TREATMENT EFFECT HETEROGENEITY IN RANDOMIZED PLACEBO-CONTROLLED TRIALS OF LAMOTRIGINE FOR ACUTE BIPOLAR DEPRESSION

A Thesis Submitted to the
College of Graduate and Postdoctoral Studies
In Partial Fulfillment of the Requirements
For the Degree of Master of Science
In the College of Medicine
University of Saskatchewan
Saskatoon

By

Evyn Peters

PERMISSION TO USE

In presenting this thesis/dissertation in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis/dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis/dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

DISCLAIMER

Reference in this thesis/dissertation to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement, recommendation, or favoring by the University of Saskatchewan. The views and opinions of the author expressed herein do not state or reflect those of the University of Saskatchewan, and shall not be used for advertising or product endorsement purposes.

Requests for permission to copy or to make other uses of materials in this thesis/dissertation in whole or part should be addressed to:

Dean of the College of Medicine
4A20 Health Science Building, 107 Wiggins Road
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5E5 Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9 Canada

ABSTRACT

Lamotrigine is used to treat depressive episodes in bipolar disorder, but evidence from clinical trials has been inconsistent. There is some evidence that certain patient subgroups may benefit more than others. The goal of this thesis was to further consider the possibility of heterogeneous treatment effects by conducting a pooled analysis of five randomized, double-blind, placebo-controlled trials of lamotrigine for acute bipolar depression. The results are presented in two manuscripts. In the first study, a prespecified analysis was conducted to determine if patients with melancholic depression were more responsive than nonmelancholic patients. A subgroup with higher scores on depression scale items representing melancholic features had numerically larger treatment effects, but interaction tests were nonsignificant, and a melancholic diagnosis from structured clinical interviews was not associated with larger effects. Furthermore, baseline depression severity was inconsistently associated with response depending on which depression scale was used to define severity. The second study was an exploratory analysis that examined treatment effects on individual depression scale items and then attempted to use these items to create subgroup variables that could identify patients with larger overall treatment effects. On two depression scales, there were larger and statistically significant effects on items representing depressed mood/sadness, decreased interest/anhedonia, fatigue/anergia, and pessimism/guilt. The items with larger effects tended to be more prevalent in the sample at baseline, suggesting a floor effect limited the sensitivity of other items. Patients with higher scores on the mood/sadness and interest/anhedonia items also tended to have larger overall treatment effects compared to the rest of the sample. Taken together, in contrast to previous research, the results did not clearly support the hypothesis that patients with melancholic depression benefit more from lamotrigine compared to placebo, nor did baseline depression severity moderate treatment effects in a straightforward manner. Consistent with previous research, targeted assessments of core depressive symptoms that were more prevalent in the sample at baseline appeared to be more sensitive to change compared to total depression sum scores. More research is needed to establish the precise role of lamotrigine in the treatment of bipolar disorder.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor, Dr. Lloyd Balbuena, for his patient support and guidance. I would like to thank my committee members, Dr. Rohit Lodhi, Dr. Yanbo Zhang, Dr. Cindy Feng, and Dr. Hua Li, for volunteering their time and expertise. Dr. Balbuena, Dr. Lodhi, Dr. Zhang, Dr. Feng, and Dr. Li have contributed meaningfully to each manuscript and the research process. I would also like to thank Dr. Marilyn Baetz and Dr. Marla Davidson for their encouragement and support, and Dr. Bonnie Janzen for acting as an external examiner.

I would like to thank the Department of Psychiatry and the College of Medicine, University of Saskatchewan, and the trustees of the Laura E. Chapman Award, for their financial support. I would also like to thank GlaxoSmithKline and clinicalstudydatarequest.com for providing access to the clinical trial data.

Lastly, I would like to thank Dr. Rudy Bowen for his unconditional support and mentorship.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA	Analysis of variance
CI	Confidence interval
DSM	Diagnostic and Statistical Manual of Mental Disorders
FDA	U.S. Food and Drug Administration
HAMD	Hamilton Depression Rating Scale
HAMD-6	Hamilton Depression Rating Scale, 6-item subscale
HAMD-17	Hamilton Depression Rating Scale, 17-item version
HAMD-31	Hamilton Depression Rating Scale, 31-item version
HPA	Hypothalamic-pituitary-adrenal
LTG	Lamotrigine
MADRS	Montgomery–Åsberg Depression Rating Scale
MADRS-6	Montgomery–Åsberg Depression Rating Scale, 6-item subscale
NNT	Number needed to treat
OR	Odds ratio
PBO	Placebo
SCID	Structured Clinical Interview for DSM-IV
SD	Standard deviation
SE	Standard error

CHAPTER 1

Introduction

Bipolar disorder is one of several mood disorders in the Diagnostic and Statistical Manual of Mental Disorders (DSM) (1). It is characterised by recurrent episodes of mania (elevated or irritable mood with increased energy, arousal, and goal-directed activity) and depression (decreased mood, energy, and goal-directed behaviour) (1). The lifetime prevalence is only 0.9%, but it is one of the leading causes of disability worldwide (2). The risk of suicide is 20-30 times higher than in the general population (3). Depressive episodes contribute significantly to this burden and are difficult to treat, partially due to the limited tolerability of existing medications (4, 5).

Lamotrigine is a sodium- and calcium-channel blocker that inhibits the release of glutamate from presynaptic neurons (6, 7). It is widely used to treat bipolar depression (8, 9). Lamotrigine has a favourable side-effect profile; it stands alone as the only recommended treatment that does not cause metabolic syndrome or sexual dysfunction, does not require regular bloodwork, is considered safer in pregnancy, and may not require polypharmacy for mania prophylaxis (5, 10). Unfortunately, the evidence supporting its efficacy is inconsistent, at best.

1.1 Lamotrigine for bipolar depression

Lamotrigine was first approved as an antiepileptic in the early 1990s (11). Valproic acid and carbamazepine, two other antiepileptics, were already being used to treat bipolar disorder (11). This coincidence, along with two positive case reports and an epileptology-inspired kindling theory of bipolar disorder, convinced GlaxoSmithKline to sponsor what was arguably the most ambitious and comprehensive bipolar disorder drug development program to date (11). However, despite promising results from a preliminary open-label trial (12), five randomized placebo-controlled trials for acute bipolar depression did not show consistent benefit (13), although in two maintenance trials, lamotrigine reduced relapse times compared to placebo (14, 15). As a result, in 2003, it was approved by the U.S. Food and Drug Administration (FDA) for the prevention of depressive episodes, but not for acute treatment (6).

Prior to the FDA decision, a small non-industry trial found lamotrigine monotherapy to be more effective than placebo (16). More recently, the large, multi-site CEQUEL study reported that lamotrigine added to 7-14 days of quetiapine was more effective than the addition of placebo (17). Previous placebo-controlled adjunct trials were also generally supportive, albeit smaller and less rigorous in design than CEQUEL (18, 19).

Secondary analyses of industry-sponsored trial data have also uncovered efficacy signals. A reanalysis of two placebo-controlled trials (one including unipolar patients) conducted by Mitchell et al. (20) noted larger and statistically significant treatment effects on some depressive symptoms (i.e., depressed mood and cognitions, and psychomotor retardation) than on total depression scores (20). In addition, a meta-analysis of the five bipolar depression monotherapy trials conducted by Geddes et al. (21) found lower placebo response rates in patients with baseline depression scores above the sample mean, resulting in larger and statistically significant treatment effects. Taken together, these studies suggest lamotrigine may only be more effective than placebo for certain patients (20, 21). Treatment effect heterogeneity could also explain the negative results from the industry-sponsored monotherapy trials, as treatment effects would be smaller and harder to detect in the broader samples recruited.

1.2 Melancholia

The classical features of melancholic depression—sudden, unprovoked, and often recurrent episodes of extreme low mood, psychomotor retardation, mood-congruent delusions, and neurovegetative symptoms such as weight loss, decreased libido, and interrupted sleep—have been identified since antiquity, and throughout most of the 19th and 20th centuries, melancholic depression was considered a distinct subtype, more commonly associated with bipolar disorder (i.e., manic-depressive illness) than unipolar depression (22-24).

Arguments for classifying melancholia as a distinct subtype are partially motivated by its differential response to treatment; placebo and psychotherapy are believed to be less effective, relative to patients with nonmelancholic depression, such that greater benefit is derived from medications and electroconvulsive therapy (22-26). Furthermore, several decades of research have consistently associated melancholic depression with hyperactivity of the hypothalamic-pituitary-adrenal (HPA) axis, most commonly with the dexamethasone-suppression test (27-29), which by itself also predicts nonresponse to placebo (30) and psychotherapy (31). Central hypercortisolism also likely explains why melancholic features have been consistently associated with worse performance on tests of memory, attention, and executive function (32-36), as the deleterious effect of exogenous corticosteroids on cognition have been well-documented (37).

1.3 Lamotrigine for melancholic depression

There is some indirect evidence to suggest melancholic depression may be more responsive to lamotrigine. The pattern of results in the Geddes et al. (21) meta-analysis for

patients with above-average depression scores (i.e., lower placebo response rates without a corresponding reduction in drug response rates) is the same pattern that has been reported for patients with melancholic depression and a positive dexamethasone-suppression test being treated with other medications (26, 30); because melancholic patients tend to have higher depression scores (38), it is not unreasonable to assume that they would have been over-represented in this subgroup. Furthermore, the depressive symptoms that Mitchell et al. (20) found to be more responsive to lamotrigine tend to be more pronounced in melancholic depression, particularly psychomotor retardation (22-25).

Lamotrigine may also preferentially target pathophysiological processes underlying melancholic depression. Makatsori et al. (39) found that lamotrigine pretreatment significantly decreased the HPA-axis response to a social-stress test in healthy individuals, compared to placebo. This is consistent with the notion that glutamatergic neurotransmission activates the HPA axis (40, 41). Quetiapine is used to treat unipolar and bipolar depression, and like lamotrigine, it suppresses the HPA axis (42, 43). Accordingly, we conducted a pooled-analysis of four placebo-controlled trials and found a larger treatment effect in unipolar depressed patients with melancholic symptoms at baseline compared to those without (44).

Two placebo-controlled trials conducted by Brown et al. (45, 46) found that lamotrigine significantly improved declarative memory in nonpsychiatric patients receiving chronic corticosteroid therapy. This suggests lamotrigine could also attenuate downstream effects caused by sustained central hypercortisolism in melancholic depression.

1.4 Thesis overview

Broadly speaking, the purpose of this thesis was to examine further the possibility of treatment effect heterogeneity in the five industry-sponsored trials of lamotrigine for acute bipolar depression. It contains two manuscripts.

The first manuscript is presented in Chapter 2. This study was designed to test the hypothesis that lamotrigine may be more effective for patients with melancholic depression. It followed a prespecified analysis plan and was designed to improve upon a pilot study we conducted prior to beginning the Master of Science program (38). To meet the space requirements of the target journal, some information has been placed in appendices that will eventually be submitted as online supplementary material.

The second manuscript is contained in Chapter 3. This manuscript reports a series of exploratory analyses that were conducted on the same data after completion of the first study. Individual depression scale items were analyzed to determine if certain depressive symptoms were more responsive to lamotrigine treatment. These scale items were then used to create subgroup variables that could reliably identify patients who, overall, responded better to lamotrigine compared to placebo. The analysis and proceeding discussion aimed to move away from the question of melancholic depression as a treatment effect moderator and attempted to situate the results within the broader literature concerning symptom-level moderators in depressed patients more broadly.

Chapter 4 contains a general discussion and concluding remarks.

1.5 References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. Arlington, VA: American Psychiatric Association, 2013.

2. Moreira ALR, Van Meter A, Genzlinger J, et al. Review and meta-analysis of epidemiologic studies of adult bipolar disorder. J Clin Psychiatry 2017;78(9):e1259-e1269.

3. Pompili M, Gonda X, Serafini G, et al. Epidemiology of suicide in bipolar disorders: a systematic review of the literature. Bipolar Disord 2013;15(5):457-490.

4. Baldessarini RJ, Vázquez GH, Tondo L. Bipolar depression: a major unsolved challenge. Int J Bipolar Disord 2020;8(1):1.

5. Yatham LN, Kennedy SH, Parikh SV, et al. Canadian network for mood and anxiety treatments (CANMAT) and international society for bipolar disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder. Bipolar Disord 2018;20(2):97-170.

6. Goldsmith DR, Wagstaff AJ, Ibbotson T, et al. Lamotrigine. Drugs 2003;63(19):2029-2050.

7. Ketter TA, Manji HK, Post RM. Potential mechanisms of action of lamotrigine in the treatment of bipolar disorders. J Clin Psychopharmacol 2003;23(5):484-495.

8. Kessing LV, Vradi E, Andersen PK. Nationwide and population-based prescription patterns in bipolar disorder. Bipolar Disord 2016;18(2):174-182.

9. Karanti A, Kardell M, Lundberg U, Landén M. Changes in mood stabilizer prescription patterns in bipolar disorder. J Affect Disord 2016;195:50-56.

10. Bowden CL, Asnis GM, Ginsberg LD, et al. Safety and tolerability of lamotrigine for bipolar disorder. Drug Saf 2004;27(3):173-184.

11. Weisler RH, Calabrese JR, Bowden CL, et al. Discovery and development of lamotrigine for bipolar disorder: a story of serendipity, clinical observations, risk taking, and persistence. J Affect Disord 2008;108(1-2):1-9.

12. Calabrese JR, Bowden CL, McElroy SL, et al. Spectrum of activity of lamotrigine in treatment-refractory bipolar disorder. Am J Psychiatry 1999;156(7):1019-1023.

13. Calabrese JR, Huffman RF, White RL, et al. Lamotrigine in the acute treatment of bipolar depression: results of five double-blind, placebo-controlled clinical trials. Bipolar Disord 2008;10:323-333.

14. Calabrese JR, Bowden CL, Sachs G, et al. A placebo-controlled 18-month trial of lamotrigine and lithium maintenance treatment in recently depressed patients with bipolar I disorder. J Clin Psychiatry 2003;64(9):1013-1024.

15. Calabrese JR, Suppes T, Bowden CL, et al. A double-blind, placebo-controlled, prophylaxis study of lamotrigine in rapid-cycling bipolar disorder. J Clin Psychiatry 2000;61(11):841-850.

16. Frye MA, Ketter TA, Kimbrell TA, et al. A placebo-controlled study of lamotrigine and gabapentin monotherapy in refractory mood disorders. J Clin Psychopharmacol 2000;20(6):607-614.

17. Geddes JR, Gardiner A, Rendell J, et al. Comparative evaluation of quetiapine plus lamotrigine combination versus quetiapine monotherapy (and folic acid versus placebo) in bipolar depression (CEQUEL): a 2× 2 factorial randomised trial. Lancet Psychiatry 2016;3(1):31-39.

18. van der Loos ML, Mulder PG, Hartong EGTM, et al. Efficacy and safety of lamotrigine as add-on treatment to lithium in bipolar depression: a multicenter, double-blind, placebo-controlled trial. J Clin Psychiatry 2009;70(2):223-231.

19. Normann C, Hummel B, Scharer LO, et al. Lamotrigine as adjunct to paroxetine in acute depression: a placebo-controlled, double-blind study. J Clin Psychiatry 2002;63(4):337-344.

20. Mitchell PB, Hadzi-Pavlovic D, Evoniuk G, et al. A factor analytic study in bipolar depression, and response to lamotrigine. CNS Spectr 2013;18(4):214-224.

21. Geddes JR, Calabrese JR, Goodwin GM. Lamotrigine for treatment of bipolar depression: independent meta-analysis and meta-regression of individual patient data from five randomised trials. Brit J Psychiatry 2009;194(1):4-9.

22. Shorter E. The doctrine of the two depressions in historical perspective. Acta Psychiatr Scand 2007;115:5-13.

23. Fink M, Taylor MA. Resurrecting melancholia. Acta Psychiatr Scand 2007;115:14-20.

24. Kendler KS. The genealogy of major depression: symptoms and signs of melancholia from 1880 to 1900. Mol Psychiatry 2017;22(11):1539-1553.

25. Parker G, Fink M, Shorter E, et al. Issues for DSM-5: whither melancholia? The case for its classification as a distinct mood disorder. Am J Psychiatry 2010;167(7):745-747.

26. Valerio MP, Szmulewicz AG, Martino DJ. A quantitative review on outcome-to-antidepressants in melancholic unipolar depression. Psychiatr Res 2018;265:100-110.

27. Carroll BJ. The dexamethasone suppression test for melancholia. Brit J Psychiatry 1982;140(3):292-304.

28. Gold PW, Chrousos GP. Organization of the stress system and its dysregulation in melancholic and atypical depression: high vs low CRH/NE states. Mol Psychiatry 2002;7(3):254-275.

29. Juruena MF, Bocharova M, Agustini B, et al. Atypical depression and non-atypical depression: is HPA axis function a biomarker? A systematic review. J Affect Disord 2018;233:45-67.

30. Ribeiro SC, Tandon R, Grunhaus L, et al. The DST as a predictor of outcome in depression: a meta-analysis. Am J Psychiatry 1993;150(11):1618-1629.

31. Fischer S, Strawbridge R, Vives AH, et al. Cortisol as a predictor of psychological therapy response in depressive disorders: systematic review and meta-analysis. Brit J Psychiatry 2017;210(2):105-109.

32. Bosaipo NB, Foss MP, Young AH, et al. Neuropsychological changes in melancholic and atypical depression: a systematic review. Neurosci Biobehav Rev 2017;73:309-325.

33. Egeland J, Lund A, Landrø NI, et al. Cortisol level predicts executive and memory function in depression, symptom level predicts psychomotor speed. Acta Psychiatr Scand 2005;112(6):434-441.

34. Keller J, Gomez R, Williams G, et al. HPA axis in major depression: cortisol, clinical symptomatology and genetic variation predict cognition. Mol Psychiatry 2017;22(4):527-536.

35. Travis SG, Coupland NJ, Hegadoren K, et al. Effects of cortisol on hippocampal subfields volumes and memory performance in healthy control subjects and patients with major depressive disorder. J Affect Disord 2016;201:34-41.

36. Zaninotto L, Solmi M, Veronese N, et al. A meta-analysis of cognitive performance in melancholic versus non-melancholic unipolar depression. J Affect Disord 2016;201:15-24.

37. Prado CE, Crowe SF. Corticosteroids and cognition: a meta-analysis. Neuropsychol Rev 2019:1-25.

38. Peters EM, Bowen R, Balbuena L. Melancholic symptoms in bipolar II depression and responsiveness to lamotrigine in an exploratory pilot study. J Clin Psychopharmacol 2018;38(5):509-512.

39. Makatsori A, Duncko R, Moncek F, et al. Modulation of neuroendocrine response and non-verbal behavior during psychosocial stress in healthy volunteers by the glutamate release-inhibiting drug lamotrigine. Neuroendocrinology 2004;79:34–42.

40. Herman JP, Mueller NK, Figueiredo H. Role of GABA and glutamate circuitry in hypothalamo-pituitary-adrenocortical stress integration. Ann N Y Acad Sci 2004;1018(1):35-45.

41. Zelena D, Mergl Z, Makara GB. Glutamate agonists activate the hypothalamic–pituitary–adrenal axis through hypothalamic paraventricular nucleus but not through vasopressinerg neurons. Brain Res 2005;1031(2):185-193.

42. Cohrs S, Röher C, Jordan W, et al. The atypical antipsychotics olanzapine and quetiapine, but not haloperidol, reduce ACTH and cortisol secretion in healthy subjects. Psychopharmacology 2006;185(1):11-8.

43. Sarubin N, Nothdurfter C, Schmotz C, et al. Impact on cortisol and antidepressant efficacy of quetiapine and escitalopram in depression. Psychoneuroendocrinology 2014;39:141-151.

44. Peters EM, Bowen R, Balbuena L. Melancholic depression and response to quetiapine: a pooled analysis of four randomized placebo-controlled trials. J Affect Disord 2020;276:696-698.

45. Brown ES, Wolfshohl J, Shad MU, et al. Attenuation of the effects of corticosteroids on declarative memory with lamotrigine. Neuropsychopharmacology 2008;33(10):2376-2383.

46. Brown ES, Sayed N, Choi C, et al. A randomized, double-blind, placebo-controlled trial of lamotrigine for prescription corticosteroid effects on the human hippocampus. Eur Neuropsychopharmacol 2019;29(3):376-383.

CHAPTER 2

Manuscript 1

Melancholic features in bipolar depression and response to lamotrigine: a pooled analysis of five randomized placebo-controlled trials

ABSTRACT

A pilot study suggested lamotrigine may be more effective for bipolar depression with melancholic features. We tested this hypothesis in a pooled analysis of five randomized double-blind placebo-controlled trials of lamotrigine for acute bipolar depression. The pooled sample consisted of 1,072 adult outpatients. Depressive symptoms were assessed for 7-10 weeks with the Hamilton Depression Rating Scale and the Montgomery–Åsberg Depression Rating Scale. The outcome measure was end-trial response (score reduction ≥ 50%). Melancholic features were assessed with both the Structured Clinical Interview for DSM-IV (SCID) and baseline depression scale items, according to DSM criteria. The item-based melancholic specifier was associated with numerically larger treatment effects, although subgroup-treatment interactions in logistic regression models did not reach statistical significance. The small subgroup of patients with severe psychomotor retardation also appeared to benefit from lamotrigine. However, the SCID melancholic specifier was not associated with larger treatment effects. Baseline depression severity was inconsistently associated with response depending on which scale was used to define severity. The two melancholia variables had poor agreement despite having similar prevalences. Our results do not clearly support the original hypothesis but do reinforce the importance of replicating secondary analyses of clinical trials with additional data.

Keywords: anticonvulsant; endogenous depression; mood disorder; psychomotor retardation

2.1 Introduction

Previously we hypothesized lamotrigine may be more effective for melancholic depression (1). The rationale was that lamotrigine inhibits presynaptic glutamate release (2) which appears to blunt the cortisol stress response (3) and mitigate the deleterious neuropsychological effects of chronic corticosteroid exposure (4), while melancholic features are associated with hypothalamic-pituitary-adrenal (HPA) axis hyperactivity (5, 6) and cognitive dysfunction (7), possibly mediated by glutamatergic excitotoxicity (6). Furthermore, lamotrigine appears to improve psychomotor retardation (8)—a cardinal melancholic sign (9, 10).

In a pilot study, we reanalyzed one of five industry-sponsored trials of lamotrigine for acute bipolar depression after using baseline scale items to split the sample into melancholic and nonmelancholic subgroups (1). There was more separation of lamotrigine from placebo in the melancholic subgroup, but not all results were significant, and we were unable to adequately control for baseline severity (1). The latter was a notable limitation as Geddes et al. (11) had reanalyzed all five trials and found a larger effect in patients with baseline depression scores above the sample mean (11). Here we conducted a pooled analysis of all five trials to replicate the pilot results in a larger sample with a standard melancholia assessment, to consider models with melancholia and depression severity included together, and to explore whether baseline psychomotor retardation alone was associated with larger treatment effects.

2.2 Methods

2.2.1 Data acquisition

Data was provided by the sponsor via clinicalstudydatarequest.com. We obtained ethical approval and preregistered prior to accessing the data. Each trial is summarized in Table 2.1. They were all randomized, double-blind, placebo-controlled, parallel-group, monotherapy trials (12). All data was pooled, except for the 50 mg/day arm, following Geddes et al. (11).

2.2.2 Participants

The pooled sample consisted of 1,072 adult outpatients aged 18 years or older with a diagnosis of bipolar disorder I or II currently experiencing an acute major depressive episode. Notable exclusion criteria were: prior lamotrigine treatment; active mania or suicidality; rapid cycling; social phobia, panic disorder, obsessive-compulsive disorder, or bulimia nervosa in the last year; substance abuse or dependence in the last month or year, respectively;

pregnancy/breastfeeding; other psychotropic medication; recent psychotherapy; thyroid disease; and epilepsy (12).

2.2.3 Instruments

The Structured Clinical Interview for DSM-IV (SCID) (13) was used to confirm study eligibility. Depressive symptoms were measured at baseline and weekly with the 17- and 31-item versions of the Hamilton Depression Rating Scale (HAMD) (14, 15) and the Montgomery–Åsberg Depression Rating Scale (MADRS) (16). We used all three scales because the HAMD-17 and MADRS overrepresent melancholic symptoms, whereas the HAMD-31 includes atypical symptoms (e.g., increased sleep, increased appetite) (14-16).

2.2.4 Outcome measures

The outcome measures were end-trial response rates (score reduction ≥ 50% from baseline) on the HAMD-17, HAMD-31, and MADRS. Response rates were chosen over change scores because they are less susceptible to floor effects (17), and we anticipated melancholic depression would be associated with higher baseline scores. We did not prioritize one scale as the primary outcome but were looking for consistency across scales.

2.2.5 Melancholic features

We used two methods to define melancholic depression. The first was the SCID melancholic specifier. Second, we used baseline HAMD and MADRS item scores to split the sample into melancholic and nonmelancholic subgroups according to the DSM criteria. Because there is no accepted threshold for diagnosing melancholic symptoms with item scores, we planned to adjust the diagnostic algorithm so that the prevalence roughly approximated that returned by the SCID (see Appendix A for more details). This was done to avoid having to rely on prespecified criteria that turned out to be inappropriately selective or inclusive. Psychomotor retardation alone was assessed with HAMD item 8.

2.2.6 Analysis

Missing HAMD and MADRS scores were imputed by carrying forward the last nonmissing score to the end of the trial, prior to calculating response rates. This was done to make the analysis comparable to Geddes et al. (11) and the original trials (12). The analysis then proceeded in four sequential steps. First, the scale-derived melancholia variable was created and compared to the SCID variable. Second, treatment effects were estimated by calculating response-rate differences (i.e., risk differences) between treatment groups, with Wilson-

Newcombe confidence intervals and number-needed-to-treat (NNT) statistics. This was done stratified by melancholic status and baseline depression severity. Third, logistic regression was used to test for treatment effect heterogeneity across melancholic subgroups, if suggested by the response rates. Models were stratified by melancholic status, with treatment condition predicting response, and then repeated unstratified with a melancholia-by-treatment interaction. We also included baseline severity as a covariate in the models. No other covariates (e.g., age, gender) were considered. Fourth, the sequence above was repeated with psychomotor retardation as the subgroup variable instead of melancholic status.

2.3 Results

2.3.1 Melancholic subgroups

The process for creating a scale-derived melancholia variable is described in Appendix A. The final criteria are presented in Table 2.2, with baseline factors stratified by both variables. Melancholic status was associated with higher baseline depression scores, as expected, but not age, sex, or bipolar type. The two variables had poor agreement (61.1%; $\kappa = .22$) despite having similar prevalences (see Table 2.2).

2.3.2 Response rates

Response rates stratified by melancholic status and baseline depression scale scores (dichotomized at the mean) are presented in Table 2.3. As a point of reference, for the entire sample, the response-rate differences were 8.4 (NNT = 11.9), 7.0 (NNT = 14.2), and 8.9 (NNT = 11.3) with the HAMD-17, HAMD-31, and MADRS, respectively. SCID melancholia was associated with lower response rates in both treatment conditions such that the response-rate differences were comparable across subgroups, similar to those in the entire sample. In contrast, while the scale-derived melancholic subgroup also had lower lamotrigine response rates, the placebo response rates were even lower, resulting in somewhat larger treatment effects (though less pronounced with the HAMD-31).

Patients with higher HAMD-17 scores had larger response-rate differences across scales, but the opposite pattern was found with severity defined by the HAMD-31 and MADRS. Although scale-derived melancholia and HAMD-17 severity were both associated with numerically larger response-rate differences, this benefit appeared to be largely confined to patients meeting both criteria (see Figure 2.1). However, we did not test this further as it was not part of the analysis plan.

2.3.3 Logistic regressions

Proceeding with the scale-derived variable, in stratified models, the main effects of lamotrigine were statistically significant in the melancholic subgroup (HAMD-17, *OR* = 1.71, *SE* = .33, *p* = .01; HAMD-31, *OR* = 1.47, *SE* = .28, *p* = .04; MADRS, *OR* = 1.70, *SE* = .32, *p* = .01) but not in the nonmelancholic subgroup (HAMD-17, *OR* = 1.24, *SE* = .21, *p* = .20; HAMD-31, *OR* = 1.25, *SE* = .21, *p* = .17; MADRS, *OR* = 1.27, *SE* = .21, *p* = .15). In full models, the melancholic-treatment interactions did not reach statistical significance (HAMD-17, *OR* = 1.38, *SE* = .35, *p* = .21; HAMD-31, *OR* = 1.17, *SE* = .29, *p* = .54; MADRS, *OR* = 1.33, *SE* = .33, *p* = .25) and did not change after adjusting for baseline severity (HAMD-17, *OR* = 1.36, *SE* = .35, *p* = .23; HAMD-31, *OR* = 1.17, *SE* = .29, *p* = .54; MADRS, *OR* = 1.34, *SE* = .33, *p* = .25).

2.3.4 Psychomotor retardation

Response rates over psychomotor retardation scores are presented in Appendix B (Figure B.1). Response-rate differences tended to decrease slightly as scores increased from 0 (no retardation, *n* = 201) to 1 (slight retardation, *n* = 420) to 2 (obvious retardation, *n* = 400). Only at scores ≥ 3 (interview difficult/stupor, *n* = 51) were the response-rate differences larger (10.6-18.5, NNT = 5.4-9.5) compared to the rest of the sample (6.78-8.30, NNT = 12.0-14.7). We did not use this cut-off score to conduct additional analyses because the subgroup was so small.

2.3.5 Sensitivity analyses

These were conducted post hoc to test the robustness of the treatment effect in the scale-derived melancholic subgroup. It remained significant sequentially excluding data from each study (except on HAMD-31 response) and was comparable between bipolar types (see Tables B.1 and B.2, Appendix B). There was modest agreement between scale-derived melancholia at baseline and a week earlier at the screening visit (78.4%; κ = .56). The screening-visit variable also agreed poorly with the SCID (59.4%; κ = .19) and was associated with significant treatment effects (*OR* = 1.57-1.58, all *p* = .02). An alternate scale-derived variable was created with a different algorithm (see Appendix A) to see if results were sensitive to the method by which the original variable was specified. It had a similar prevalence (47.7%) and poor agreement with the SCID (61.9%; κ = .24), and the response rates were nearly identical, with significant effects in the melancholic subgroup only (see Table B.3, Appendix B).

2.4 Discussion

We were able to replicate the pilot study result in that treatment effects were numerically larger in patients with baseline scale item scores consonant with the DSM melancholic specifier, although the subgroup-treatment interactions did not reach statistical significance. Overall, the hypothesis was not clearly supported—melancholic depression diagnosed with the SCID, a standard measure, was not associated with larger treatment effects, and the scale-derived variable had poor agreement with the SCID. Moreover, there did not appear to be a monotonic relationship between psychomotor retardation and the response-rate difference, although the small subgroup with the highest scores seemed to benefit from lamotrigine over placebo.

Our interpretation is that the scale-derived variable selected patients who were less likely to respond to placebo, resulting in somewhat larger treatment effects, but they did not necessarily meet the SCID melancholic specifier, and we cannot know if they would have been classified as melancholic with another diagnostic system. Interestingly, the effect of baseline HAMD-17 severity on response rates, previously reported by Geddes et al. (11), seemed to be more pronounced in the scale-derived melancholic subgroup. Considering the other depression scales, higher baseline scores per se did not appear to result in larger treatment effects. Therefore, the present results have expanded upon the Geddes et al. meta-analysis by suggesting certain scale items are more relevant in this regard.

Several points are pertinent for future research. First, it cannot be assumed that a melancholic diagnosis made post hoc with scale items will necessarily agree with the SCID. A notable difference between measures is that the SCID considers symptoms during the entire depressive episode (which may have resolved by the time of assessment) whereas the scales measure currently active symptoms (13-16). Second, sum scores from different depression scales may not always be associated with treatment outcomes in the same way. A substantial amount of research has examined whether antidepressant effects are moderated by baseline HAMD-17 sum scores (18), although this study supports the notion that some items should be weighed more heavily (19). In general, the results reinforce the importance of replicating secondary analyses of clinical trials with additional data, using a variety of methods, before drawing firm conclusions.

A major limitation is there was only one standard assessment of melancholia that relied on the DSM criteria. Although common, these criteria have been criticized, often for over-diagnosing melancholia (9, 10). It is possible other diagnostic measures would have produced different results. Psychomotor retardation was also assessed with a single item; this part of the

analysis was exploratory and cannot speak to the importance of psychomotor retardation as a melancholic sign (9, 10). As with most industry-sponsored trials, the extensive list of exclusion criteria limits generalizability. It is also possible that differences by melancholic status would have been more pronounced in an inpatient sample (5). The hypothesis itself could be criticized because the rationale was that lamotrigine might benefit patients with HPA axis hyperactivity, but rather than testing this directly, we focused on a clinical diagnosis known to be associated with HPA axis hyperactivity. Having to infer the presence of biological abnormalities from clinical diagnoses, knowing these are unlikely to be present in every case, is a frustrating limitation that unfortunately also still pervades much of contemporary psychiatry.

Table 2.1

Descriptions of clinical trials used in the pooled analysis

	SCAB2001	SCAA2010	SCA40910	SCA100223	SCA30924
Sample	Bipolar I	Bipolar I/II	Bipolar I	Bipolar II	Bipolar I
	LTG: *n* = 63	LTG: *n* = 103	LTG: *n* = 133	LTG: *n* = 111	LTG: *n* = 131
	PBO: *n* = 66	PBO: *n* = 103	PBO: *n* = 124	PBO: *n* = 110	PBO: *n* = 128
Final dose	200 mg	100-400 mg	200 mg	200 mg	200 mg
Duration	7 weeks	10 weeks	8 weeks	8 weeks	8 weeks
Inclusion criteria	HAMD-17 score ≥ 18	HAMD-17 score ≥ 18	HAMD-17 score ≥ 18	HAMD-17 score ≥ 18	HAMD-17 score ≥ 18
	MDE ≥ 2 weeks	MDE ≥ 2 weeks	MDE ≥ 2 weeks	MDE ≥ 8 weeks	MDE ≥ 8 weeks
	Last 10 years:	Last 10 years:	Last 5 years:	HAMD-17 score ≥ 3 on items 1 or 7	HAMD-17 score ≥ 3 on items 1 or 7
	≥ 2 mood episodes, 1 manic or mixed	Type I: ≥ 2 mood episodes, 1 manic or mixed	≥ 2 mood episodes, 1 manic or mixed		Past hospitalization for mood disorder, or incarceration for mania
		Type II: ≥ 1 MDE and ≥ 2 hypomanic episodes			

Note. SCA100223 and SCA30924 also excluded patients with prior nonresponse to ≥ 2 antidepressants, or if HAMD-17 scores changed more than 20% from screening to baseline. More detailed descriptions can be found at https://www.gsk-studyregister.com. LTG = lamotrigine; PBO = placebo; MDE = major depressive episode; HAMD-17 = 17-item Hamilton Depression Rating Scale.

Table 2.2

Baseline descriptive statistics stratified by melancholic status

Variable	Melancholia (SCID)		Melancholia (scale-derived)	
	Present	Absent	Present	Absent
n (%)	518 (48.3)	554 (51.7)	479 (44.7)	593 (55.3)
LTG, *n* (%)	253 (48.8)	288 (52.0)	243 (50.7)	298 (50.3)
Female, *n* (%)	302 (58.3)	324 (58.5)	279 (58.3)	347 (58.5)
Age, *M* (SD)	39.5 (12.1)	38.5 (11.7)	39.0 (11.6)	39.0 (12.2)
Bipolar I, *n* (%)	371 (71.6)	396 (71.5)	344 (71.8)	423 (71.3)
HAMD-17, *M* (SD)	25.3 (3.88)	23.5 (3.57)	26.0 (3.76)	23.1 (3.37)
	$F = 66.2$***, $R^2 = .06$		$F = 181.0$***, $R^2 = .14$	
HAMD-31, *M* (SD)	36.9 (6.52)	34.1 (6.35)	37.5 (6.65)	33.9 (6.07)
	$F = 50.8$***, $R^2 = .05$		$F = 86.7$***, $R^2 = .07$	
MADRS, *M* (SD)	30.8 (5.45)	28.3 (5.82)	33.0 (4.29)	26.6 (5.26)
	$F = 52.6$***, $R^2 = .05$		$F = 449.4$***, $R^2 = .30$	

Note. The criteria used to establish melancholic status with scale items were anhedonia (MADRS item 8 ≥ 4) or nonreactive mood (MADRS items 1 or 2 ≥ 5), and at least three of: psychomotor disturbance (HAMD items 8 or 9 ≥ 1), guilt (HAMD item 2 ≥ 1), late insomnia (HAMD item 6 ≥ 1), or appetite/weight loss (HAMD items 12 or 16 = 2). One-way ANOVA was used to test depression score differences by melancholic status. LTG = lamotrigine; HAMD = Hamilton Depression Rating Scale (17- and 31-item versions); MADRS = Montgomery–Åsberg Depression Rating Scale.

***$p < .001$

Table 2.3

Response rates stratified by melancholic status and baseline depression scale scores dichotomized at the sample mean

	HAMD-17				HAMD-31				MADRS			
Subgroup	LTG	PBO	Diff (95% CI)	NNT	LTG	PBO	Diff (95% CI)	NNT	LTG	PBO	Diff (95% CI)	NNT
SCID-MEL +	42.3	34.0	8.3 (-0.01-16.5)	12.0	42.7	34.3	8.4 (-0.01-16.6)	12.0	44.3	36.6	7.7 (-0.8-16.0)	13.0
SCID-MEL –	47.2	39.1	8.1 (-0.1-16.2)	12.3	48.3	42.9	5.4 (-2.9-13.6)	18.5	51.4	41.7	9.7 (1.3-17.8)	10.3
Scale-MEL +	42.4	30.1	12.3 (3.7-20.6)	8.1	42.0	33.1	8.9 (0.3-17.4)	11.2	46.5	33.9	12.6 (3.8-21.1)	7.9
Scale-MEL –	47.0	41.7	5.3 (-2.7-13.2)	18.9	48.7	43.1	5.6 (-2.4-13.5)	17.8	49.3	43.4	5.9 (-2.1-13.8)	16.8
HAMD-17 ≥ 25	43.6	30.0	13.7 (5.0-22.0)	7.3	43.6	30.4	13.2 (4.6-21.6)	7.6	45.7	31.2	14.5 (5.8-22.8)	6.9
HAMD-17 ≤ 24	46.0	41.8	4.1 (-3.8-12.0)	24.2	47.3	45.2	2.1 (-5.9-10.0)	48.1	50.0	45.6	4.4 (-3.6-12.4)	22.6
HAMD-31 ≥ 36	39.2	36.3	3.0 (-5.5-11.3)	33.8	42.0	38.6	3.3 (-5.2-11.8)	30.2	42.4	38.2	4.1 (-4.4-12.5)	24.4
HAMD-31 ≤ 35	50.0	36.8	13.2 (5.0-21.1)	7.6	49.0	38.6	10.4 (2.2-18.4)	9.6	53.1	40.0	13.1 (4.9-21.1)	7.6
MADRS ≥ 30	40.5	34.7	5.9 (-2.1-13.7)	17.0	39.5	36.1	3.4 (-4.6-11.3)	29.5	45.4	37.2	8.1 (0.04-16.1)	12.3
MADRS ≤ 29	50.0	38.5	11.5 (2.8-19.9)	8.7	52.8	41.2	11.6 (2.9-20.0)	8.7	51.2	41.2	10.0 (1.3-18.4)	10.0

Note. Response rates and differences were proportions converted to percentages. HAMD = Hamilton Depression Rating Scale (17- and 31-item versions); MADRS = Montgomery–Åsberg Depression Rating Scale; LTG = lamotrigine; PBO = placebo; Diff = risk difference; NNT = number needed to treat; SCID = Structured Clinical Interview for DSM-IV; MEL = melancholic status (present or absent); Scale = scale-derived variable.

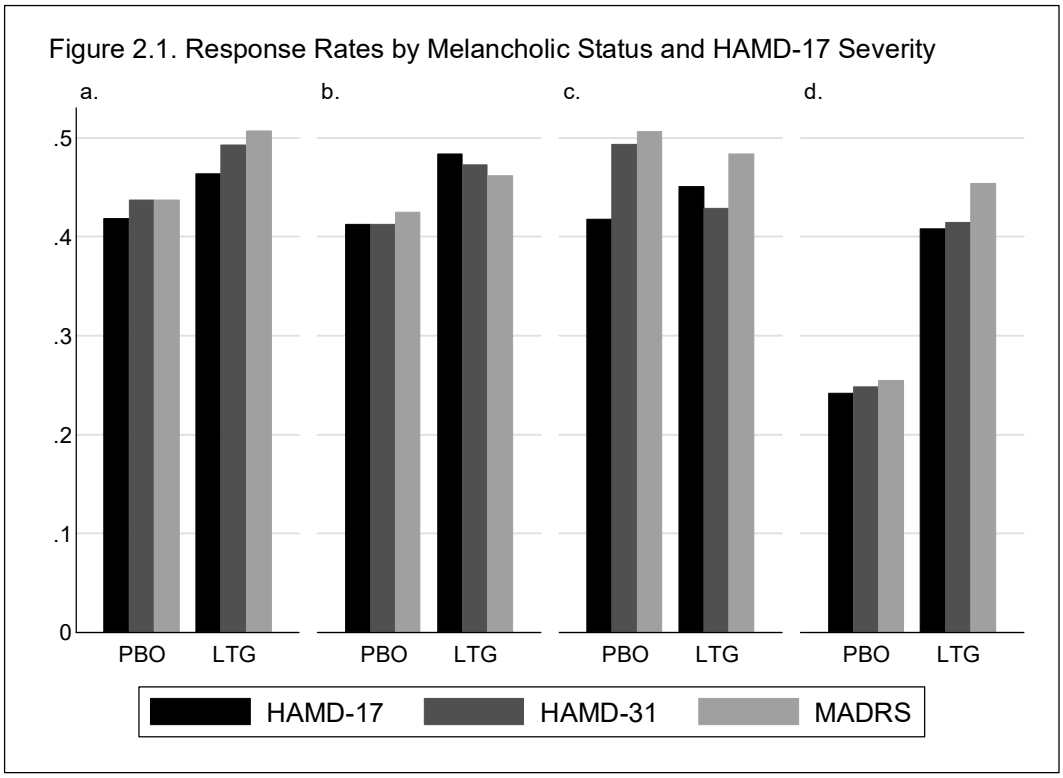

Figure 2.1. Response rates (y-axis) are presented as proportions. These were calculated with three different rating scales: the HAMD-17, the HAMD-31, and the MADRS. Subgroups were (a) nonmelancholic and baseline HAMD-17 ≤ 24 (n = 422), (b) nonmelancholic and baseline HAMD-17 ≥ 25 (n = 171), (c) melancholic and baseline HAMD-17 ≤ 24 (n = 170), and (d) melancholic and baseline HAMD-17 ≥ 25 (n = 309). In the subgroup with scale-derived melancholia and HAMD-17 scores ≥ 25, the response-rate differences were between 16.6 (NNT = 6.0) and 19.9 (NNT = 5.0). HAMD = Hamilton Depression Rating Scale (17- and 31-item versions); PBO = placebo; LTG = lamotrigine; MADRS = Montgomery–Åsberg Depression Rating Scale.

2.5 References

1. Peters EM, Bowen R, Balbuena L. Melancholic symptoms in bipolar II depression and responsiveness to lamotrigine in an exploratory pilot study. J Clin Psychopharmacol 2018;38(5):509-512.

2. Ketter TA, Manji HK, Post RM. Potential mechanisms of action of lamotrigine in the treatment of bipolar disorders. J Clin Psychopharmacol 2003;23:484-495.

3. Makatsori A, Duncko R, Moncek F, et al. Modulation of neuroendocrine response and non-verbal behavior during psychosocial stress in healthy volunteers by the glutamate release-inhibiting drug lamotrigine. Neuroendocrinology 2004;79:34-42.

4. Brown ES, Wolfshohl J, Shad MU, et al. Attenuation of the effects of corticosteroids on declarative memory with lamotrigine. Neuropsychopharmacology 2008;33:2376-2383.

5. Carroll BJ. The dexamethasone suppression test for melancholia. Brit J Psychiatry 1982;140:292-304.

6. Gold P. The organization of the stress system and its dysregulation in depressive illness. Mol Psychiatry 2015;20:32-47.

7. Bosaipo NB, Foss MP, Young AH, et al. Neuropsychological changes in melancholic and atypical depression: a systematic review. Neurosci Biobehav Rev 2017;73:309-325.

8. Mitchell PB, Hadzi-Pavlovic D, Evoniuk G, et al. A factor analytic study in bipolar depression, and response to lamotrigine. CNS Spectr 2013;18:214-224.

9. Martino DJ, Szmulewicz AG, Valerio MP, et al. Melancholia: an attempt at definition based on a review of empirical data. J Nerv Ment Dis 2019;207:792-798.

10. Taylor MA, Fink M. Restoring melancholia in the classification of mood disorders. J Affect Disord 2008;105:1-14.

11. Geddes JR, Calabrese JR, Goodwin GM. Lamotrigine for treatment of bipolar depression: independent meta-analysis and meta-regression of individual patient data from five randomised trials. Brit J Psychiatry 2009;194:4-9.

12. Calabrese JR, Huffman RF, White RL, et al. Lamotrigine in the acute treatment of bipolar depression: results of five double-blind, placebo-controlled clinical trials. Bipolar Disord 2008;10:323-333.

13. First M, Gibbon M, Spitzer RL, et al. Structured Clinical Interview for DSM-IV Axis I Disorders Research Version (SCID-I). New York: Biometric Research, New York State Psychiatric Institute, 1996.

14. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;23:56-62.

15. Williams J, Link M, Rosenthal N, et al. Structured Interview Guide for the Hamilton Depression Rating Scale-Seasonal Affective Disorder Version (SIGH-SAD). New York: New York State Psychiatric Institute, 1988.

16. Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. Brit J Psychiatry 1979;134:382-389.

17. Vöhringer PA, Ghaemi SN. Solving the antidepressant efficacy question: effect sizes in major depressive disorder. Clin Ther 2011;33:B49-B61.

18. Furukawa TA, Maruo K, Noma H, et al. Initial severity of major depression and efficacy of new generation antidepressants: individual participant data meta-analysis. Acta Psychiatr Scand 2018;137:450-458.

19. Trivedi MH, South C, Jha MK, et al. A novel strategy to identify placebo responders: prediction index of clinical and biological markers in the EMBARC Trial. Psychother Psychosom 2018;87:285-295.

2.6 Appendix A

It was deemed necessary to combine MADRS and HAMD items to create a melancholic variable based on DSM criteria because HAMD items 1 and 7 measure low mood and reduced interest but they do not explicitly assess lack of pleasure/reactivity, and item 7 is also confounded by functional impairment (i.e., unable to work due to illness) that may or may not be due to anhedonia. In contrast, MADRS item 8 (inability to feel) more clearly describes consumptive anhedonia, and items 1 and 2 describe mood/affect nonreactivity at scores > 4, but the MADRS does not have dedicated items for psychomotor retardation and terminal insomnia. Therefore, we used MADRS items to assess Criterion A and HAMD items for Criterion B (see Table A.1).

Table A.1

Scale items used to assess melancholic features

DSM melancholic features	Corresponding scale items
Criterion A (one of): ▪ Lack of reactivity to normally pleasurable stimuli ▪ Loss of pleasure in all or most activities	Montgomery–Åsberg Depression Rating Scale 1. Observed despondency/despair in facial expressions, speech, posture; rated by depth and inability to brighten 2. Subjectively reported depressed mood; rated by intensity, duration, and lack of reaction to external events 8. Reduced interest in surroundings or normally pleasurable activities; lack of emotional reaction to circumstances
Criterion B (at least three of): ▪ Excessive guilt ▪ Early-morning awakening ▪ Psychomotor agitation or retardation ▪ Anorexia or weight loss	Hamilton Depression Rating Scale 2. Feelings of guilt (incl. ideas, delusions, hallucinations) 6. Late insomnia (waking up early, cannot return to sleep) 8. Slowness of thinking, speech, and activity (observed) 9. Agitation and restlessness (observed) 12. Loss of appetite, needs urging to eat, requires laxatives 16. Weight loss (reported or measured)

Note. The preliminary criteria were MADRS item 8 ≥ 4, or items 1 or 2 = 6, with at least three of: HAMD item 2 ≥ 2, items 8 or 9 ≥ 2, items 12 or 16 = 2, and item 6 ≥ 1. The DSM Criterion B feature that mood has a "distinct quality" of profound despondency, despair, or moroseness was not assessed separately, as it has no dedicated scale item. Mood being regularly worse in the morning could also not be assessed as the time of day was not recorded in all datasets.

As per the analysis plan, the criteria for the scale-derived variable were modified until the prevalence was between 40-60%, similar to the SCID (48.3%). The preliminary criteria (see Table A.1) resulted in a prevalence of 28.9%. We then examined each item and found only 12 patients scored 6 on MADRS items 1 or 2, so this cut-off seemed unrealistically high (by comparison, other features were present in 24.4-81.2% of the sample). The two items were loosened to ≥ 5, but still the prevalence was only 29.8%. We did not loosen the Criterion A items further because lower scores indicate the retained ability to feel pleasure and mood being reactive to external events.

At this point we could either loosen the HAMD item scores or require only two Criterion B features be present. The former approach was considered more desirable because the DSM requires at least three Criterion B features. However, the latter was thought to be defensible because these patients would have had high MADRS item 1 and 2 scores (on which high scores are supposed to reflect intense despondency/misery, different from ordinary low spirits) such that they could arguably be described as suffering from "profound despondency, despair, or moroseness" bringing the total number of Criterion B features to three.

Proceeding with the former approach, we sequentially loosened HAMD item scores to ≥ 1 (first items 8 and 9, then 2) until the prevalence was above 40%, the minimum specified a priori in the analysis plan. This resulted in a variable with a prevalence of 44.7%, but it had poor agreement with the SCID (61.1%; $\kappa = .22$).

We then attempted to create another variable that required only two Criterion B features, and with each HAMD item score ≥ 2 the prevalence was 47.7%, very close to the SCID. However, the agreement was still poor (61.9%; $\kappa = .24$). Therefore, we retained the first variable, as it was closer to the preliminary criteria, although we re-estimated treatment effects with the second variable (see Appendix B) to see if the results were sensitive to the chosen method (the choice seeming somewhat arbitrary at this point, considering both methods resulted in equally poor agreement with the SCID).

2.7 Appendix B

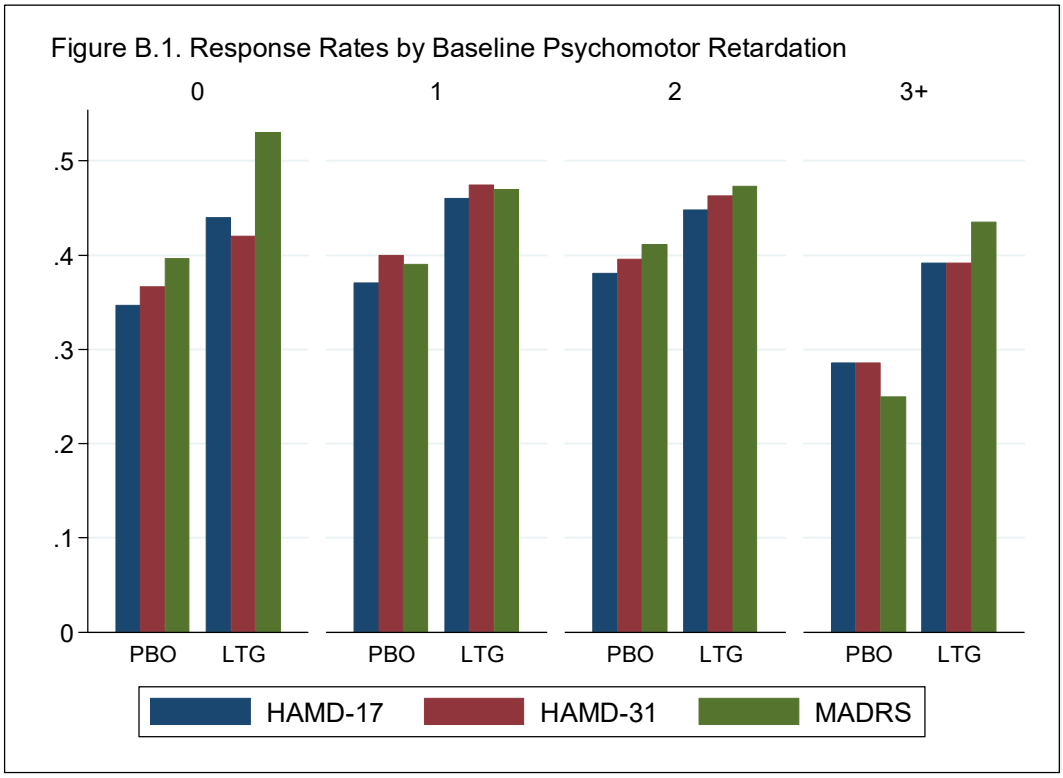

Figure B.1. Response rates (y-axis) are presented as proportions. The numbers above each subgraph indicate baseline HAMD item 8 scores. Response rates were calculated with three different rating scales: the HAMD-17, the HAMD-31, and the MADRS. PBO = placebo; LTG = lamotrigine; HAMD = Hamilton Depression Rating Scale (17- and 31-item versions); MADRS = Montgomery–Åsberg Depression Rating Scale.

Table B.1

Treatment effects in the scale-derived melancholic subgroup sequentially excluding one trial

Excluded	HAMD-17			HAMD-31			MADRS		
	OR (*SE*)	*z*	*p*	*OR* (*SE*)	*z*	*p*	*OR* (*SE*)	*z*	*p*
SCAB2001	1.67 (.34)	2.52	.012	1.41 (.28)	1.70	.088	1.53 (.30)	2.15	.031
SCAA2010	1.74 (.36)	2.65	.008	1.45 (.30)	1.81	.070	1.66 (.34)	2.48	.013
SCA40910	1.72 (.37)	2.51	.012	1.51 (.32)	1.94	.052	1.86 (.40)	2.93	.003
SCA100223	1.60 (.36)	2.12	.034	1.38 (.30)	1.46	.146	1.64 (.36)	2.28	.023
SCA30924	1.84 (.42)	2.67	.007	1.61 (.36)	2.11	.035	1.85 (.42)	2.73	.006

Note. Treatment effects were estimated with logistic regression models with treatment condition predicting response on three different rating scales: the HAMD-17, the HAMD-31, and the MADRS. HAMD = Hamilton Depression Rating Scale (17- and 31-item versions); MADRS = Montgomery–Åsberg Depression Rating Scale.

Table B.2

Response rates in the scale-derived melancholic subgroup by bipolar type

	Bipolar I (*n* = 344)				Bipolar II (*n* = 135)			
Outcome	LTG	PBO	Diff (95% CI)	NNT	LTG	PBO	Diff (95% CI)	NNT
HAMD-17	41.7	29.0	12.7 (2.6-22.5)	7.9	44.1	32.8	11.3 (-5.0-26.8)	8.9
HAMD-31	40.6	30.8	9.8 (-0.3-19.6)	10.2	45.6	38.8	6.8 (-9.7-22.7)	14.7
MADRS	45.7	33.1	12.6 (2.2-22.5)	8.0	48.5	35.8	12.7 (-3.9-28.3)	7.9

Note. Response rates and differences were proportions converted to percentages. LTG = lamotrigine; PBO = placebo; HAMD = Hamilton Depression Rating Scale (17- and 31-item versions); MADRS = Montgomery–Åsberg Depression Rating Scale; Diff = risk difference; NNT = number needed to treat.

Table B.3

Response rates in the alternate scale-derived melancholic subgroup

	Melancholic				Nonmelancholic			
Outcome	LTG	PBO	Diff (95% CI)	NNT	LTG	PBO	Diff (95% CI)	NNT
HAMD-17	42.2	29.4	12.8 (4.4-20.8)	7.83	47.5	42.8	4.7 (-3.5-12.9)	21.2
HAMD-31	42.6	31.0	11.5 (3.2-19.7)	8.67	48.6	45.2	3.3 (-4.9-11.5)	30.0
MADRS	45.2	33.1	12.2 (3.7-20.4)	8.21	50.7	44.5	6.2 (-2.1-14.3)	16.1

Note. Response rates and differences were proportions converted to percentages. LTG = lamotrigine; PBO = placebo; HAMD = Hamilton Depression Rating Scale (17- and 31-item versions); MADRS = Montgomery–Åsberg Depression Rating Scale; Diff = risk difference; NNT = number needed to treat.

Transition Note 1

The prespecified analysis reported in Manuscript 1 unfortunately raised more questions than it answered. Because there is no way to establish the validity of the scale-derived variable, we cannot interpret the results as supporting the hypothesis. The most puzzling finding, however, was the contradictory results across measures of baseline severity. It is also somewhat concerning because the larger treatment effect reported in the Geddes et al. meta-analysis is still the most convincing evidence supporting lamotrigine monotherapy for acute bipolar depression. When we inspected response rates stratified by scale-derived melancholic status and HAMD-17 severity, it appeared that only patients in both subgroups experienced larger treatment effects. One possible explanation is that only certain scale items were associated with response, and that we used one or some of these items to create the scale-derived melancholic variable. If the dichotomized severity variables captured different proportions of patients scoring higher on these items then this could also explain the contradictory results across the three severity variables. Therefore, we obtained permission from the sponsor and independent review panel to conduct a series of exploratory analyses to try to determine which specific scale items were associated with response, with the hope of providing some clarity to the question of symptom-level treatment effect moderators in the pooled dataset. In addition to considering response rates, we also examined treatment effects on individual item scores in an attempt to establish consistency across outcome measures. Based on the breadth of the analysis, it was decided that these should be presented in a separate manuscript, which is presented in the next chapter.

CHAPTER 3

Manuscript 2

Response to lamotrigine for acute bipolar depression: an exploratory item-level analysis

ABSTRACT

Lamotrigine is used to treat bipolar depression despite inconsistent evidence. This study was an exploratory analysis of pooled data from five randomized placebo-controlled trials of lamotrigine for acute bipolar depression. The goal was to determine if certain depression scale items were more responsive to lamotrigine treatment, and if patients scoring higher on these items at baseline experienced larger overall treatment effects. The pooled sample contained 1,072 adult outpatients treated for up to 7-10 weeks. Depressive symptoms were measured with the Hamilton Depression Rating Scale and the Montgomery–Åsberg Depression Rating Scale. Change scores on individual scale items were compared between treatment groups. End-trial response rates were calculated with total depression scale scores as a measure of overall treatment response. There were statistically significant effects on items assessing depressed mood/sadness, lack of interest/anhedonia, pessimism/guilt, and anergia/fatigue, on both scales. However, there was marked variation in the baseline symptom prevalence, and items with higher scores at baseline tended to have larger and statistically significant treatment effects. Graphing response rates over baseline item scores indicated that patients with more severe depressed mood and/or anhedonia at baseline tended to have larger overall response-rate differences. Given the exploratory nature of the analysis, firm conclusions cannot be drawn, although the pattern of item-level effects was consistent with past research. This suggested that a floor effect limited the sensitivity of certain scale items, and that relying on total depression scale sum scores over targeted assessments of core depressive symptoms may have impeded signal detection in the original trials.

Keywords: anticonvulsant; clinimetrics; rating scale; mood disorder

3.1 Introduction

Lamotrigine is used to treat acute bipolar depression despite inconsistent evidence (1). Five industry-sponsored monotherapy trials initially failed to find a significant benefit over placebo (2), yet maintenance trials (3) and more recent adjunct trials (4, 5) were supportive. Additional research has suggested specific symptoms and patient subgroups with higher baseline depression scores (that were not assessed separately in the monotherapy trials) may be more responsive to lamotrigine treatment (6-10). This could explain the apparent clinical utility, as evidenced by increasing lamotrigine prescriptions (11, 12), despite overall negative results from the monotherapy trials. Here we report the results of an exploratory item-level analysis of the five monotherapy trials in an attempt to identify specific depressive symptoms associated with responsiveness to lamotrigine over placebo.

3.2 Methods

3.2.1 Design

This study was a pooled analysis of five randomized, double-blind, placebo-controlled, parallel-group, monotherapy trials of lamotrigine for acute bipolar depression (2). Data access was requested from the original trial sponsors through clinicalstudydatarequest.com. The analyses reported here were conducted post hoc after another study (reported elsewhere) with permission from the sponsor and an independent review panel. The trials are described briefly here and in more detail elsewhere (2). In all trials, lamotrigine was administered with a fixed-dose titration up to 200 mg/day, except in trial SCAA2010 in which it was dosed flexibly at 100-400 mg/day. We did not include data from the fixed-dose 50 mg/day arm in trial SCAB2001 as this dose is considered subtherapeutic (1, 7). Double-blind treatment lasted for 7 to 10 weeks (depending on the trial), and we included all data up to the latest time point.

3.2.2 Sample

The pooled sample consisted of 1,072 adult outpatients (age ≥ 18 years, $M = 39.0$, $SD = 11.9$; 58.4% female) currently experiencing an acute major depressive episode (duration ≥ 2 or 8 weeks, depending on the trial) with a diagnosis of bipolar disorder type I ($n = 767$) or II ($n = 305$). Diagnoses were confirmed with the Structured Clinical Interview for DSM-IV (13). A 17-item Hamilton Depression Rating Scale (HAMD-17) (14) score ≥ 18 was also required. Additional inclusion criteria varied between trials (described elsewhere) (2). Common exclusion criteria were previous lamotrigine treatment; concurrent or recent (within five half-lives)

psychotropic medication (except some benzodiazepines); psychotherapy (recently started); abnormal thyroid tests; epilepsy; active suicidality; panic disorder, bulimia nervosa, obsessive-compulsive disorder, or social phobia in the last 12 months; DSM rapid-cycling subtype; substance abuse/dependence; and medical conditions that could interfere with treatment (2).

3.2.3 Instruments

Depressive symptoms were measured with the 17- and 31-item versions of the Hamilton Depression Rating Scale (HAMD-17 and HAMD-31) (14, 15) and the 10-item Montgomery–Åsberg Depression Rating Scale (MADRS) (16). For part of the analysis, each item (41 in total) was considered in isolation. HAMD-17 and MADRS sum scores were also used to estimate treatment effects. Missing scores were filled by carrying forward the last observation, consistent with the original trials (2).

3.2.4 Analysis

There were three components to the analysis. First, we attempted to determine which scale items were more responsive to lamotrigine treatment compared to placebo. To do this, change scores (baseline – final score) for each depression scale item were compared between treatment groups with Wilcoxon rank-sum tests. A nonparametric test was chosen in case some items (particularly those with restricted ranges) were not normally distributed. For each item, we also calculated mean baseline scores and the proportion of patients scoring zero at baseline to gauge the extent that treatment effects were related to baseline prevalence (17, 18).

For the second part of the analysis, we attempted to determine if the baseline items that were more responsive to lamotrigine treatment could also be used to identify *patients* who experienced a larger overall treatment effect. To do this, we plotted response rates over baseline scores on relevant items to see if these were positively correlated with the degree of separation between lamotrigine and placebo. A response was defined as an end-trial score reduction $\geq 50\%$ from baseline on the HAMD-17 and MADRS. The rationale for requiring a score reduction on both scales was to identify items that were relevant regardless of which scale was being used. Prior to creating plots, baseline item scores were collapsed to ordinal categories such that the tail scores contained at least roughly 5% of the sample (otherwise, in many cases, these would have contained $< 1\%$ of the sample).

After visually inspecting the graphs, we attempted to use baseline item cut-off scores to create binary subgroup variables that selected patients with increased response-rate differences,

compared to the rest of the sample. The subgroup-treatment effects were then tested in logistic regression models predicting response. Each model contained three terms: the main effect of lamotrigine, the main effect of subgroup status, and the subgroup-treatment interaction. A statistically significant interaction term was considered evidence of effect heterogeneity. Consistency across the five trials was also examined, but this was done only by inspecting response rates (i.e., looking for a numerical advantage) because interaction tests would likely be underpowered to detect an interaction (19).

In the third part of the analysis, we examined the association between subgroup status and baseline depression score severity. A previous meta-analysis suggested patients with higher baseline depression scores experienced larger response rates with lamotrigine compared to placebo (7). Therefore, we used total depression scale sum scores to create comparably sized patient subgroups, to determine if the items identified in the second part of the analysis could select treatment-responsive patients as effectively as total scores from the parent scale.

3.3 Results

3.3.1 Item-level change scores

HAMD-17 item-level baseline and change scores are presented in Table 3.1. Significant treatment effects were detected on items 1 (depressed mood), 2 (guilt), 7 (work and interest), and 13 (general somatic/fatigue). The respective Cohen's *d* effect sizes were 0.15, 0.14, 0.24, and 0.15. Among the additional HAMD-31 items, the only marginally significant effect in favor of lamotrigine was on item 28 (motor retardation; $p = .08$, $d = .09$). MADRS item-level scores (see Table 3.2) were generally consistent, with significant effects on items 1 (apparent sadness; $d = 0.15$), 2 (reported sadness; $d = 0.14$), 7 (lassitude; $d = 0.17$), 8 (inability to feel; $d = 0.20$), and 9 (pessimistic thoughts; $d = 0.15$). In general, items with significant or marginally significant effects tended to have higher baseline scores, with some exceptions (e.g., HAMD items 10 and 11 vs. 8 and 15; MADRS item 6 vs. 7 and 8; see Tables 3.1 and 3.2).

The items with significant or marginally significant effects roughly matched the HAMD-6 "melancholia" subscale (20-23). Because the HAMD-6 has established psychometric and clinimetric advantages over the HAMD-17 (20, 23), we decided to calculate change scores and response rates for this subscale, and a subscale of the six corresponding MADRS items, which are presented in Table 3.3. Each subscale had a larger effect size than the parent scale when

calculated with change scores, but with response rates, there was less difference between subscale and parent scale (i.e., the lamotrigine-placebo differences were similar).

3.3.2 Response rates by baseline item scores

From here, the graphical analysis included the HAMD-6 items and corresponding MADRS items. Response rates over baseline scores for each item are presented in Figure 3.1. Only for HAMD items 1 (depressed mood) and 7 (work and interests) did there appear to be a clear monotonic relationship between baseline score and the size of the treatment effect. Considering the analogous MADRS items, 1 (apparent sadness) and 8 (inability to feel), a monotonic relationship was less obvious, but at scores > 4 (or > 3 for item 8) the treatment effects appeared to be larger. There also appeared to be a slight advantage at higher scores of HAMD item 10 (psychic anxiety), but not with the analogous MADRS item, 3 (inner tension). For the remainder of the analysis, we decided to focus on the mood and interest/anhedonia items because of the consistency across scales.

3.3.3 Creation of subgroup variables

Dichotomous subgroup variables representing severe low mood or anhedonia were created with HAMD item scores (1 or 7 = 4) and the analogous MADRS items (1 or 8 > 4). A high score on only one item was required to maximize subgroup size, although these were still small (n = 138 and 160, respectively). Similarly, we created a third variable with all four items (only one high score required) to try to capture a larger subgroup (n = 240) as well as an alternate MADRS variable with loosened criteria that allowed item 8 scores to be > 3 (n = 638).

3.3.4 Subgroup treatment effects

Response rates stratified by the four subgroups are presented in Table 3.4. All variables selected patients with larger treatment effects, the loosened MADRS variable less so. The logistic regression subgroup-treatment interactions were statistically significant with the HAMD variable (OR = 2.83, SE = 1.14, p = .01) and the combined variable (OR = 2.49, SE = 0.81, p = .01) but did not reach statistical significance with the MADRS variables (OR = 1.82, SE = 0.70, p = .12; loosened criteria, OR = 1.56, SE = .40, p = .08) despite the numerical advantage.

Across all five trials, the response-rate differences were numerically larger in the selected subgroups, with one exception (a single trial using the loosened MADRS criteria). There was, however, considerable trial-to-trial variability, as can be seen in Figure 3.2. It is also worthwhile to point out that in trial SCA100223 and SCA30924, a score > 2 on HAMD item 1 or 7 was

required for inclusion, and this was not the case in the other trials; accordingly, the difference between HAMD-derived subgroups was the smallest in these trials.

3.3.5 Comparison with full depression scale scores

Patients with high scores on HAMD item 1 or 7 (which was the most exclusive subgroup with the largest subgroup-treatment effect) had higher baseline HAMD-17 scores, on average (M = 27.1, SD = 4.34 vs. M = 24.0, SD = 3.59; t = 9.22, $p < .001$). However, these ranged from 18 to 36, and only 43.4% scored > 27, an established cut-off for "severe" depression (24). Creating comparably sized subgroups by dichotomizing HAMD-17 scores did not select patients with comparably larger response-rate differences compared to the rest of the sample (HAMD-17 > 27, n = 229, difference = 9.83 vs. 7.66; HAMD-17 > 28, n = 173, difference = 4.80 vs. 8.21; HAMD-17 > 29, n = 120, difference = 4.82 vs. 8.54). We also tried using HAMD-6 sum scores, but there was no improvement (HAMD-6 > 14, n = 269, difference = 8.06 vs. 8.54; HAMD-6 > 15, n = 122, difference = 7.74 vs. 8.52).

We considered that baseline HAMD-17 and/or HAMD-6 sum scores could be more strongly associated with the response-rate difference if treated as continuous variables, and not dichotomized to create subgroups. However, in logistic regression models, neither significantly interacted with treatment condition to predict response (HAMD-17, OR = 1.01, SE = .03, p = .80; HAMD-6, OR = 1.05, SE = 0.7, p = .41). In contrast, the interaction with items 1 and 7, summed together and treated as a continuous variable, was statistically significant (OR = 1.42, SE = .22, p = .02). These interactions are depicted in Figure 3.3. Although there appeared to be a slight (nonsignificant) advantage at higher HAMD-6 scores, the predicted effects were homogenous across HAMD-17 scores.

3.4 Discussion

The most pertinent finding was that treatment effects on change scores for symptoms assessed by the HAMD-6 melancholia factor (in particular, anhedonia/interest, mood, energy, and guilt) were larger, with consistency across scales. This finding is important not because it is novel but because it has already been demonstrated with striking consistency for numerous second-generation antidepressants, albeit mostly in unipolar depression (20-23, 25-32). Although this analysis was exploratory, convergence with past research is reassuring. Previously it has been suggested that the lamotrigine monotherapy program failed, at least in part, due to the short trial durations combined with slow dose titrations, and suboptimal end-trial doses (1). The

current results suggest that relying on HAMD-17 and MADRS sum scores over targeted assessments of core depressive symptoms also impeded signal detection in the original trials.

The results derived from item-level change scores are also consistent with a prior reanalysis of two placebo-controlled lamotrigine trials (unipolar and bipolar depression) that found significant effects on factors assessing depressive cognitions and psychomotor retardation (8). However, in contrast to the current results, the anergia factor (containing the work/interest and general somatic items) did not clearly improve, possibly because it also contained the insight and libido items (8). Past research has suggested that lamotrigine may dampen affective instability and related symptoms such as irritability in patients with mood disorders (6, 9). Although these were not assessed in the current study, the positive effect on MADRS item 8 (inability to feel) argues against emotional blunting in this sample. It is not clear why HAMD paranoia item scores improved more with placebo than lamotrigine, although the effect was only marginally significant and could have been due to chance.

The items with the largest effects also tended to be more prevalent in the sample at baseline, although not invariably. This suggests a floor effect, to some extent, may have limited the sensitivity of other items. Consistently, there was no advantage using the HAMD-6 to calculate response rates, which, unlike change scores, are defined relative to baseline scores and are less susceptible to floor effects (18, 33). Alternatively, the lack of prominent effects on the items assessing sleep and appetite/weight would also be in keeping with lamotrigine being weight-neutral and not overly sedating (2, 34). The marked variation of baseline prevalence across the spectrum of depressive symptoms was similar to that reported in a recent pooled analysis of trial data from patients with unipolar depression (17). This should continue to be noted for future clinical trials, particularly if symptom-level effects are being examined—it would be difficult to detect an advantage over placebo if the majority of the sample is without the symptom in question.

We also found that patients with more severe low mood and decreased interest/anhedonia had larger response-rate differences in favor of lamotrigine, partially because their placebo response rates were lower. This is consistent with a recent machine-learning study that identified anhedonia and melancholia (which involves prominent anhedonia and severe low mood) as inverse predictors of placebo response in unipolar depression (35), an association for melancholia that is fairly well established (36). In a previous study, mood and anhedonia items

along with several others (sleep, appetite, agitation, etc.) were used to select a melancholic subgroup which appeared to benefit more from lamotrigine over placebo (10). In hindsight, it appears as if the same result could have been obtained with only two items, calling into question the assumption that these patients were necessarily melancholic, as per the DSM definition.

Given the inconsistent results regarding the efficacy of lamotrigine for bipolar depression, the larger treatment effect in patients with more severe depressed mood and decreased interest/anhedonia, who seemed to respond quite poorly to placebo, provides some reassurance that lamotrigine does possess an antidepressant effect. This interpretation echoes the conclusion drawn by Geddes et al. (7), who analyzed the same dataset and reported a similar finding in patients with baseline HAMD-17 scores above the sample mean. In this study, we did not find clear evidence of a subgroup-treatment effect when higher HAMD-17 scores were used to split the sample, nor was there a significant interaction between treatment condition and continuous HAMD-17 scores. However, given the exploratory nature of the analysis, it would be premature to suggest that the two HAMD items are a better index of severity and/or placebo nonresponse than total HAMD-17 sum scores. The results reported here need to be considered hypothesis-generating until replicated with additional data. At most, they illustrate the general notion that it may be possible to use individual symptom scores to predict treatment outcomes. It may be useful for future research in this area to consider mood and anhedonia items in isolation, and not only total depression scale scores.

The results also underscore the utility of understanding placebo response moderators. Precision medicine to guide the selection of antidepressants is an aspirational goal in psychiatry, but an index that predicts nonresponse to placebo, regardless of what medication is being prescribed, would be a useful tool for drug development. To the extent that some of the placebo response is a function of spontaneous recovery over time (37), it could also potentially help clinicians choose between active treatment and watchful waiting, although this would need to be tested directly.

Another limitation with the analysis was that it was fairly rudimentary. There are more sophisticated methods of model building and variable selection that could have produced different results. We also did not consider other potentially relevant baseline variables (age, episode duration, etc.). However, our goal was not to build the most comprehensive model, but specifically, to determine which scale items were associated with response. Based on past

research (7, 8, 10), we knew some items would likely be identified, and in this sense, the chosen method may be considered appropriate, although it remains unclear how these items might interact with other baseline variables. Not every item was considered in the creation of subgroups because the analysis was designed to move from broadly considering all items to narrowly focusing on key items relevant across outcomes. The results ultimately pointed toward the items that assess the only two symptoms necessary for a diagnosis of major depression, and it seemed reasonable to concentrate on these further. The use of single items to assess symptoms is another limitation, although it does permit comparison with a substantial number of industry-sponsored trials.

3.5 References

1. Yatham LN, Kennedy SH, Parikh SV, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder. Bipolar Disord 2018;20(2):97-170.

2. Calabrese JR, Huffman RF, White RL, et al. Lamotrigine in the acute treatment of bipolar depression: results of five double-blind, placebo-controlled clinical trials. Bipolar Disord 2008;10(2):323-333.

3. Goodwin GM, Bowden CL, Calabrese JR, et al. A pooled analysis of 2 placebo-controlled 18-month trials of lamotrigine and lithium maintenance in bipolar I disorder. J Clin Psychiatry 2004;65(3):432-441.

4. Geddes JR, Gardiner A, Rendell J, et al. Comparative evaluation of quetiapine plus lamotrigine combination versus quetiapine monotherapy (and folic acid versus placebo) in bipolar depression (CEQUEL): a 2× 2 factorial randomised trial. Lancet Psychiatry 2016;3(1):31-39.

5. van der Loos ML, Mulder PG, Hartong EGTM, et al. Efficacy and safety of lamotrigine as add-on treatment to lithium in bipolar depression: a multicenter, double-blind, placebo-controlled trial. J Clin Psychiatry 2009;70(2):223-231.

6. Bowen RC, Balbuena L, Baetz M. Lamotrigine reduces affective instability in depressed patients with mixed mood and anxiety disorders. J Clin Psychopharmacol 2014;34(6):747-749.

7. Geddes JR, Calabrese JR, Goodwin GM. Lamotrigine for treatment of bipolar depression: independent meta-analysis and meta-regression of individual patient data from five randomised trials. Brit J Psychiatry 2009;194(1):4-9.

8. Mitchell PB, Hadzi-Pavlovic D, Evoniuk G, et al. A factor analytic study in bipolar depression, and response to lamotrigine. CNS Spectr 2013;18(4):214-224.

9. Peters EM, Bowen R. Irritability in a mixed sample of patients with unipolar and bipolar II depression predicts responsiveness to lamotrigine. Pers Med Psychiatry 2018;7:27-30.

10. Peters EM, Bowen R, Balbuena L. Melancholic symptoms in bipolar II depression and responsiveness to lamotrigine in an exploratory pilot study. J Clin Psychopharmacol 2018;38(5):509-512.

11. Karanti A, Kardell M, Lundberg U, Landén M. Changes in mood stabilizer prescription patterns in bipolar disorder. J Affect Disord 2016;195:50-56.

12. Kessing LV, Vradi E, Andersen PK. Nationwide and population-based prescription patterns in bipolar disorder. Bipolar Disord 2016;18(2):174-182.

13. First M, Gibbon M, Spitzer RL, et al. Structured Clinical Interview for DSM-IV Axis I Disorders Research Version (SCID-I). New York: Biometric Research, New York State Psychiatric Institute, 1996.

14. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;23:56-62.

15. Williams J, Link M, Rosenthal N, et al. Structured Interview Guide for the Hamilton Depression Rating Scale-Seasonal Affective Disorder Version (SIGH-SAD). New York: New York State Psychiatric Institute, 1988.

16. Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. Brit J Psychiatry 1979;134:382-389.

17. Hieronymus F, Lisinski A, Nilsson S, Eriksson E. Influence of baseline severity on the effects of SSRIs in depression: an item-based, patient-level post-hoc analysis. Lancet Psychiatry 2019;6(9):745-752.

18. Vöhringer PA, Ghaemi SN. Solving the antidepressant efficacy question: effect sizes in major depressive disorder. Clin Ther 2011;33:B49-B61.

19. Brookes ST, Whitely E, Egger M, et al. Subgroup analyses in randomized trials: risks of subgroup-specific analyses: power and sample size for the interaction test. J Clin Epidemiol 2004;57(3):229-236.

20. Carrozzino D, Patierno C, Fava GA, Guidi J. The Hamilton rating scales for depression: a critical review of clinimetric properties of different versions. Psychother Psychosom 2020;89(3):133-150.

21. Bech P. Rating scales in depression: limitations and pitfalls. Dialogues Clin Neurosci 2006;8(2):207-215.

22. Østergaard SD, Bech P, Trivedi MH, et al. Brief, unidimensional melancholia rating scales are highly sensitive to the effect of citalopram and may have biological validity: implications for the research domain criteria (RDoC). J Affect Disord 2014;163:18-24.

23. Timmerby N, Andersen JH, Søndergaard S, et al. A systematic review of the clinimetric properties of the 6-item version of the Hamilton Depression Rating Scale (HAM-D6). Psychother Psychosom 2017;86(3):141-149.

24. Müller MJ, Szegedi A, Wetzel H, Benkert O. Moderate and severe depression: gradations for the Montgomery–Åsberg depression rating scale. J Affect Disord 2000;60(2):137-140.

25. Lisinski A, Hieronymus F, Näslund J, et al. Item-based analysis of the effects of duloxetine in depression: a patient-level post hoc study. Neuropsychopharmacology 2020;45(3):553-560.

26. Faries D, Herrera J, Rayamajhi J, et al. The responsiveness of the Hamilton depression rating scale. J Psychiatr Res 2000;34(1):3-10.

27. Bech P, Tanghøj P, Andersen H, Overø K. Citalopram dose-response revisited using an alternative psychometric approach to evaluate clinical effects of four fixed citalopram doses compared to placebo in patients with major depression. Psychopharmacology 2002;163(1):20-25.

28. Schneider LS, Nelson JC, Clary CM, et al. An 8-week multicenter, parallel-group, double-blind, placebo-controlled study of sertraline in elderly outpatients with major depression. Am J Psychiatry 2003;160(7):1277-1285.

29. Bech P, Tanghøj P, Cialdella P, et al. Escitalopram dose–response revisited: an alternative psychometric approach to evaluate clinical effects of escitalopram compared to citalopram and placebo in patients with major depression. Int J Neuropsychopharmacol 2004;7(3):283-290.

30. Entsuah R, Shaffer M, Zhang J. A critical examination of the sensitivity of unidimensional subscales derived from the Hamilton depression rating scale to antidepressant drug effects. J Psychiatr Res 2002;36(6):437-448.

31. Bech P, Boyer P, Germain J-M, et al. HAM-D17 and HAM-D6 sensitivity to change in relation to desvenlafaxine dose and baseline depression severity in major depressive disorder. Pharmacopsychiatry 2010;43(7):271-276.

32. Hieronymus F, Emilsson JF, Nilsson S, Eriksson E. Consistent superiority of selective serotonin reuptake inhibitors over placebo in reducing depressed mood in patients with major depression. Mol Psychiatry 2016;21(4):523-530.

33. 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):623-627.

34. Bowden CL, Asnis GM, Ginsberg LD, et al. Safety and tolerability of lamotrigine for bipolar disorder. Drug Saf 2004;27(3):173-184.

35. Trivedi MH, South C, Jha MK, et al. A novel strategy to identify placebo responders: prediction index of clinical and biological markers in the EMBARC trial. Psychother Psychosom 2018;87(5):285-295.

36. Valerio MP, Szmulewicz AG, Martino DJ. A quantitative review on outcome-to-antidepressants in melancholic unipolar depression. Psychiatr Res 2018;265:100-110.

37. Rutherford BR, Mori S, Sneed JR, et al. Contribution of spontaneous improvement to placebo response in depression: a meta-analytic review. J Psychiatr Res 2012;46(6):697-702.

Table 3.1

HAMD items with baseline prevalence and change scores

Item	Baseline		Change scores, *M* (*SD*)		
	M (*SD*)	%$_{Zero}$	LTG	PBO	*z*
1. Depressed mood	2.9 (.50)	0.0	1.23 (1.2)	1.05 (1.2)	2.51*
2. Guilt	2.0 (.80)	5.7	0.94 (1.1)	0.78 (1.1)	2.21*
3. Suicide	1.0 (.80)	32.7	0.52 (.86)	0.47 (.87)	0.89
4. Insomnia, early	1.5 (.79)	18.3	0.49 (.98)	0.50 (1.0)	-0.23
5. Insomnia, middle	1.4 (.77)	17.5	0.43 (.97)	0.47 (.98)	-0.64
6. Insomnia, late	1.2 (.85)	27.6	0.45 (1.0)	0.46 (.97)	-0.02
7. Work and interests	2.9 (.58)	0.3	1.37 (1.3)	1.06 (1.2)	3.84***
8. Retardation	1.3 (.82)	18.8	0.66 (.92)	0.55 (.93)	1.89†
9. Agitation	1.1 (.84)	27.0	0.33 (.93)	0.37 (.87)	-0.61
10. Anxiety, psychic	2.3 (.71)	2.0	0.79 (1.2)	0.68 (1.1)	1.59
11. Anxiety, somatic	1.7 (.94)	12.9	0.67 (1.1)	0.61 (1.1)	0.67
12. Somatic, GI	0.8 (.76)	39.5	0.31 (.88)	0.36 (.82)	-0.55
13. Somatic, general	1.7 (.53)	4.0	0.76 (.91)	0.62 (.90)	2.47*
14. Genital	1.3 (.81)	23.0	0.45 (.92)	0.45 (.89)	-0.09
15. Hypochondriasis	0.7 (.76)	47.6	0.29 (.82)	0.19 (.80)	1.80†
16. Weight loss	0.3 (.59)	75.5	0.07 (.67)	0.13 (.71)	-1.29
17. Insight	0.2 (.41)	83.8	0.03 (.50)	0.04 (.40)	-0.05
18. Diurnal variation	0.9 (.83)	37.4	0.46 (.87)	0.38 (.93)	1.33
19. Depersonalization	0.4 (.72)	67.8	0.20 (.75)	0.24 (.71)	-0.74
20. Paranoid	0.4 (.59)	63.1	0.09 (.60)	0.17 (.63)	-1.82†
21. Obsessional/compulsive	0.4 (.59)	70.5	0.09 (.54)	0.11 (.54)	-0.34
22. Hypersomnia, early	0.5 (.80)	69.8	0.23 (.81)	0.18 (.83)	0.72
23. Hypersomnia, oversleep	0.4 (.72)	75.7	0.16 (.73)	0.14 (.77)	0.34
24. Hypersomnia, napping	0.6 (.81)	59.6	0.18 (.89)	0.20 (.91)	-0.03
25. Increased appetite	0.3 (.65)	77.9	0.14 (.67)	0.13 (.71)	0.07
26. Weight gain	0.3 (.60)	79.8	0.11 (.67)	0.10 (.64)	0.01
27. Psychic retardation	1.1 (.84)	26.1	0.55 (.90)	0.49 (.91)	0.86

Table 3.1 Continued

28. Motor retardation	1.0 (.84)	30.9	0.54 (.86)	0.46 (.91)	1.73†
29. Helplessness	1.4 (.88)	18.2	0.63 (1.1)	0.55 (1.1)	1.41
30. Hopelessness	1.6 (.90)	11.9	0.70 (1.1)	0.65 (1.1)	0.66
31. Worthlessness	1.7 (.86)	10.6	0.72 (1.0)	0.65 (1.0)	1.31

Note. Change scores were compared between treatment conditions with Wilcoxon rank-sum tests. A negative *z* score indicates the change score was larger in the placebo group. HAMD = Hamilton Depression Rating Scale; LTG = lamotrigine; PBO = placebo. *p < .05; **p < .01; ***p < .001; $^{\dagger}p$ < .10.

Table 3.2

MADRS items with baseline prevalence and change scores

Item	Baseline		Change scores, *M* (*SD*)		
	M (*SD*)	$\%_{Zero}$	LTG	PBO	*z*
1. Apparent sadness	3.5 (.90)	0.4	1.58 (1.6)	1.35 (1.6)	2.37*
2. Reported sadness	3.8 (.80)	0.3	1.68 (1.7)	1.44 (1.6)	2.46*
3. Inner tension	3.1 (.95)	1.8	1.04 (1.5)	0.87 (1.4)	1.84†
4. Reduced sleep	3.2 (1.5)	10.1	0.96 (1.8)	1.07 (1.8)	-0.81
5. Reduced appetite	1.6 (1.5)	40.8	0.63 (1.7)	0.71 (1.6)	-0.61
6. Concentration	3.3 (1.0)	2.2	1.31 (1.7)	1.18 (1.5)	1.29
7. Lassitude	3.4 (1.1)	1.9	1.53 (1.6)	1.24 (1.7)	2.72**
8. Inability to feel	3.4 (1.1)	1.8	1.64 (1.6)	1.31 (1.7)	3.16**
9. Pessimistic thoughts	2.7 (1.1)	3.0	1.22 (1.5)	0.99 (1.5)	2.44*
10. Suicidal thoughts	1.4 (1.1)	27.2	0.68 (1.2)	0.66 (1.3)	0.40

Note. Change scores were compared between treatment conditions with Wilcoxon rank-sum tests. A negative *z* score indicates the change score was larger in the placebo group. MADRS = Montgomery–Åsberg Depression Rating Scale; LTG = lamotrigine; PBO = placebo. *$p < .05$; **$p < .01$; ***$p < .001$; $^{\dagger}p < .10$.

Table 3.3

Depression scale change scores and response rates by treatment condition

Scale	Change scores, *M* (*SD*)		*F*	*d*	Response, %		Diff. (95% CI)
	LTG	PBO			LTG	PBO	
HAMD-6[a]	5.77 (4.97)	4.73 (4.89)	11.7***	.21	43.8	35.0	8.78 (2.95-14.6)
HAMD-17	9.80 (8.43)	8.80 (8.32)	3.82†	.12	44.9	36.5	8.38 (2.52-14.2)
HAMD-31	14.6 (12.3)	13.2 (12.7)	3.18†	.11	45.7	38.6	7.05 (1.15-12.9)
MADRS-6[b]	8.69 (7.80)	7.21 (7.76)	9.69**	.19	48.2	37.5	10.7 (4.80-16.6)
MADRS	12.3 (11.3)	10.8 (11.1)	4.42*	.13	48.1	39.2	8.89 (2.97-14.8)

Note. One-way analysis-of-variance tests were used to compare mean change scores between treatment groups (see *F* statistics) with Cohen's *d* as a measure of effect size. HAMD = Hamilton Depression Rating Scale; MADRS = Montgomery–Åsberg Depression Rating Scale; LTG = lamotrigine; PBO = placebo; Diff. = difference. *$p < .05$; **$p < .01$; ***$p < .001$; $^{\dagger}p < .10$.

[a]This subscale is composed of HAMD items 1 (depressed mood), 2 (guilt), 7 (work and interest), 8 (retardation), 10 (psychic anxiety) and 13 (general somatic/fatigue).

[b]This subscale is composed of MADRS items 1 (apparent sadness), 2 (reported sadness), 3 (inner tension), 7 (lassitude), 8 (inability to feel), and 9 (pessimistic thoughts).

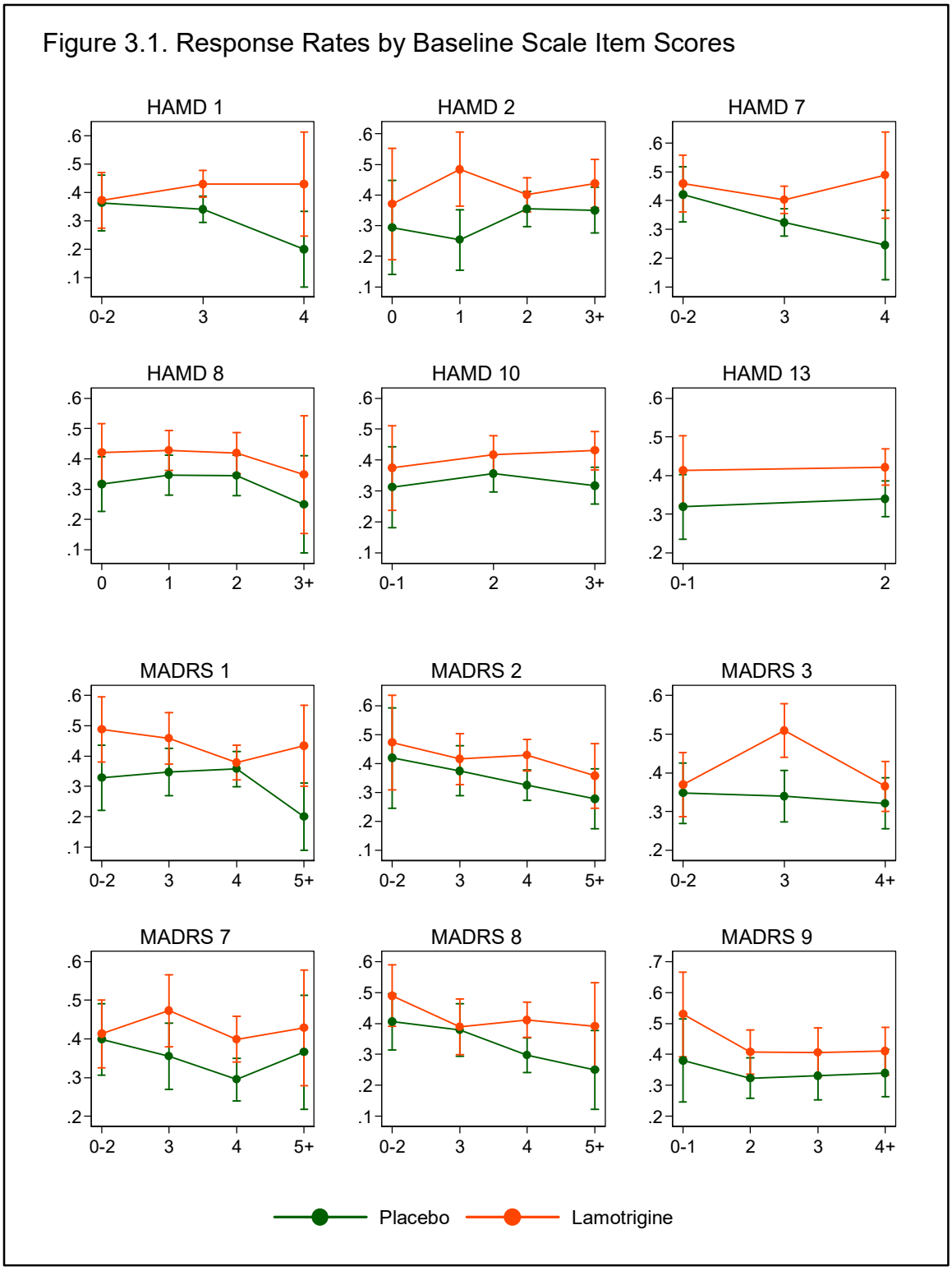

Figure 3.1. Each graph depicts the proportion of responders on the y-axis over baseline scores for the indicated depression scale item on the x-axis. HAMD = Hamilton Depression Rating Scale; MADRS = Montgomery–Åsberg Depression Rating Scale.

Table 3.4

Response rates stratified by subgroups derived from baseline scale item cut-off scores

Subgroup criteria	Met			Not met		
	LTG	PBO	Diff. (95% CI)	LTG	PBO	Diff. (95% CI)
1. H1 or H7 = 4	49.2	21.3	27.9 (12.4-43.3)	41.0	35.5	5.48 (-0.75-11.7)
2. M1 or M8 > 4	39.3	21.1	18.2 (4.34-32.1)	42.5	35.7	6.77 (0.45-13.1)
3. 1 or 2	41.9	19.5	22.4 (11.0-33.7)	42.0	37.7	4.24 (-2.41-10.9)
4. M1 > 4 or M8 > 3	41.0	28.3	12.7 (5.42-20.0)	43.5	40.5	2.95 (-6.35-12.2)

Note. Response rates are presented as percentages. H1 = HAMD item 1 (depressed mood); H7 = HAMD item 7 (work and interests); M1 = MADRS item 1 (apparent sadness); M8 = MADRS item 8 (inability to feel); LTG = lamotrigine; PBO = placebo; Diff. = difference; HAMD = Hamilton Depression Rating Scale; MADRS = Montgomery–Åsberg Depression Rating Scale.

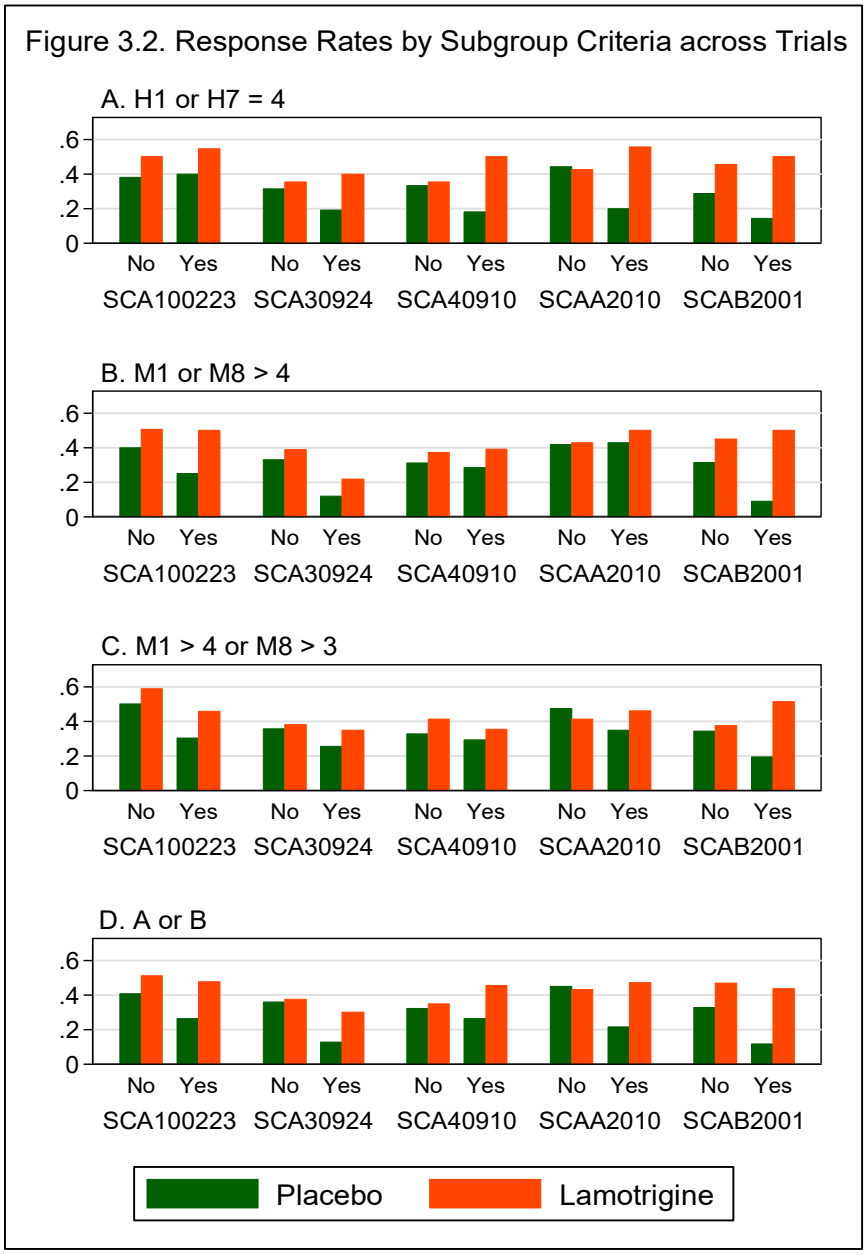

Figure 3.2. The graphs depict the proportion of responders stratified by trial number and subgroup status determined by baseline scores on depression scale items. Each graph (A through D) used different subgroup criteria. H1 = HAMD item 1 (depressed mood); H7 = HAMD item 7 (work and interests); M1 = MADRS item 1 (apparent sadness); M8 = MADRS item 8 (inability to feel); HAMD = Hamilton Depression Rating Scale; MADRS = Montgomery–Åsberg Depression Rating Scale.

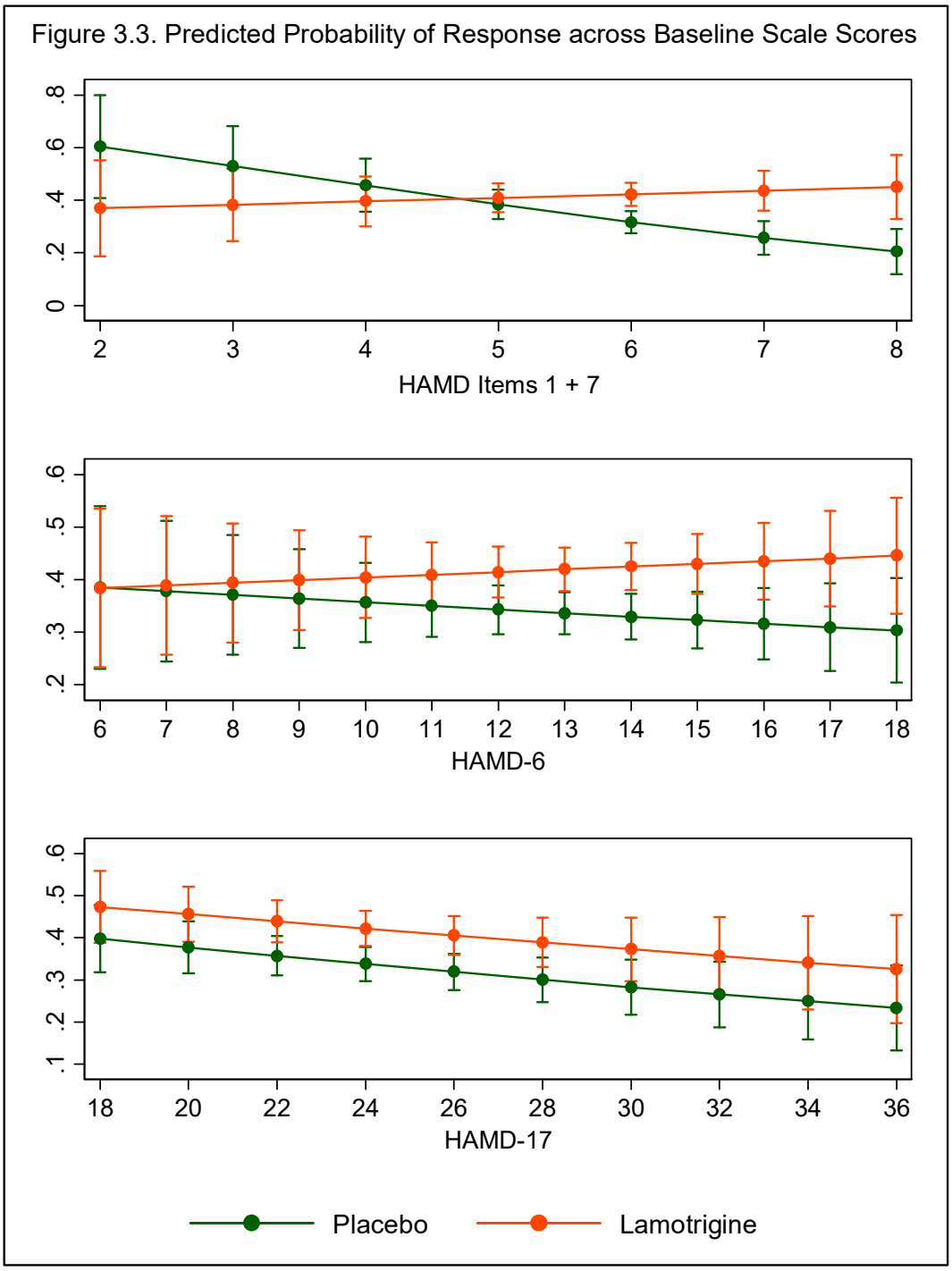

Figure 3.3. Each graph depicts the predicted probability of response across baseline sum scores on different measures derived from the Hamilton Depression Rating Scale (HAMD). These were created by estimating logistic regression models predicting response with an interaction between treatment and continuous scale scores, and main effects (not shown); the marginal effects of treatment condition across baseline scores were then calculated and graphed. HAMD-6 = 6-item subscale (items 1, 2, 7, 8, 10, and 13); HAMD-17 = full 17-item scale.

Transition Note 2

The second study analysis was not prespecified, meaning there was essentially no limit on the number of analyses that could have been conducted. At a certain point, however, these would have become redundant and unnecessarily complicated since firm conclusions cannot be drawn from exploratory analyses. Therefore, an attempt was made to keep the analysis relatively simple while addressing some of the discrepancies from the first study. To some extent, this was accomplished.

It was unclear from the first study why severity on the HAMD-17, but not the HAMD-31 or MADRS, was associated with larger response-rate differences. In the second study, two categorical and one continuous severity variables derived from HAMD-17 sum scores were not associated with increased response-rate differences, suggesting HAMD-17 scores actually do not have a straightforward effect on treatment response.

Patients scoring higher on mood and anhedonia items had larger treatment effects, with consistency across scales. This association appeared to be responsible for the results found with the scale-derived melancholia variable in the first study. Although some of the patients scoring higher on these items would have been melancholic, the effect was also clear when scores were treated dimensionally, arguing against a categorical melancholic-nonmelancholic distinction underlying the first study results.

By considering change scores at the item level, we replicated previous findings that core depressive symptoms (e.g., low mood, anhedonia, guilt, psychomotor retardation) were somewhat more responsive to lamotrigine treatment than total depression sum scores. Previously, we interpreted this as evidence that lamotrigine may preferentially benefit melancholic patients, for whom these symptoms are more prominent. However, the increased change-score differences appeared to be largely due to a statistical floor effect, arguing against the notion that lamotrigine preferentially targets these symptoms. Although this interpretation is contrary to the original hypothesis, it does reinforce the decision to use response rates as the primary outcome in the first study.

CHAPTER 4

Discussion and Conclusions

This thesis presented the results of two studies in which pooled data from five randomized placebo-controlled trials of lamotrigine for acute bipolar depression was analyzed with the goal of identifying treatment effect moderators. The main findings are summarized in the next sections, along with general limitations and recommendations for future research.

4.1 Melancholic depression as a predictor of lamotrigine response

In the first study, we were able to replicate the result from the pilot study (1) with the scale-derived melancholia variable, but it was not strong enough to produce a statistically significant interaction, and there was no apparent moderating effect when melancholic symptoms were assessed with another established measure. Without being able to confirm the validity of the scale-derived variable, the results cannot be used to support the hypothesis.

In the second study, only two of the items used to create the scale-derived melancholia variable were needed to produce the same effect. Furthermore, in line with Mitchell et al. (2), we also found larger treatment effects on scale items that assess core depressive symptoms. However, this appeared to be largely due to a floor effect, arguing against our initial assumption that lamotrigine may be more effective at treating these symptoms, and by extension, melancholic patients. Overall, the original hypothesis was not supported.

4.2 Depression severity as a predictor of lamotrigine response

The first study replicated the Geddes et al. (3) results in that patients scoring above the sample mean on the HAMD-17 had larger response-rate differences, although this was not the case when HAMD-31 and MADRS scores were used to split the sample. Furthermore, in the second study, HAMD-17 scores were not linearly related to the response-rate difference and we did not find a similar effect when HAMD-17 scores were dichotomized at higher levels. Taken together, the results do not support a straightforward relationship between baseline depression score severity and the lamotrigine-placebo response-rate difference.

4.3 Responsiveness of depression scale items to lamotrigine

In the second study, we found significant treatment effects on individual scale items assessing core depressive symptoms including mood, interest/anhedonia, guilt/pessimism, and energy/fatigue. This finding is consistent with the results reported by Mitchell et al. (2), who examined item factors in a smaller sample that included unipolar depressed patients. Accordingly, using subscales composed of these items to calculate change scores resulted in larger treatment effects compared to those derived from the full scales. This pattern has been

repeatedly demonstrated across decades of research on other depressed samples being treated with other medications (4-7). The main implication is that the original monotherapy trials may have found stronger efficacy signals had treatment effects been calculated with core subscales instead of the full scales.

4.4 Limitations

A major limitation with both studies was the post-hoc design, which poses an increased risk of type I and type II error (8). We attempted to mitigate this in the first study by pooling data and performing interaction tests (8), but it still would have been prudent to replicate positive results in a prospective trial with stratified sampling and multiplicity adjustments. We did not adjust for multiple comparisons in the first study because we did not conceptualize melancholic depression and severe depression as distinct subpopulations, but rather, competing and largely overlapping definitions of similar patient populations. In the second study, multiple comparisons were more of a concern, particularly regarding the item-level analysis, so we cannot rule out the possibility that some significant effects occurred by chance. Moreover, the subgroup analysis from the second study needs to be considered entirely hypothesis-generating.

As discussed in the first manuscript, designing a study around the concept of melancholia is problematic because there is no gold-standard diagnostic measure. We focused on the DSM criteria because these are the most common, and were accessible from the data, but there are other measures that are arguably superior. For example, the Newcastle index (9) includes similar melancholic symptoms but also considers premorbid personality, prior episodes, and preceding life events; the CORE scale (10) focuses on observable psychomotor disturbance and affective signs, and does not rely on the patient's subjective emotional descriptions or neurovegetative symptoms that could have another cause (e.g., medications). Accordingly, these scales may be more strongly associated with dexamethasone nonsuppression than the DSM criteria (11-13) and may have been more useful for establishing treatment effect heterogeneity.

Conceptual and methodological ambiguity further limited both studies in that severity and treatment response have multiple accepted definitions. Furthermore, there are over a dozen different depression scales, and each one is unique in some way (14, 15). It is likely some results were dependent upon methodological decisions. This was apparent in the first study, where the moderating effect of severity was in opposite directions depending on which scale was used to define severity. We also found discrepant results depending on how treatment effects were

calculated (e.g., larger effects with HAMD-6 change scores but not response rates) and the scale used (e.g., smaller and less robust results with HAMD-31 response rates, in the first study). The results could have been even more convoluted had we considered alternate measures of severity and/or treatment response that do not rely on depression scale scores (e.g., functional impairment, global improvement ratings), or alternate statistical techniques (e.g., mixed models for repeated measures).

Another significant limitation, often not discussed, is that most industry-sponsored trials employ numerous sampling criteria that exclude roughly 85% of depressed outpatients, and virtually all inpatients (16). Presumably, it is easier to detect treatment effects in homogenous populations with less within-group differences, but the results still cannot be generalized to many patients encountered in clinical practice.

4.5 Recommendations for future research

Future research on melancholic depression as a moderator of treatment response could benefit from using one of several biomarkers (e.g., dexamethasone nonsuppression) to strengthen clinical diagnoses. These are not universally accepted as valid diagnostic tests because they do not map perfectly onto to the clinical criteria (17) but could potentially help isolate more homogenous subgroups than the clinical criteria alone. An alternative strategy would be to isolate prototypical melancholic patients with obvious symptoms and a positive diagnosis from multiple measures, and a clearly nonmelancholic comparison group. This would avoid having to split samples with imperfect clinical diagnostic criteria, potentially making subgroup differences easier to detect, although some patients would remain unclassified.

It would be interesting for future research to consider baseline scores on the mood and anhedonia items as predictors of treatment response, to both lamotrigine and other medications. In this study we found some consistency across scales (which was not apparent using total depression scores) and trials. Several pooled analyses have already attempted to determine if baseline depression sum scores moderate response to second-generation antidepressants in unipolar depression, with negative results (18-20). It is worth exploring whether the findings reported here generalize to other medications. If so, it may be possible to use these items to refine trial inclusion criteria such that treatment effects are easier to detect.

Finally, future research should be aware of, and attempt to account for, methodological considerations that could influence trial results and their interpretation. Using outcome measures

that are standardized by baseline scores would minimize the problem of floor effects (21). Alternatively, the HAMD-6 may be less susceptible to floor effects compared to the full HAMD-17 (5). This would be particularly useful for research examining treatment effects across subgroups with unequal depression scores at baseline.

4.6 Conclusions

It remains unclear if patients with bipolar depression and melancholic features benefit more from lamotrigine compared to nonmelancholic patients. There continues to be evidence that it is possible to identify subgroups of treatment-responsive patients in the pooled monotherapy trial data, but conceptual and methodological limitations preclude drawing firm conclusions about the nature of these relationships. The depression scales used in the original trials may have underestimated treatment effects relative to core-symptom subscales. Further research is needed to better understand lamotrigine's antidepressant effect and its precise role in the treatment of bipolar disorder.

4.7 References

1. Peters EM, Bowen R, Balbuena L. Melancholic symptoms in bipolar II depression and responsiveness to lamotrigine in an exploratory pilot study. J Clin Psychopharmacol 2018;38(5):509-512.

2. Mitchell PB, Hadzi-Pavlovic D, Evoniuk G, et al. A factor analytic study in bipolar depression, and response to lamotrigine. CNS Spectr 2013;18(4):214-224.

3. Geddes JR, Calabrese JR, Goodwin GM. Lamotrigine for treatment of bipolar depression: independent meta-analysis and meta-regression of individual patient data from five randomised trials. Brit J Psychiatry 2009;194(1):4-9.

4. Carrozzino D, Patierno C, Fava GA, Guidi J. The Hamilton rating scales for depression: a critical review of clinimetric properties of different versions. Psychother Psychosom 2020;89(3):133-150.

5. Hieronymus F, Lisinski A, Nilsson S, Eriksson E. Influence of baseline severity on the effects of SSRIs in depression: an item-based, patient-level post-hoc analysis. Lancet Psychiatry 2019;6(9):745-752.

6. Lisinski A, Hieronymus F, Näslund J, et al. Item-based analysis of the effects of duloxetine in depression: a patient-level post hoc study. Neuropsychopharmacology 2020;45(3):553-560.

7. Timmerby N, Andersen JH, Søndergaard S, et al. A systematic review of the clinimetric properties of the 6-item version of the Hamilton Depression Rating Scale (HAM-D6). Psychother Psychosom 2017;86(3):141-149.

8. Brookes ST, Whitely E, Egger M, et al. Subgroup analyses in randomized trials: risks of subgroup-specific analyses: power and sample size for the interaction test. J Clin Epidemiol 2004;57(3):229-236.

9. Carney MW, Roth M, Garside RF. The diagnosis of depressive syndromes and the prediction of ECT response. Brit J Psychiatry 1965;111(477):659-674.

10. Parker G, McCraw S. The properties and utility of the CORE measure of melancholia. J Affect Disord 2017;207:128-135.

11. Davidson J, Lipper S, Zung WW, et al. Validation of four definitions of melancholia by the dexamethasone suppression test. Am J Psychiatry 1986;141(10):1220-1223.

12. Mitchell PH. Validity of the CORE: I. A neuroendocrinological strategy. Melancholia: A disorder of movement and mood. Cambridge University Press, Cambridge. 1996;29:138-148.

13. Philipp M, Maier W, Holsboer F. Psychopathological correlates of plasma cortisol after dexamethasone suppression: a polydiagnostic approach. Psychoneuroendocrinology 1986;11(4):499-507.

14. Fried EI. The 52 symptoms of major depression: lack of content overlap among seven common depression scales. J Affect Disord 2017;208:191-197.

15. Fried EI, Nesse RM. Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. BMC Med 2015;13(1):1-11.

16. Zimmerman M, Balling C, Chelminski I, Dalrymple K. Have treatment studies of depression become even less generalizable? Applying the inclusion and exclusion criteria in placebo-controlled antidepressant efficacy trials published over 20 years to a clinical sample. Psychother Psychosom 2019;88(3):165-170.

17. Fink M. Should the dexamethasone suppression test be resurrected?. Acta Psychiatr Scand 2005;112(4):245-249.

18. Furukawa TA, Maruo K, Noma H, et al. Initial severity of major depression and efficacy of new generation antidepressants: individual participant data meta-analysis. Acta Psychiatr Scand 2018;137(6):450-458.

19. Gibbons RD, Hur K, Brown CH, et al. 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):572-579.

20. Rabinowitz J, Werbeloff N, Mandel FS, et al. Initial depression severity and response to antidepressants v. placebo: patient-level data analysis from 34 randomised controlled trials. Brit J Psychiatry 2016;209(5):427-428.

21. Vöhringer PA, Ghaemi SN. Solving the antidepressant efficacy question: effect sizes in major depressive disorder. Clin Ther 2011;33(12):B49-61.