar RMSE values. However, the chosen model (unadjusted model 2) had the lowest BIC and the smallest degrees of freedom among the models with baseline as a covariate (Appendix S2). For completeness, the parameters of all models were computed (Appendix S3).

Tests for the assumptions of the linearity of the best fitting model suggested that a linear model was reasonable (Appendix S4). Notably, in this model, the two-way interaction of treatment \times baseline YMRS score was statistically significant ($p=0.013$, Table 2).

Figure 1 shows a scatterplot of all observed and predicted YMRS change scores at three weeks (predicted scores were used for patients with less than three weeks treatment), superimposed with regression lines for the olanzapine and placebo arms. The magnitude of the difference in YMRS change scores between olanzapine and placebo increased as baseline YMRS scores increased. The mean score difference between olanzapine and placebo at three weeks was estimated to be 2.6 points for patients with a baseline YMRS score from 20 to 25, 4.7 points for patients with a baseline YMRS score from 25 to 35, and 8 points for patients with a baseline score up to 60 (Table 3).

The between-group effect size

The between-group effect size was 0.35 for patients with a baseline YMRS score from 20 to 25, 0.58 for those with a baseline score from 25-35, and 0.70 for those with a baseline score up to 60 (Table 3).

The number needed to treat to achieve a minimal important change

Using a 10% improvement in YMRS score as the MIC, a NNT of 8 can be expected for patients with a baseline YMRS score of 20 and a NNT of 6 for those with a baseline score of 30 or 45. For a 20% YMRS improvement as the MIC, a NNT of 8 can be expected for patients with a baseline YMRS score of 20, a NNT of 4 for those with a baseline score of 30, and a NNT of 5 for those with a baseline score of 45. For either a 30% or a 40% YMRS improvement as the MIC, a NNT of 7 can be expected for patients with a baseline YMRS score of 20, a NNT of 4 for those with a baseline score of 30, and a NNT of 3 for those with a baseline score of 45. Finally, for a 50% YMRS improvement as the MIC, a NNT of 7 can be expected for patients with a baseline YMRS score of 20, a NNT of 5 for those with a baseline score of 30, and a NNT of 4 for those with a baseline score of 45 (Figure 2 and Table 3).

Discussion

This study examined the association between initial severity and efficacy in the treatment of bipolar I disorder. Its results are based on IPD and suggest that the difference in symptom reduction between the antipsychotic olanzapine and placebo increases significantly as the

initial symptom severity increases. Using group- or study-level aggregated data, previous studies in both antipsychotics and mood stabilizers failed to identify a relationship between initial severity and drug-placebo difference²²⁻²⁴, but these studies were subject to the risk of ecological fallacy.⁴⁷⁻⁴⁹ An IPD analysis of divalproex sodium versus placebo in two RCTs (N=357) also found no significant relationship⁵⁰, but this discrepancy might be explained by the fact that divalproex is less efficacious than olanzapine;⁵¹⁻⁵³ thus, a larger sample of patients would be needed in order to detect a possible significant differentiation in symptom improvement between drug and placebo across the initial symptom severity range (type II error).

Some other studies examined the relationship between initial severity and symptom improvement (rather than drug-placebo difference in symptom improvement) and reported a significant relationship.^{23,50,54,55} Nevertheless, this relationship could be explained by Wilder's law of initial value⁵⁶ which suggests that greater change is expected for higher initial values. We selected drug-placebo difference, and not simply symptom improvement, as a variable of interest because the potential confounding role of that law would apply both for drug and placebo arms and could not be used to explain their in-between difference in symptom improvement. In addition, it is the superiority of drug over placebo that it is important for patients. Regression to the mean is frequently erroneously cited as another analytic obstacle in this kind of studies but its effect is 'cancelled out' by random allocation,^{57,58} multiple measurements at different times,⁵⁹ and it is expected to affect both drug and placebo arms equally.

The main strength of our analysis is the use of IPD from five RCTs with a considerable total sample size of 939 patients which makes our results rather robust. However, there are also important limitations. Only olanzapine was included in the analysis and, despite the large sample of patients, replication of the results examining other drugs is needed. There are many more placebo-controlled RCTs, but the ones we used were all the available relevant trials in the secure Internet cloud-based platform 'ClinicalStudyDataRequest.com'; the RCTs were not pre-selected to show a baseline severity effect. In addition, factors that could have affected olanzapine or placebo response, like the efficacy of blinding (potential bias in favour of olanzapine) or benzodiazepine use (potential bias in favour of placebo), could not be examined due to lack of data. Nevertheless, as we were interested in the effect of baseline severity on the difference of symptom change between olanzapine and placebo, such factors were not expected to bias our results towards finding a significant effect of baseline severity. Moreover, the patients in our analysis were highly selected since they came from double-blind, placebo-controlled randomized trials with strict inclusion and exclusion criteria (Table 1). No patient had initial severity score below 20 points in YMRS (entry cut-off for all studies) and only few patients had initial score above 40 points and even fewer above 50 points in YMRS. Research has documented that such selection criteria allow for restricted generalizations from clinical trial data to the general population.⁶⁰ For example, one epidemiological study found that more than 5 of 10 participants with bipolar depression or mania would have been excluded by at least 1 criterion if a standard set of eligibility criteria was applied to a national, representative, USA survey of 43,093 patients.⁶¹ Therefore, like all clinical trial data, caution is warranted regarding the generalizability of the current results beyond clinical treatment trial settings. To

better inform clinical practice, replications in large scale, naturalistic studies and examination of other drugs are needed. Finally, it is appropriate to note that the included patients with a diagnosis of a mixed episode according to DSM-IV and DSM-IV-TR criteria are very different from patients whose diagnosis has the new, DSM-V specifier 'mixed features'. This specifier, although used to replace the previous diagnoses of a mixed episode, is not *specific* for bipolar I disorder and thus it is not equivalent to the 'mixed episode' as previously defined by DSM-IV and DSM-IV-TR criteria.^{62,63}

Notwithstanding these limitations, a significant relationship between the efficacy of olanzapine over placebo and the initial symptom severity was demonstrated. Clinicians should be aware and expect greater efficacy among patients with more severe initial symptoms in bipolar I disorder. Nevertheless, benefits of olanzapine treatment were important for the full spectrum of examined baseline symptom severity; even toward the mildest end of the spectrum, effect size was estimated around 0.35 and NNT for the different definitions of MIC ranged between 7 and 8 which could not be considered negligible. Furthermore, clinicians should consider that there is limited evidence supporting or refuting medication change after response to the acute phase treatment;⁶⁴ thus, continuing long-term treatment with an antipsychotic shown to be effective during the acute phase would seem the most reasonable maintenance treatment strategy.² This treatment strategy, coupled with the evidence provided by our current study, may suggest that drugs with milder side-effects profile and in lower dosage should be considered in the acute phase treatment of patients with less severe mania because such patients derive smaller benefit while being at risk for the same degree of side-effect burdens.

For example, in a maintenance trial, olanzapine was shown slightly superior to lithium but produced significant excess weight gain.⁶⁵ Therefore, a careful initial choice of drug treatment during the acute episode is crucial for long-term treatment as well although relative merits of such therapies need to be assessed in trials of maintenance treatments per se. However, if the episode of acute mania was not extremely severe, and in particular if serious side-effects occur during maintenance treatment, switching from olanzapine to a more benign drug might be considered.⁶⁶

For researchers, the results of our study could be used to guide the design of future trials in bipolar I disorder. The usual current practice of setting the YMRS entry cut-off at 20 points could result in false-negative trials and hinder the detection of a real difference between the examined treatments, especially when small samples of patients are included. On the other hand, higher entry cut-offs could facilitate the detection of a signal, but could also complicate the recruitment process. Moreover, the effect of baseline severity on drug-placebo difference in symptom improvement seems to be present in bipolar I disorder as it was shown to be in schizophrenia and autism.^{20,21} Concerning major depressive disorder, the relationship between baseline severity and drug-placebo difference remains complex; findings of different studies provide inconsistent results that preclude any straightforward answer.⁵⁻¹³ In part, this inconsistency may be explained by limitations of aggregated data meta-analysis which use mean values of studies instead of IPD. This type of meta-analysis is frequently limited in detecting subtle moderators, because there is narrow variability in the observed means. Moreover, as mentioned before, there is the risk of ecological fallacy.⁴⁷⁻⁴⁹ But the inconsistency

of results remains even in the subset of studies that employed only IPD meta-analysis.^{8,9,11,13}

Another possible explanation could be the lower efficacy of antidepressants in depression (SMD from 0.31 to 0.33^{7,67,68}) compared to the efficacy of the antipsychotics amisulpride, olanzapine and risperidone in schizophrenia (SMD from 0.56 to 0.66⁶⁹), risperidone in autism (SMD=0.86⁷⁰) and olanzapine in bipolar mania (SMD from 0.44 to 0.48^{51,53,54}) for which the effect of baseline severity has been shown in previous studies^{20,21} and the current one. In that case, a larger sample of patients would be needed to detect a possible significant effect of baseline severity. Whether there are other explanations as well, it remains to be found. The baseline effect should be also examined in other mental disorders and general medical diseases. More studies based on IPD and in different conditions are needed to clarify this important question.

In conclusion, the current results based on 939 participants with bipolar I disorder from five placebo controlled RCTs showed that olanzapine is efficacious for the treatment of acute manic or mixed episodes across the initial symptom severity spectrum. Nevertheless, the difference in symptom improvement between olanzapine and placebo increased as the baseline severity increased. Toward the mildest end of baseline severity range, clinicians should balance the benefits and risks from treatment with an antipsychotic drug like olanzapine, but also take account of its potential prophylactic effect in the long-term.

Research in context

Evidence before this study

Starting with antidepressants, some studies suggested they may be less efficacious for the milder spectrum of the disorder compared to placebo whereas others did not, leading to a considerable debate among scientists and extreme titles in public newspapers. Continuing with antipsychotics, the effect of baseline severity on drug-placebo difference in symptom improvement was shown to be present in schizophrenia and autism disorder. But for mania, previous studies failed to identify a relationship between initial severity and drug-placebo difference. Nevertheless, these studies used group- or study-level aggregated data; thus, they were subject to the risk of ecological fallacy. To shed light on this topic, we searched ClinicalStudyDataRequest.com on 2 February 2016 for randomized controlled trials (RCTs) that compared any antipsychotic with placebo for the acute treatment of manic or mixed episodes associated with bipolar I disorder. ClinicalStudyDataRequest.com is a secure Internet cloud-based platform that allows researchers to analyze patient-level clinical trial data provided by several pharmaceutical companies. Our only search term was the medical condition “bipolar”. No language restriction was applied. We identified 31 trials, from which we selected 5 RCTs (n=939 patients) for inclusion. All included RCTs were sponsored by Eli Lilly and compared olanzapine with placebo. On the basis of these individual participant data (IPD), we assessed the influence of initial severity of acute mania on the efficacy of olanzapine.

Added value of this study

This is the first IPD meta-analysis to examine the association between initial symptom severity and the efficacy of an antipsychotic drug like olanzapine for the treatment of manic or mixed episodes in bipolar disorder. Our results show that the interaction between baseline symptom severity and treatment was statistically significant ($P = 0.013$) meaning that the greater the baseline severity, the greater the magnitude of the differences between olanzapine and placebo.

Implications of all the available evidence

Benefits of an antipsychotic drug like olanzapine can be expected for the full initial severity spectrum of patients with an acute manic or mixed episode, but less severely ill patients seem

to benefit less than more severely ill patients although they experience the same side effects. Therefore, toward the mildest end of baseline severity range, clinicians should be more careful with antipsychotic treatment, start with lower doses and less side-effect prone drugs, but also take account of the potential prophylactic antipsychotic effect in the long-term. The baseline severity effect should be also examined in other mental disorders and general medical diseases.

Acknowledgment section

Contributors

Drs Samara and Goldberg are joint first authors. Drs Samara and Leucht had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Samara, Leucht.

Acquisition, analysis, or interpretation of data: All authors.

Drafting of the manuscript: Samara, Leucht.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Goldberg, Levine.

Administrative, technical, or material support: Samara, Leucht.

Study supervision: Samara, Levine, Leucht.

Declaration of Interests

Stephen Z. Levine has received research support from SHIRE DEVELOPMENT LLC in the past 3 years. Toshi A. Furukawa has received lecture fees from Eli Lilly, Janssen, Meiji, MSD, Otsuka, Pfizer and Tanabe-Mitsubishi, and consultancy fees from Takeda Science Foundation. He has received royalties from Igaku-Shoin and Nihon Bunka Kagaku-sha publishers. He has received research support from Mochida and Tanabe-Mitsubishi. John R. Geddes is an NIHR Senior Investigator. In 2015 Andrea Cipriani was expert witness for Accord Healthcare for a patent issue about quetiapine extended release. In the past 3 years, Stefan Leucht reports receiving

honoraria for lectures from Eli Lilly, Lundbeck (Institute), Pfizer, Janssen, BMS, Johnson and Johnson, Otsuka, Roche, Sanofi, ICON, AbbVie, AOP Orphan, and Servier; for consulting/advisory boards from Roche, Janssen, Lundbeck, Eli Lilly, Otsuka, and TEVA; and for the preparation of educational material and publications from Lundbeck Institute and Roche. Eli Lilly has provided medication for a clinical trial led by Dr Leucht as the principal investigator. The other authors have no conflict of interest/financial support to declare.

Acknowledgments

This study received no funding. Eli Lilly allowed use of their participant-level data through the secure Internet cloud-based platform 'ClinicalStudyDataRequest.com'. Andrea Cipriani is supported by the NIHR Oxford Cognitive Health Clinical Research Facility

References

1. Grunze H, Vieta E, Goodwin GM, et al. The World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for the biological treatment of bipolar disorders: update 2012 on the long-term treatment of bipolar disorder. *World J Biol Psychiatry* 2013; **14**(3): 154-219.
2. Yatham LN, Kennedy SH, Parikh SV, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) collaborative update of CANMAT guidelines for the management of patients with bipolar disorder: update 2013. *Bipolar Disord* 2013; **15**(1): 1-44.
3. Malhi GS, Bassett D, Boyce P, et al. Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for mood disorders. *Aust N Z J Psychiatry* 2015; **49**(12): 1087-206.
4. Goodwin GM, Haddad PM, Ferrier IN, et al. Evidence-based guidelines for treating bipolar disorder: Revised third edition recommendations from the British Association for Psychopharmacology. *J Psychopharmacol* 2016; **30**(6): 495-553.
5. Khan A, Leventhal RM, Khan SR, Brown WA. Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. *J Clin Psychopharmacol* 2002; **22**(1): 40-5.
6. Khan A, Brodhead AE, Kolts RL, Brown WA. Severity of depressive symptoms and response to antidepressants and placebo in antidepressant trials. *J Psychiatr Res* 2005; **39**(2): 145-50.
7. Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med* 2008; **5**(2): e45.
8. Fournier JC, DeRubeis RJ, Hollon SD, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA* 2010; **303**(1): 47-53.
9. Nelson JC, Delucchi KL, Schneider LS. Moderators of outcome in late-life depression: a patient-level meta-analysis. *Am J Psychiatry* 2013; **170**(6): 651-9.
10. Melander H, Salmonson T, Abadie E, van Zwieten-Boot B. A regulatory Apologia--a review of placebo-controlled studies in regulatory submissions of new-generation antidepressants. *Eur Neuropsychopharmacol* 2008; **18**(9): 623-7.
11. Gibbons RD, Hur K, Brown CH, Davis JM, Mann JJ. Benefits from antidepressants: synthesis of 6-week patient-level outcomes from double-blind placebo-controlled randomized trials of fluoxetine and venlafaxine. *Arch Gen Psychiatry* 2012; **69**(6): 572-9.
12. Locher C, Kossowsky J, Gaab J, Kirsch I, Bain P, Krummenacher P. Moderation of antidepressant and placebo outcomes by baseline severity in late-life depression: A systematic review and meta-analysis. *J Affect Disord* 2015; **181**: 50-60.
13. Rabinowitz J, Werbeloff N, Mandel FS, Menard F, Marangell L, Kapur S. Initial depression severity and response to antidepressants v. placebo: patient-level data analysis from 34 randomised controlled trials. *Br J Psychiatry* 2016; **209**(5): 427-8.
14. Hyman M. Why Antidepressants Don't Work for Treating Depression. *The Huffington Post*, 2010. <http://www.huffingtonpost.com> (accessed).
15. Mercola J. Antidepressant drugs do not work, and they have serious side effects. *Health Impact News*, 2016. <http://healthimpactnews.com> (accessed).
16. Pietrowski A. Research: Antidepressants Don't Work and May Make Depression Worse. *Waking Times*, 2015. <http://www.wakingtimes.com> (accessed).
17. Gotzsche PC, Young AH, Crace J. Does long term use of psychiatric drugs cause more harm than good? *BMJ* 2015; **350**: h2435.
18. Menand L. Head case: can psychiatry be a science? . *The New Yorker*, 2010. <http://www.newyorker.com> (accessed).
19. Caplan P. Psychiatry's bible, the DSM, is doing more harm than good. *The Washington Post*, 2012. <https://www.washingtonpost.com> (accessed).

20. Furukawa TA, Levine SZ, Tanaka S, et al. Initial severity of schizophrenia and efficacy of antipsychotics: participant-level meta-analysis of 6 placebo-controlled studies. *JAMA Psychiatry* 2015; **72**(1): 14-21.
21. Levine SZ, Kodesh A, Goldberg Y, et al. Initial severity and efficacy of risperidone in autism: Results from the RUPP trial. *Eur Psychiatry* 2016; **32**: 16-20.
22. McIntyre RS, Konarski JZ, Jones M, Paulsson B. Quetiapine in the treatment of acute bipolar mania: efficacy across a broad range of symptoms. *J Affect Disord* 2007; **100 Suppl 1**: S5-14.
23. Yildiz A, Vieta E, Tohen M, Baldessarini RJ. Factors modifying drug and placebo responses in randomized trials for bipolar mania. *Int J Neuropsychopharmacol* 2011; **14**(7): 863-75.
24. Glue P, Herbison P. Comparative efficacy and acceptability of combined antipsychotics and mood stabilizers versus individual drug classes for acute mania: Network meta-analysis. *Aust N Z J Psychiatry* 2015; **49**(12): 1215-20.
25. Stewart LA, Clarke MJ. Practical methodology of meta-analyses (overviews) using updated individual patient data. Cochrane Working Group. *Stat Med* 1995; **14**(19): 2057-79.
26. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Eval Health Prof* 2002; **25**(1): 76-97.
27. Tohen M, Jacobs TG, Grundy SL, et al. Efficacy of olanzapine in acute bipolar mania: a double-blind, placebo-controlled study. The Olanzapine HGGW Study Group. *Arch Gen Psychiatry* 2000; **57**(9): 841-9.
28. Tohen M, Kryzhanovskaya L, Carlson G, et al. Olanzapine versus placebo in the treatment of adolescents with bipolar mania. *Am J Psychiatry* 2007; **164**(10): 1547-56.
29. Tohen M, Sanger TM, McElroy SL, et al. Olanzapine versus placebo in the treatment of acute mania. Olanzapine HGEH Study Group. *Am J Psychiatry* 1999; **156**(5): 702-9.
30. Tohen M, Vieta E, Goodwin GM, et al. Olanzapine versus divalproex versus placebo in the treatment of mild to moderate mania: a randomized, 12-week, double-blind study. *J Clin Psychiatry* 2008; **69**(11): 1776-89.
31. Katagiri H, Takita Y, Tohen M, Higuchi T, Kanba S, Takahashi M. Efficacy and safety of olanzapine in the treatment of Japanese patients with bipolar I disorder in a current manic or mixed episode: a randomized, double-blind, placebo- and haloperidol-controlled study. *J Affect Disord* 2012; **136**(3): 476-84.
32. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry* 1978; **133**: 429-35.
33. Bowden CL, Grunze H, Mullen J, et al. A randomized, double-blind, placebo-controlled efficacy and safety study of quetiapine or lithium as monotherapy for mania in bipolar disorder. *J Clin Psychiatry* 2005; **66**(1): 111-21.
34. McIntyre RS, Brecher M, Paulsson B, Huizar K, Mullen J. Quetiapine or haloperidol as monotherapy for bipolar mania--a 12-week, double-blind, randomised, parallel-group, placebo-controlled trial. *Eur Neuropsychopharmacol* 2005; **15**(5): 573-85.
35. Cohen J. Statistical Power Analysis in the Behavioral Sciences. 2nd ed ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
36. Lukasiewicz M, Gerard S, Besnard A, et al. Young Mania Rating Scale: how to interpret the numbers? Determination of a severity threshold and of the minimal clinically significant difference in the EMBLEM cohort. *Int J Methods Psychiatr Res* 2013; **22**(1): 46-58.
37. Zemlan FP, Hirschowitz J, Garver DL. Mood-incongruent versus mood-congruent psychosis: differential antipsychotic response to lithium therapy. *Psychiatry Res* 1984; **11**(4): 317-28.
38. Keck PE, Jr., Versiani M, Potkin S, et al. Ziprasidone in the treatment of acute bipolar mania: a three-week, placebo-controlled, double-blind, randomized trial. *Am J Psychiatry* 2003; **160**(4): 741-8.

39. Chengappa KN, Baker RW, Shao L, et al. Rates of response, euthymia and remission in two placebo-controlled olanzapine trials for bipolar mania. *Bipolar Disord* 2003; **5**(1): 1-5.
40. Vieta E, Bourin M, Sanchez R, et al. Effectiveness of aripiprazole v. haloperidol in acute bipolar mania: double-blind, randomised, comparative 12-week trial. *Br J Psychiatry* 2005; **187**: 235-42.
41. Sysko R, Walsh BT. A systematic review of placebo response in studies of bipolar mania. *J Clin Psychiatry* 2007; **68**(8): 1213-7.
42. Hedeker D, Gibbons RD. Longitudinal Data Analysis. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2006.
43. Schwarz G, Gideon E. Estimating the dimension of a model. *Annals of Statistics* 1978; **6**(2): 461-4.
44. Ihaka R, Gentleman R. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 1996; **5**(3): 299-314.
45. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. nlme: Linear and nonlinear mixed effects models. R package version 3.1-117. Vienna: R Foundation for Statistical Computing; 2014.
46. Canty A, Ripley B. boot: Bootstrap R (S-Plus) functions. *R package version* 2012; **1**(7).
47. Robinson WS. Ecological Correlations and the Behavior of Individuals. *American Sociological Review* 1950; **15**(3): 351-7.
48. Simmonds MC, Higgins JP. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. *Stat Med* 2007; **26**(15): 2982-99.
49. Riley RD, Lambert PC, Staessen JA, et al. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Stat Med* 2008; **27**(11): 1870-93.
50. Welge JA, Keck PE, Jr., Meinhold JM. Predictors of response to treatment of acute bipolar manic episodes with divalproex sodium or placebo in 2 randomized, controlled, parallel-group trials. *J Clin Psychopharmacol* 2004; **24**(6): 607-12.
51. Cipriani A, Barbui C, Salanti G, et al. Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple-treatments meta-analysis. *Lancet* 2011; **378**(9799): 1306-15.
52. Macritchie K, Geddes JR, Scott J, Haslam D, de Lima M, Goodwin G. Valproate for acute mood episodes in bipolar disorder. *Cochrane Database Syst Rev* 2003; (1): CD004052.
53. Yildiz A, Nikodem M, Vieta E, Correll CU, Baldessarini RJ. A network meta-analysis on comparative efficacy and all-cause discontinuation of antimanic treatments in acute bipolar mania. *Psychol Med* 2015; **45**(2): 299-317.
54. Yildiz A, Vieta E, Leucht S, Baldessarini RJ. Efficacy of antimanic treatments: meta-analysis of randomized, controlled trials. *Neuropsychopharmacology* 2011; **36**(2): 375-89.
55. Tarr GP, Herbison P, de la Barra SL, Glue P. Study design and patient characteristics and outcome in acute mania clinical trials. *Bipolar Disord* 2011; **13**(2): 125-32.
56. Wilder J. The law of initial value in neurology and psychiatry; facts and problems. *J Nerv Ment Dis* 1957; **125**(1): 73-86.
57. Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. *Int J Epidemiol* 2005; **34**(1): 215-20.
58. Torgerson C. 2 The Nature of Evidence in Effectiveness Research. Systematic reviews. London: Continuum International Publishing Group; 2003.
59. Chiolero A, Paradis G, Rich B, Hanley JA. Assessing the Relationship between the Baseline Value of a Continuous Variable and Subsequent Change Over Time. *Front Public Health* 2013; **1**: 29.
60. Licht RW. Limits of the applicability and generalizability of drug trials in mania. *Bipolar Disord* 2002; **4 Suppl 1**: 66-8.
61. Hoertel N, Le Strat Y, Lavaud P, Dubertret C, Limosin F. Generalizability of clinical trial results for bipolar disorder to community samples: findings from the National Epidemiologic Survey on Alcohol and Related Conditions. *J Clin Psychiatry* 2013; **74**(3): 265-70.

62. Malhi GS. Diagnosis of bipolar disorder: who is in a mixed state? *Lancet* 2013; **381**(9878): 1599-600.
63. Malhi GS, Lampe L, Coulston CM, et al. Mixed state discrimination: a DSM problem that wont go away? *J Affect Disord* 2014; **158**: 8-10.
64. Fountoulakis KN, Yatham L, Grunze H, et al. The International College of Neuro-Psychopharmacology (CINP) Treatment Guidelines for Bipolar Disorder in Adults (CINP-BD-2017), Part 2: Review, Grading of the Evidence, and a Precise Algorithm. *Int J Neuropsychopharmacol* 2016.
65. Tohen M, Greil W, Calabrese JR, et al. Olanzapine versus lithium in the maintenance treatment of bipolar disorder: a 12-month, randomized, double-blind, controlled clinical trial. *Am J Psychiatry* 2005; **162**(7): 1281-90.
66. Grande I, Bernardo M, Bobes J, Saiz-Ruiz J, Alamo C, Vieta E. Antipsychotic switching in bipolar disorders: a systematic review. *Int J Neuropsychopharmacol* 2014; **17**(3): 497-507.
67. Barbui C, Furukawa TA, Cipriani A. Effectiveness of paroxetine in the treatment of acute major depression in adults: a systematic re-examination of published and unpublished data from randomized trials. *CMAJ* 2008; **178**(3): 296-305.
68. Dunlop BW, Thase ME, Wun CC, et al. A meta-analysis of factors impacting detection of antidepressant efficacy in clinical trials: the importance of academic sites. *Neuropsychopharmacology* 2012; **37**(13): 2830-6.
69. Leucht S, Cipriani A, Spineli L, et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *Lancet* 2013.
70. Fung LK, Mahajan R, Nozzolillo A, et al. Pharmacologic Treatment of Severe Irritability and Problem Behaviors in Autism: A Systematic Review and Meta-analysis. *Pediatrics* 2016; **137** Suppl 2: S124-35.

Figure legends

Figure 1. Observed and estimated changes in the Young Mania Rating Scale (YMRS) following 3-week treatment of acute manic or mixed episodes.

Each straight line represents the expected reduction in YMRS score on olanzapine (red solid line) or placebo (blue solid line) predicted at 3 weeks on the vertical axis for corresponding baseline score on the horizontal axis.

Each dot represents an individual patient either on olanzapine (red circle) or placebo (blue triangle), with his/her reduction score at 3 weeks on the vertical axis and his/her baseline YMRS score at baseline on the horizontal axis.

Figure 2. Expected Numbers Needed to Treat (NNTs) for different definitions of the Minimal Important Change (MIC).

Each straight line represents the expected NNT for 5 different definitions of the MIC (10%, 20%, 30%, 40% and 50% YMRS score reduction) at 3 weeks on the vertical axis for corresponding baseline YMRS score on the horizontal axis.

Table 1. Included studies

| | | |
|-----------------------------|---------|---|
| Katagiri 2012 ³¹ | Methods | Double-blind, placebo- and active-controlled RCT Duration: 3 weeks (study period II) |
|-----------------------------|---------|---|

| | | |
|--------------------------|---------------|--|
| | Participants | <p>Diagnosis and main criteria for inclusion: bipolar I disorder, displaying a current manic or mixed episode, with or without psychotic features (DSM-IV-TR); YMRS total score ≥ 20 and ≤ 30; aged 20 to 65 years</p> <p>N at randomisation: 224 (olanzapine 105; [haloperidol 20;] placebo 99)</p> <p>Sex: 100 males, 121 females (includes only the safety sample: 221)</p> <p>Age: 43.4 ± 11.6 years (includes only the safety sample: 221)</p> <p>Episode: 203 manic, 18 mixed (40 out of 221 with psychotic features and 16 out of 221 with rapid-cycling course; includes only the safety sample: 221)</p> |
| | Interventions | Olanzapine 5-20 mg/d [versus haloperidol 2.5-10 mg/d] versus placebo, mean modal dose of olanzapine 11.2 ± 4.4 mg/d [and of haloperidol 5 ± 2 mg/d] |
| | Rating Scale | YMRS |
| | Sponsor | Eli Lilly Company |
| Tohen 1999 ²⁹ | Methods | Double-blind, placebo-controlled RCT Duration: 3 weeks |
| | Participants | <p>Diagnosis and main criteria for inclusion: bipolar I disorder, displaying an acute manic or mixed episode of at least 2 weeks' duration, with or without psychotic features (DSM-IV); YMRS total score ≥ 20; aged 18 to 65 years</p> <p>N at randomisation: 139 (olanzapine 70; placebo 69)</p> <p>Sex: 72 males, 67 females</p> <p>Age: 39.5 ± 11 years</p> <p>Episode: 115 manic, 24 mixed (74 out of 139 with psychotic features and 45 out of 139 with rapid-cycling course)</p> |
| | Interventions | Olanzapine 5-20 mg/d versus placebo, mean modal dose of olanzapine 14.9 ± 5.0 mg/d |
| | Rating scale | YMRS |
| | Sponsor | Eli Lilly Company |
| Tohen 2000 ²⁷ | Methods | Double-blind, placebo-controlled RCT Duration: 4 weeks |
| | Participants | <p>Diagnosis and main criteria for inclusion: bipolar I disorder, currently displaying an acute manic or mixed episode, with or without psychotic features (DSM-IV); history of at least 1 previous manic episode; YMRS total score ≥ 20; aged 18 to 70 years</p> <p>N at randomisation: 115 (olanzapine 55; placebo 60)</p> <p>Sex: 58 males, 57 females</p> <p>Age: 38.6 ± 10.4 years</p> <p>Episode: 66 manic, 49 mixed (64 out of 115 with psychotic features and 45 out of 115 with rapid-cycling course)</p> |
| | Interventions | Olanzapine 5-20 mg/day versus placebo, mean modal dose of olanzapine 16.4 ± 4.2 mg/d |
| | Rating scale | YMRS |
| | Sponsor | Eli Lilly Company |
| Tohen 2007 ²⁸ | Methods | Double-blind, placebo-controlled RCT |

| | | |
|--------------------------|---------------|--|
| | | Duration: 3 weeks |
| | Participants | <p>Diagnosis and main criteria for inclusion: bipolar I disorder, currently displaying an acute manic or mixed episode, with or without psychotic features (DSM-IV-TR); YMRS total score ≥ 20; aged 13 to 17 years</p> <p>N at randomisation: 161 (olanzapine 107; placebo 54)</p> <p>Sex: 85 males, 76 females</p> <p>Age: 15.2 ± 1.25 years</p> <p>Episode: 75 manic, 86 mixed (29 out of 160 with psychotic features and 30 out of 144 with rapid-cycling course)</p> |
| | Interventions | Olanzapine 2.5-20 mg/day versus placebo, mean modal dose of olanzapine 10.7 ± 4.5 mg/d |
| | Rating scale | Adolescent-structured YMRS |
| | Sponsor | Eli Lilly Company |
| Tohen 2008 ³⁰ | Methods | <p>Double-blind, placebo- and active-controlled RCT</p> <p>Duration: 3 weeks (study period II)</p> |
| | Participants | <p>Diagnosis and main criteria for inclusion: bipolar I disorder, currently displaying an acute mild or moderate manic or mixed episode, without a rapid-cycling course or psychotic features (DSM-IV-TR); YMRS total score ≥ 20 and ≤ 30; aged 18 to 65 years</p> <p>N at randomisation: 521 (olanzapine 215; [divalproex 201;] placebo 105)</p> <p>Sex: 232 males, 254 females (includes only the ITT sample: 486)</p> <p>Age: 39.6 ± 12.0 years</p> <p>Episode: 141 manic, 323 mixed (includes only the patients whose current episode is defined as recent episode manic or mixed, based on the ITT sample)</p> |
| | Interventions | Olanzapine 5-20 mg/day [versus divalproex 500-2500mg/d] versus placebo, mean modal dose of olanzapine 11.4 ± 2.49 mg/d [and of divalproex 848.4 ± 135.62 mg/d] |
| | Rating scale | YMRS |
| | Sponsor | Eli Lilly Company |

DSM-IV, -IV-TR = different versions of the Diagnostic and Statistical Manual of Mental Disorders; ITT: intention-to-treat; mg/d: milligram/day; N: Number; RCT: randomized controlled trial; YMRS: Young Mania Rating Scale

[drug groups in squared brackets were not used in any analysis]

Table 2. Parameters of model with best fit

| Parameters | Value | Standard error | T-value | 95% CI | p-value |
|-----------------------|--------------|-----------------------|----------------|---------------|----------------|
| Intercept | 8.63 | 2.23 | 3.87 | 4.27, 12.99 | 0.00011 |
| Baseline | -0.20 | 0.08 | -2.54 | -0.35, -0.05 | 0.011 |
| Olanzapine | -4.21 | 2.51 | -1.68 | -9.13, 0.71 | 0.094 |
| Time | -2.51 | 0.75 | -3.34 | -3.99, -1.04 | 0.00087 |
| Baseline x Olanzapine | 0.22 | 0.09 | 2.49 | 0.05, 0.39 | 0.013 |
| Baseline x Time | 0.16 | 0.03 | 6.11 | 0.11, 0.21 | <0.0001 |
| Olanzapine x Time | 0.80 | 0.33 | 2.44 | 0.16, 1.44 | 0.015 |

CI: Confidence Interval from bootstrap; p-value based on 1000 sized permutation tests.

Table 3. Baseline YMRS Score and Expected Reduction in YMRS Score, Effect Size, Expected % of Patients with 20% and 50% YMRS Reduction and NNT at 3 Weeks

| Baseline YMRS score | Expected Reduction in YMRS score | | Effect Size | 95% CI | p-value | Expected % of Patients Showing 20% YMRS Reduction from Baseline to Endpoint | | NNT | 95% CI | p-value | Expected % of Patients Showing 50% YMRS Reduction from Baseline to Endpoint | | NNT | 95% CI | p-value |
|---------------------|----------------------------------|---------|-------------|------------|---------|---|---------|-----|--------|---------|---|---------|-----|--------|---------|
| | Olanzapine | Placebo | | | | Olanzapine | Placebo | | | | Olanzapine | Placebo | | | |
| [20, 25] | 9.26 | 6.70 | 0.35 | 0.11, 0.60 | 0.003 | 79 | 67 | 8 | 4, 30 | 0.015 | 48 | 34 | 7 | 4, 19 | 0.004 |
| (25, 35] | 14.25 | 9.51 | 0.58 | 0.34, 0.86 | 0.001 | 83 | 60 | 4 | 3, 10 | 0.001 | 49 | 28 | 5 | 3, 12 | 0.002 |
| (35, 60] | 21.72 | 13.71 | 0.70 | 0.31, 1.23 | 0.002 | 90 | 71 | 5 | 2, 20 | 0.005 | 43 | 21 | 4 | 2, 18 | 0.010 |

YMRS: Young Mania Rating Scale; NNT: number needed to treat; CI: Confidence Interval from bootstrap; p-value based on 1000 sized permutation tests.