Published in final edited form as:

Lancet Psychiatry. 2016 May ; 3(5): 436–442. doi:10.1016/S2215-0366(16)00006-7.

Can research in severe and treatment-resistant mood disorders be conducted safely? Outcome data from 12 years of clinical trials at the US National Institute of Mental Health

Allison C Nugent, PhD¹, Nicolas D Iadarola, BA², Frank G Miller, PhD¹, David A Luckenbaugh, MA¹, and Carlos A Zarate Jr, MD¹

¹Experimental Therapeutics and Pathophysiology Branch, NIMH, NIH, Bethesda, USA

²The Department of Bioethics, Clinical Center, NIH, Bethesda, USA

Abstract

Background—This study sought to assess the cumulative effects of research on individuals with mood disorders and to provide data to potentially address ethical concerns regarding research in this population. Empirical data were collected from 18 clinical trials conducted from December 2001 to January 2014 at the US National Institute of Mental Health (NIMH) exploring drug tapers, placebo arms, and the cumulative effects of research.

Methods—Subjects included 504 individuals with either major depressive disorder or bipolar disorder who gave written informed consent to a treatment trial in either our inpatient or outpatient behavioral health research clinics at the National Institutes of Health. Over 70% of these subjects were enrolled in trials requiring treatment resistance, and mean severity at screening was in the moderate to severe range. Data were reviewed to assess the cumulative effects of research on our patient population, including the effects of drug taper/washout and placebo administration on mood state.

This manuscript version is made available under the CC BY-NC-ND 4.0 license.

Corresponding Author: Allison C Nugent, PhD, Experimental Therapeutics and Pathophysiology Branch, National Institute of Mental Health, National Institutes of Health, 10 Center Drive, MSC 1282, Bethesda, MD 20892-1282, Phone: 301-451-8863, Fax: 301-480-3610, nugenta@mail.nih.gov.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Declaration of Interests: A patent application for the use of ketamine in depression has been submitted listing Dr. Zarate among the inventors; he has assigned his rights on the patent to the U.S. government, but will share a percentage of any royalties that may be received by the government. All other authors have no conflict of interest to disclose, financial or otherwise.

Contributors

AN was lead investigator for the study; oversaw and assisted with data compilation; performed data analysis; interpreted the results; prepared the manuscript; revised the manuscript; approved the final version of the manuscript. NI assisted with data collection; revised the manuscript; approved the final version of the manuscript. FM provided guidance on the ethical aspects of study design and interpretation; revised the manuscript; approved the final version of the manuscript. DL assisted with statistical analyses; revised the manuscript; approved the final version of the manuscript. CZ oversaw collection of the source data; provided scientific guidance for the project; interpreted the results; revised the manuscript; approved the final version of the manuscript.

Trial Registration: NCT00024635

Findings—Full participation in research—including drug tapers, drug-free periods, and placebo-controlled trials—carried a relatively low risk of symptom exacerbation. Subjects undergoing drug taper showed a mean increase in symptoms of 4.2%, while subjects receiving placebo improved on average. Research participation at the NIMH, in general, was not detrimental to health and safety, and conferred benefit in many cases. This was true not only in our entire research population, but also in treatment-resistant subgroups as well as subgroups with a history of suicidality. Six serious adverse events were reported; this included one suicide attempt that occurred during the standard treatment phase and not an experimental trial.

Interpretation—These data provide evidence to guide ethical analysis of issues in psychiatric research and support continued scientific investigation.

Keywords

depression; bipolar disorder; placebo; washout; clinical trial; ethics

Introduction

Mood disorders, including major depressive disorder and bipolar disorder, are highly prevalent, frequently disabling, and sometimes fatal. Currently approved pharmacologic treatments for major depressive disorder are characterized by delayed onset of efficacy, with maximal effects not seen for several weeks or longer¹. In addition, first-line agents generally result in remission in less than one third of all cases, leading to high prevalence of treatment-resistant depression and residual symptoms^{2, 3}. Furthermore, meta-analyses show that the most widely prescribed FDA-approved antidepressants outperform placebo by an approximate margin of only two points on the 17-item Hamilton Depression Rating Scale^{4, 5}.

Given the relatively modest benefit of existing antidepressants, new treatment targets are needed. Relatively few options outside of research exist for severely ill and/or treatment-resistant subjects who might have exhausted all approved treatment options. Placebo-controlled, medication-free trials have long been considered a key research component in the study of mood disorders and relevant treatment mechanisms. However, ethical concerns have been raised by such research, leading to an ongoing debate as to whether placebo controls are ethically acceptable. In particular, placebo controls are most frequently criticized for denying patients proven effective treatments⁶, with one author going so far as to state that research should not be an “invitation to practice substandard medicine”⁷. Other authors have suggested that severely ill patients should never be included in research^{8, 9}. Nevertheless, data regarding how patients fare during drug-free research periods has been missing from the debate over the ethics of study design in mood disorders.

This study presents data collected from patients with either major depressive disorder or bipolar disorder enrolled in a treatment trial in either an inpatient or outpatient behavioural health research clinic conducted at the US National Institute of Mental Health (NIMH) over a 12-year period. This study examined whether mechanistic trials of novel treatments can be performed safely even in a severely ill and treatment-resistant population. We tested several hypotheses, primarily based on our clinical observations. First, that tapering ineffective

psychiatric drugs would not, on average, seriously exacerbate symptoms in our patient population. Second, that placebo would, on average, result in symptomatic improvement, although not necessarily to the same degree as active treatments. Third, that over the course of research participation at our institution, subjects' mood states would not, in general, decline.

Method

Subjects

Five hundred and four subjects were enrolled in one or more of 18 clinical trials conducted from December 2001 to January 2014 in NIMH inpatient and outpatient psychiatric research clinics. All studies were approved by an NIMH or NIH intramural IRB. All patients gave written informed consent to a screening protocol at the time of admission, and also when they enrolled in their respective clinical trial. Additional details can be found in the appendix.

Data Collection

Research databases were examined to obtain demographic data, clinical characteristics, and mood rating scores. Diagnoses were taken from the SCID (structured clinical interview for DSM-IV) results and clinician judgment. The primary outcome measure for most studies was Montgomery-Asberg Depression Rating Scale (MADRS) score¹⁰, although some trials used the Hamilton Depression Rating Scale¹¹ or the Young Mania Rating Scale¹². Two subgroups were examined: subjects enrolled in trials explicitly requiring treatment resistance (*TR trials*) and subjects with a current or past history of suicidal ideation/behaviour (*Suicidality+* vs. *Suicidality*). For more information, see the appendix.

Research participation was divided into five discrete time periods: 1) screening, 2) drug taper and washout, 3) prospective drug trial (to establish if unmedicated subjects were treatment-responsive), 4) placebo lead-in, and 5) main clinical trial. Although some drug trials were conducted on an outpatient basis, most drug tapers were performed on an inpatient research unit with high staffing ratios that allowed for careful monitoring of subjects to ensure participant safety. Following research participation, all participants were given the opportunity to receive up to three months of standard treatment, frequently while on the inpatient unit.

Statistical analysis

For all statistical tests, significance was set at $p < 0.05$, using two-tailed tests. Because we conceptualized this as a study of a single cohort designed to assess the safety of research participation in our psychiatric units, we pooled all data and used standard statistical tools rather than meta-analytic methods. To compensate for different outcome measures across studies, percent change from screening was used as the primary outcome measure for statistical analysis of change in mood over study periods. For each time period considered, we assessed mood change over the interval for all subjects and within subgroups using one-sample *t*-tests to determine whether significant changes in mood existed over the interval. We also used independent samples *t*-tests to examine mood changes in *Suicidality+* subjects

compared to *Suicidality*— subjects. We assessed differences in mood changes within each interval across diagnostic groups using one-way ANOVA, and calculated Pearson correlations between mood change and severity at screening. In the main text, we report results only from those diagnostic subgroups where ANOVA results were significant; however, all diagnostic and phenotypic subgroup results appear in the Supplemental Materials. Specific notes regarding the various analyses are given below and in the Supplemental Methods.

Effects of drug taper/washout

All studies of subjects with major depressive disorder required that they be unmedicated preceding initiation of the study drug. In some studies of patients with bipolar disorder (both bipolar disorder-I and bipolar disorder-II), subjects were maintained or started on mood stabilizer monotherapy (lithium or valproate). We examined all subjects who underwent taper of at least one agent but also report on the subgroup that tapered to no medications.

Effect of placebo on mood state

Our clinical studies included open-label, parallel, and crossover trials. In open-label trials, there is no placebo condition, and all participants and researchers know that participants are receiving active drug. In parallel trials, there are two parallel arms; subjects are randomized to receive either placebo or active drug for the duration of the trial. Both participant and investigator are “blind” to the treatment condition for the duration of the study (i.e. a double-blind). For parallel trials, we assessed the entire interval from study baseline to the end of the trial. In addition to examining mood change in the placebo group over the course of the trial, we reported the difference between those randomized to placebo and those randomized to the active compound using independent samples *t*-tests.

In a standard efficacy analysis, statistical methods such as “last observation carried forward” or other imputation methods are used to predict the difference in mood scores between placebo and drug if all patients remained until the end of the trial. This, in essence, compensates for the fact that more patients drop out of placebo arms due to inadequate treatment. In the present study, because we were concerned with actual mood scores and not drug efficacy, we reported results for those subjects who completed the trials using no imputation methods, and assessed mood changes for those who withdrew separately (as described below).

In crossover trials, subjects are initially randomized to receive either drug or placebo under double-blind conditions. Following a short treatment period (and possibly a washout period), subjects “cross over” and receive the other intervention. Thus, all patients receive both active drug and placebo, in randomized order. Our crossover trials were proof-of-concept trials designed to test agents with a rapid onset of action (~ one day) but relatively short-lived effects (< two weeks). We assessed percent change in mood scores in participants receiving placebo in the first arm only, as assessment of mood change in the second arm was problematic due to the potential presence of carryover effects.

Assessment of withdrawals

For subjects who withdrew or were withdrawn from research, we recorded the reason for withdrawal and phase of withdrawal; details appear in the appendix. If a mood score was available at or near the time of withdrawal, we assessed percent change from screening to withdrawal. We also used independent samples *t*-tests to assess whether those who withdrew due to worsening showed higher MADRS scores at screening. Protocols varied in their inclusion of specific criteria imposing mandatory withdrawal from the protocol due to worsening. The most commonly used criterion was greater than 30% worsening from study baseline, although this pertained only to the treatment portion of the trial. In practice, the decision to withdraw was usually made collaboratively between the participant, the primary clinician, and a patient representative.

General assessment of the cumulative effects of research

The cumulative effects of research were examined by calculating percent change between screening and the end of the last trial a subject participated in during the first encounter with our laboratory. Information regarding serious adverse events was extracted from IRB records.

Role of the funding source

This work was funded by the Intramural Research Program at the NIMH, NIH (ZIA MH002927), by a NARSAD Independent Investigator Award to CAZ, and by a Brain and Behavior Mood Disorders Research Award to CAZ. The funders had no further role in study design; in the collection, analysis, or interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

Results

Overall, the 504 subjects possessed many characteristics of a seriously ill, complex patient group, with high rates of substance abuse, comorbidities, treatment resistance, and suicidality (Table 1). Further details for some subgroup analyses described below are given in the Supplemental Results and in Supplemental Table 1.

Effects of drug taper/washout

A total of 223 participants had at least one medication tapered and withdrawn. Of these, 22 (9.9%) withdrew due to symptomatic worsening, while 13 (5.8%) withdrew due to symptomatic improvement. For subjects completing the taper, the mean increase in symptom severity was 4.2%; this value did not differ significantly from zero as assessed by one sample *t*-test (Figure 1A, Supplemental Table 1). Similarly modest changes were observed in the subgroup tapering to no medications (percent change=5.1%, $t_{56}=2.12$, $p=0.039$), participants enrolled in *TR trials*, and those with *Suicidality+* (see Supplemental Table 1). Diagnostic subgroup affected the results ($F_{2,94}=3.88$, $p=0.024$). Participants with bipolar disorder I showed an increase in symptom score (12.7%, $t_{21}=2.77$, $p=0.01$) despite the fact that most of these subjects (17/22) were maintained on a mood stabilizer. Mood score at screening correlated negatively with percent change post-taper ($N=115$, $r=-0.61$, $p<0.0001$),

and the least ill subjects at screening tended to show the greatest increase in symptoms. Distribution of percent change across all subjects is shown in Figure 2.

Effects of placebo on mood state

Of the 208 subjects who began a placebo lead-in, seven (3.7%) withdrew due to symptomatic worsening and 11 withdrew due to improvement (5.3%). The change in mood score over the placebo lead-in period was minimal for all groups (Figure 1A and Supplemental Table 1; see also Supplemental Figure 1A for a histogram of outcomes). Change in mood differed significantly between *Suicidality+* and *Suicidality-* participants ($t_{93}=2.65$, $p=0.01$).

Of the 213 subjects entering a parallel trial, 90 received placebo (Figure 1A, Supplemental Table 1). Forty-one (19.2%) subjects withdrew due to symptomatic worsening; 36 were enrolled in a *TR trial*. In active arms, 13.8% withdrew due to worsening, while 26.7% of those beginning a placebo arm withdrew due to worsening. Of those completing the trials, percent improvement did not differ significantly between those randomized to placebo or active treatment ($t_{148}=0.343$); note that an efficacy analysis would have compensated for withdrawals, resulting in greater difference between placebo and active arms.

A total of 113 subjects entered a crossover trial. Ten subjects (eight of whom were enrolled in *TR trials*) withdrew due to symptomatic worsening; notably, all withdrawals occurred during an active arm. All subjects and all subgroups showed no significant change in symptoms while receiving placebo in the first arm of the crossover (Figure 1A, Supplemental Table 1). Histograms of individual outcomes for patients receiving placebo in either parallel or cross-over trials are shown in Supplemental Figures 1B and 1C.

Assessment of withdrawals

In total, 216 withdrawals occurred in all enrolled subjects, reflecting a 42.9% withdrawal rate (Table 2). Of the 92 withdrawals due to worsening, 75 occurred in subjects enrolled in *TR trials*. Six of the withdrawals due to worsening were associated with suicidal ideation; all occurred during the drug trial and all participants elected to receive standard care. Subjects with more severe symptoms at screening were more likely to withdraw due to worsening compared to all other outcomes ($t_{269}=3.296$, $p=0.001$). Subjects who withdrew due to symptomatic worsening showed a mean increase in symptoms of 14.75%, while subjects who withdrew for all other reasons exhibited a mean decrease in symptoms of -18.48% (Figure 1B, Supplemental Table 1, and Supplemental Figure 2 for a histogram of individual outcomes). Similar patterns were observed for those participants enrolled in protocols that required treatment resistance. *Suicidality+* participants showed less decline at withdrawal due to worsening of symptoms than *Suicidality-* participants ($t_{25}=2.03$, $p=0.053$), and *Suicidality+* participants showed significantly greater improvement compared to *Suicidality-* participants for all other withdrawal reasons ($t_{24}=3.33$, $p=0.003$).

General assessment of the cumulative effects of research

Over the interval from screening to the end of research participation (which may have included multiple trials) subjects exhibited a mean -25.38% decrease in mood symptoms

(Figure 1B, Supplemental Table 1; for a histogram of individual outcomes see Supplemental Figure 1D). Following research participation, 87% of all subjects elected to receive standard care. Subjects received standard treatment for a mean of 93 days, although many subjects were treated for far longer in the outpatient setting.

Six serious adverse events were reported; this included one suicide attempt that occurred during the standard treatment phase and not an experimental trial. These are described in detail in the appendix.

Discussion

This study demonstrated that participating in psychiatric research did not, on average, result in a dangerous or damaging decline in mood, even in treatment-resistant and severely ill populations. We believe that our large cohort of 504 well-characterized subjects participating in trials at a single institution is ideal for demonstrating that clinical research can be safely conducted even in a severely ill group of participants. These data represent the largest sample ever reported regarding drug tapers.

In the bioethics literature, drug tapers and placebo controls are questioned on the grounds that they can result in dangerous exacerbations of symptoms⁶ that might ultimately result in irreversible neurological and/or interpersonal damage¹⁵. In our dataset, subjects who completed the taper showed only a modest worsening of mood symptoms, with the scores following a near-normal distribution. We suspect that change in mood partly represents natural fluctuations in mood state combined with response to the supportive therapeutic milieu of the research unit, as evidenced by the fact that the most severe subjects tended to improve over the course of the taper. In addition, the distribution of our results might suggest that for some patients, the medications that were tapered were actually worsening symptoms; thus, the drug taper was potentially beneficial. Notably, mean declines in well-being remained modest even in subjects who were enrolled in protocols that required treatment resistance and those with a history of suicidality. Nevertheless, any one subject undergoing drug taper does face a finite risk of symptomatic worsening; for example, we observed a maximum increase in mood symptoms of 66.7% during the drug taper. Distributions of individual outcomes were generally close to normal, with many subjects showing symptomatic improvement over the course of research intervals. While the effects on individual participants should not be ignored, they should be interpreted in the context of these disorders that naturally show waxing and waning of symptoms¹⁶; even subjects receiving maintenance therapy for mood disorders relapse at rates between 16 and 24% over a one-year period^{17–19}. Nevertheless, it is not unusual for a research participant's course of illness to decline, and a determination must be made as to whether it is ethical for a patient in decline to remain in a study. Our findings do not provide guidance in this matter, particularly because such a decision is complex and multifactorial. However, our results do show that since most subjects do not precipitously decline, a threshold can be identified that both protects subjects and allows research to continue.

Our findings are consistent with those from another study that examined the taper of ineffective drugs²⁰, in which participants with major depressive disorder generally remained

stable or improved over a two-week washout period. In the present study, we observed that participants with bipolar disorder-I showed the greatest worsening of symptoms (a 12% increase) over the course of the taper, despite the fact that most were maintained on mood stabilizers. Additional attention to participant status during the taper may be warranted for this population.

The use of placebo in clinical trials has been examined both in terms of ethical and methodological concerns. The primary ethical concern is that the use of placebo denies participants access to standard treatment for the duration of the placebo administration, potentially resulting in harm due to symptomatic worsening^{21–23}, the long term sequelae of which are unknown²⁴. On average, this study found no evidence of decline over a completed placebo period (either in a lead-in, parallel, or crossover trial arm). We thus conclude that the primary burden to our subjects was the delay of potentially more effective treatment, a burden that is equally shared by subjects in the experimental treatment arm if the drug is found to be ineffective. This echoes the findings of Kahn and colleagues²⁵, Laughren²⁶, and Storosum²⁷, which showed that not only do subjects with major depressive disorder or bipolar mania tend to show symptomatic improvement on validated symptom rating scales during placebo administration, but suicide attempt rates are nominally (although non-significantly) lower.

Interestingly, we observed reasonably high rates of withdrawal due to worsening across our studies (18.5%), as well as within individual research time periods (9.9% in drug taper, 26.7% in placebo arm of a parallel trial). Because such a high rate of withdrawal is clearly in the patient's interest rather than in the researcher's interest, we interpret this as evidence that our subjects felt free to withdraw at any time, and that we successfully identified and withdrew participants who were potentially suffering as a result of the research. Nevertheless, withdrawal due to worsening (as we define, any withdrawal due to perceived or actual worsening of symptoms, or intolerability of symptoms) was associated with a 14.8% increase in symptoms, which would represent an increase of 4-6 points on the MADRS (for a subject with an average MADRS score at screening). While this is enough to be clinically significant, in most cases it would not represent a dangerous decompensation. Notably, of those withdrawing due to worsening, 92.5% elected to receive standard treatment from our clinic.

Surprisingly, fewer participants withdrew from crossover than from parallel trials, potentially because participants find them more acceptable, or perhaps simply because they were shorter. Because treatment effects were generally transient, gaining relief from depression only to have symptoms return several days later might have been more disturbing to participants than stable symptoms, potentially explaining why all withdrawals due to worsening occurred during the active rather than placebo phases. Alternately, side effects might have been such that subjects knew they had received a study drug and saw no reason to complete the trial. An additional unexpected result was that our *Suicidality+* participants actually appeared to fare better during placebo lead-in, and showed less decline in mood when withdrawing due to symptomatic worsening. We speculate that suicidality might be considered a symptom of the disorder and not necessarily an indicator of poor prognosis.

While the full significance of this finding is unclear, it supports the contention that the *Suicidality+* and *Suicidality-* populations were exposed to equal risk.

One interesting and important fact is that, in our research trials, subjects improved over the course of research from screening to end of participation (either study completion or withdrawal) by -25.4% on average. Even participants enrolled in *TR trials* or *Suicidality+* subjects improved, although to a lesser degree than all subjects combined. This leads us to conclude that even though participants had forgone standard treatment during the time they were enrolled in our trials, they generally did receive some benefit from participating in these research studies. This is particularly noteworthy for patients who had previously failed to respond to a significant number of FDA-approved treatments. We would speculate that this is in part due to the therapeutic milieu of the inpatient research unit, although some of our studies were conducted on an outpatient basis.

The study had several limitations. First, our dataset incorporated many heterogeneous studies (see Supplemental Table 2), and was thus incomplete. Second, we used only clinician-rather than self-report scales. While it has been reported that clinician- and self-rated scales can disagree with regard to severity of illness^{13, 14}, we used only clinician-rated scales in order to eliminate any ancillary effects of patient personality on ratings. Our clinicians are scrupulously trained and tested in inter-rater reliability; thus, we felt that using clinician-related scales excluded as much variance as possible from repeated measurements. Third, our clinics are located in a large, dedicated research centre, with a highly staffed research unit capable of admitting research subjects for long periods of time, which enables gradual medication tapers under close observation; thus, the results are not necessarily generalizable. We cannot establish that outcomes would be similar under less advantageous circumstances. Nevertheless, while our results do not imply that drug tapers and placebo-controlled research is *always* safe in severe mood disorders, they do suggest that it is *possible* to safely conduct research in these participants. Finally, one somewhat philosophical limitation is that empirical data can never fully answer a normative ethical question. Therefore, we intend for this manuscript to instead provide useful data to support ethical discourse. While we can assert that our findings of only minimal average declines in mood during drug tapers support the use of such research procedures, it is an open question as to how much deterioration in mood, or what proportion of subjects experiencing a “significant” decline in mood, would be required to render a study ethically unacceptable. Thus, all our findings are still subject to interpretation by ethicists, who might not all agree with our conclusions.

Taken together, our data support the conclusion that it is possible to conduct research safely in subjects with severe mood disorders. While we do not wish to deny the suffering of individual patients who may experience worsening symptoms, on average, exacerbation of symptoms was minimal, indicating that the knowledge gained by studying these disorders may outweigh the risks of drug withdrawals and placebo administration, although only a full normative analysis can determine the ethical acceptability of such research. We believe that these data will provide a useful adjunct to a full analysis of the ethical issues involved in clinical research in mood disorders.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank the 7SE research unit and staff for their support, as well as the NIH Medical Records Department. The authors also thank the patients who took part in these studies and their families. Ms. Ioline Henter (NIMH) provided invaluable editorial assistance.

Role of Funding Source: This work was funded by the Intramural Research Program at the National Institute of Mental Health, National Institutes of Health (IRP-NIMH-NIH; ZIA MH002927), by a NARSAD Independent Investigator Award to CAZ, and by a Brain and Behavior Mood Disorders Research Award to CAZ. These funding sources had no further role in study design; in the collection, analysis, or interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

References

1. Gartlehner, G.; Hansen, RA.; Morgan, LC., et al. Second-Generation Antidepressants in the Pharmacologic Treatment of Adult Depression: An Update of the 2007 Comparative Effectiveness Review. Rockville (MD): 2011.
2. Paykel ES. Partial remission, residual symptoms, and relapse in depression. *Dialogues Clin Neurosci.* 2008; 10(4):431–7. [PubMed: 19170400]
3. Pigott HE, Leventhal AM, Alter GS, Boren JJ. Efficacy and effectiveness of antidepressants: current status of research. *Psychother Psychosom.* 2010; 79(5):267–79. [PubMed: 20616621]
4. Kirsch I, Moore TJ, Scoboria A, Nicholls SS. The emperor's new drugs: an analysis of antidepressant medication data submitted to the U.S. Food and Drug Administration. *Prevention and Treatment.* 2002; 5(1)
5. Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med.* 2008; 5(2):e45. [PubMed: 18303940]
6. Lehrman NS, Sharav VH. Ethical problems in psychiatric research. *J Ment Health Adm.* 1997; 24(2):227–50. [PubMed: 9110525]
7. Glass KC. Ethical obligations and the use of placebo controls. *Can J Psychiatry.* 2008; 53(7):428–9. [PubMed: 18674397]
8. Baldwin D, Broich K, Fritze J, Kasper S, Westenberg H, Moller HJ. Placebo-controlled studies in depression: necessary, ethical and feasible. *Eur Arch Psychiatry Clin Neurosci.* 2003; 253(1):22–8. [PubMed: 12664309]
9. DuVal G. Ethics in psychiatric research: study design issues. *Can J Psychiatry.* 2004; 49(1):55–9. [PubMed: 14763679]
10. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry.* 1979; 134:382–9. [PubMed: 444788]
11. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry.* 1960; 23:56–62. [PubMed: 14399272]
12. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry.* 1978; 133:429–35. [PubMed: 728692]
13. Enns MW, Larsen DK, Cox BJ. Discrepancies between self and observer ratings of depression – The relationship to demographic, clinical and personality variables. *J Affect Disord.* 2000; 60(1): 33–41. [PubMed: 10940445]
14. Schneibel R, Brakemeier EL, Wilbertz G, Dykieriek P, Zobel I, Schramm E. Sensitivity to detect change and the correlation of clinical factors with the Hamilton Depression Rating Scale and the Beck Depression Inventory in depressed inpatients. *Psychiatry Res.* 2012; 198(1):62–7. [PubMed: 22445070]
15. Glass KC. Ethical obligations and the use of placebo controls. *Can J Psychiatry.* 2008; 53(7):428–9. [PubMed: 18674397]

16. Vergunst FK, Fekadu A, Wooderson SC, et al. Longitudinal course of symptom severity and fluctuation in patients with treatment-resistant unipolar and bipolar depression. *Psychiatry Res.* 2013; 207(3):143–9. [PubMed: 23601791]
17. Geddes JR, Carney SM, Davies C, et al. Relapse prevention with antidepressant drug treatment in depressive disorders: a systematic review. *Lancet.* 2003; 361(9358):653–61. [PubMed: 12606176]
18. Seemuller F, Meier S, Obermeier M, et al. Three-year long-term outcome of 458 naturalistically treated inpatients with major depressive episode: severe relapse rates and risk factors. *Eur Arch Psychiatry Clin Neurosci.* 2014; 264(7):567–75. [PubMed: 24590257]
19. Vazquez GH, Holtzman J, Lolic M, Ketter T, Baldessarini RJ. Rates of recurrence in bipolar disorder: systematic analysis and comparison between long-term prospective, naturalistic studies vs. randomized controlled trials (RCTs). *Bipolar Disord.* 2015; 17:91–2.
20. Grunebaum MF, Oquendo MA, Burke AK, et al. Clinical impact of a 2-week psychotropic medication washout in unipolar depressed inpatients. *J Affect Disord.* 2003; 75(3):291–6. [PubMed: 12880942]
21. Freedman B, Glass KC, Weijer C. Placebo orthodoxy in clinical research. II.: Ethical, legal, and regulatory myths. *J Law Med Ethics.* 1996; 24(3):252–9. [PubMed: 9069852]
22. Rothman KJ, Michels KB. The continuing unethical use of placebo controls. *N Engl J Med.* 1994; 331(6):394–8. [PubMed: 8028622]
23. Fritze J, Moller HJ. Design of clinical trials of antidepressants: should a placebo control arm be included? *CNS Drugs.* 2001; 15(10):755–64. [PubMed: 11602002]
24. Kim SY. Benefits and burdens of placebos in psychiatric research. *Psychopharmacology (Berl).* 2003; 171(1):13–8. [PubMed: 12684741]
25. Khan A, Warner HA, Brown WA. Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials: an analysis of the Food and Drug Administration database. *Arch Gen Psychiatry.* 2000; 57(4):311–7. [PubMed: 10768687]
26. Laughren TP. The scientific and ethical basis for placebo-controlled trials in depression and schizophrenia: an FDA perspective. *Eur Psychiatry.* 2001; 16(7):418–23. [PubMed: 11728855]
27. Storosum JG, Wohlfarth T, Gispen-de Wied CC, et al. Suicide risk in placebo-controlled trials of treatment for acute manic episode and prevention of manic-depressive episode. *Am J Psychiatry.* 2005; 162(4):799–802. [PubMed: 15800158]

Research in Context

Evidence before this study

Despite widespread IRB approval of drug washouts and placebo-controlled clinical trials of psychiatric medications across the US, there is an ongoing debate regarding the ethical acceptability of these research procedures. Few published studies have presented data on which to base these concerns. To fully investigate the topic of drug washouts, a Scopus search was conducted in December, 2015, for studies where one of the search terms “washout, taper, or discontinuation” appeared in the title or keywords, and “depress* or bipolar” appeared in the abstract, title, or keywords. Out of over 700 manuscripts, only one measured outcomes of drug tapers before clinical trials. Because most clinical trials report placebo response, in order to examine the topic of placebo and research outcomes, we additionally searched for studies with either “depress* or bipolar” in the abstract, along with “placebo” and “ethic*”. Additional references were found by examining citations. The evidence in mood disorders demonstrates that withdrawal of ineffective medications does not, on average, alter mood. In addition, most clinical trials in depression demonstrate that, on average, patients with mood disorders improve while taking placebos. There is also little to no evidence examining the cumulative effects of research from screening to study withdrawal or completion. The effects of drug withdrawal for research purposes have not been examined in treatment-resistant subgroups or those with suicidal thoughts or behaviours.

Added value of this study

This study of a large, well-characterized cohort of treatment-resistant subjects with mood disorders, including some with prior or current suicidal thoughts or behaviors, found that medication withdrawals resulted in minimal worsening of symptoms, and outcomes did not differ for more severely ill groups. This study reports on the overall effects of research and demonstrates that research participants, on average, improved symptomatically over the course of research. These data may be used to inform a normative discussion of the acceptability of drug withdrawals, placebo-controlled trials, and research in individuals with severe mood disorders.

Implications of all the available evidence

The primary implication of this study is that it is possible to conduct such research safely—including medication withdrawals and placebo controls—even in severely ill participants, provided adequate infrastructure and safeguards are in place. While these results may not necessarily be generalizable to every research institution, we anticipate that our conclusions may be useful in a full normative discussion of the ethics of drug tapers and placebo-controlled trials.

Figure 1.

Bar graphs illustrating percent change in mood over A) drug taper, placebo lead-in, and administration of placebo in parallel and crossover trials (the first arm of the crossover only). B) percent change in mood from screening to withdrawal due to worsening and withdrawal for any other reason. Change in mood from screening to the end of research participation is also shown. Error bars represent 95% confidence intervals. PCB: placebo.

^aIndicates that the difference in mood change between *Suicidity+* and *Suicidity-* participants was statistically significant.

Figure 2.
Histogram of percent change in mood score over the duration of the drug taper/washout period (positive indicates increased severity of symptoms).

Table 1

Subject population demographics and clinical characteristics.

Cohort Characteristics	Mean	Standard Deviation
All Subjects (N)	504	
Mean age (years)	43.9	11.6
Mean age at onset (years, N=489)	21.0	11.0
Mean years ill (N=489)	22.9	1.2
Mean number of past suicide attempts (N=372)	0.6	1.5
Mean number of medications (N=256 medicated at screening)	2.7	1.4
Screening MADRS (N=272)	31.4	5.6
Screening HAM-D17 (N=19)	24.7	3.4
Screening HAM-D21 (N=11)	23.7	5.9
Screening HAM-D24 (N=4)	35.8	10.4
Screening YMRS (N=7)	28	6.1
	N	%
Female subjects	263	52.2
MDD	333	66.1
BD-I*	77	15.3
BD-II	94	18.7
Depressed at the time of admission	482	95.6
Manic at the time of admission	22	4.4
Axis I comorbidity	297	58.9
Comorbid anxiety disorder	248	49.2
Comorbid eating disorder	8	1.6
Comorbid anxiety and eating disorder	37	7.3
Other comorbidity	4	0.8
Enrolled in trial requiring treatment resistance	360	71.4
Family history of mood disorders (N=477)	345	68.5
History of alcohol/substance abuse (N=498)	186	36.9
History of abuse (N=493)	207	41.1
Medication naïve	48	9.5
Suicidal ideation at screening (N=465)	58	12.5
History of suicidal ideation (N=321)	191	59.5
Medicated at screening	256	50.8

* One extremely rapid cycling BD participant was removed from further analyses due to the potential of disorder-related fluctuations in mood state that could obscure changes in mood state due to the research intervention.

Abbreviations: MDD: major depressive disorder; BD: bipolar disorder; MADRS: Montgomery-Asberg Depression Rating Scale; HAM-D: Hamilton Depression Rating Scale; YMRS: Young Mania Rating Scale

Table 2

Frequency of withdrawals by time of withdrawal and reason for withdrawal.

Withdrawal Reason		Symptoms Worse	Symptoms Better	Side Effects	BD Switch	Criteria *	Other	Total
Withdrawal Phase	Screen	2	4	0	0	0	7	13
	Prospective	0	4	0	0	0	2	6
	Taper/Titrate	23	17	1	3	1	11	55
	Baseline	0	0	0	0	7	3	10
	PCB lead-in	7	11	2	2	0	12	34
	Trial	60	1	20	2	3	11	97
	TOTAL	92	37	23	7	11	46	215

*,"Criteria" indicates subjects who were withdrawn because they did not continue to meet study criteria.
Abbreviations: BD: bipolar disorder; PCB: placebo