

Prophylactic effects or withdrawal reactions? An analysis of time-to-event data from antidepressant relapse prevention trials submitted to the FDA

Michael P. Hengartner^{ID} and Martin Plöderl

Abstract

Background: Relapse prevention trials build the scientific foundation for recommendation of antidepressant continuation and maintenance therapy. However, the validity of the evidence is disputed and may be biased due to withdrawal confounding.

Methods: We analysed survival curves from all antidepressant relapse prevention trials submitted to the United States (US) Food and Drug Administration (FDA) between 1987 and 2012 for 13 approved drugs. The main outcome was the percent of the drug effect (placebo-antidepressant difference in relapse events) at any week of the maintenance phase in relation to the total drug effect at the endpoint of the randomised maintenance phase.

Results: Altogether, 14 studies with a mean observation period of 38.9 weeks (Kaplan–Meier estimators) were analysed. At week 3, a mean of 20.6% [95% confidence interval (CI) = 10.9–30.3%] of the total drug effect was achieved. At weeks 6 and 12, the corresponding figures were 50.3% (37.3–63.3%) and 69.0% (55.1–82.8%). No further antidepressant–placebo separation was observed as of week 24 [101.0% of total drug effect (94.6–107.3%)]. This means that censoring relapse events that occurred in the first 3, 6, 12 and 24 weeks would reduce the total drug effect at study endpoint by 20.6%, 50.3%, 69.0% and 101.0%, respectively. Assuming antidepressants had a constant prophylactic effect over 38.9 weeks, we further showed that, around week 6, the antidepressant–placebo separation was about three times larger than expected.

Conclusion: The placebo–antidepressant separation was disproportionally large between weeks 3 and 6 of the randomised maintenance phase. The benefits of continuing antidepressants relative to abrupt/rapid discontinuation declined sharply after week 6. This indicates an excess of relapse events in the placebo arms during the early maintenance phase that may be due to withdrawal reactions caused by abrupt/rapid discontinuation of active treatment. If these early relapse events are due to a direct pharmacological effect, then antidepressants' true prophylactic long-term effects are substantially overestimated.

Keywords: antidepressant, bias, prophylactic effects, relapse prevention, withdrawal confounding

Received: 4 February 2021; revised manuscript accepted: 23 June 2021.

Introduction

Depression treatment guidelines, including those published by the American Psychiatric Association (APA) and the National Institute for Health and

Care Excellence (NICE), strongly recommend continuation and maintenance of antidepressant treatment after remission to prevent relapses.^{1,2} These recommendations are, by and large, based

Ther Adv Psychopharmacol

2021, Vol. 11: 1–12

DOI: 10.1177/
20451253211032051

© The Author(s), 2021.
Article reuse guidelines:
sagepub.com/journals-
permissions

Correspondence to:
Michael P. Hengartner
Department of Applied
Psychology, Zurich
University of Applied
Sciences (ZHAW), PO Box
707, Zurich, CH-8037,
Switzerland

Medical Faculty, University
of Zurich, Switzerland
**michaelpascal.
hengartner@zhaw.ch**

Martin Plöderl
Department of Crisis
Intervention and Suicide
Prevention, University
Clinic for Psychiatry,
Psychotherapy, and
Psychosomatics, Christian
Doppler Clinic, Paracelsus
Medical University
Salzburg, Austria

on relapse prevention trials, where patients with remitted depression randomised to continued antidepressant treatment experience less relapse events than patients rapidly (often abruptly) discontinued from active treatment and switched to placebo.^{3,4} In these studies, the antidepressant–placebo difference in relapse events, which are assessed mostly *via* clinician depression rating scales, is then interpreted as a prophylactic drug effect. However, as detailed below, there is growing scepticism as to whether this outcome reflects a prophylactic drug effect in the antidepressant group or rather a withdrawal reaction in the placebo group.^{5–9}

Withdrawal confounding in relapse prevention trials

There is strong evidence from double-blind placebo-controlled randomised trials that abruptly/rapidly stopping antidepressant treatment can cause severe withdrawal symptoms.^{10–12} These withdrawal symptoms often resemble depression symptoms, for example, low mood, anxiety, sleep problems, agitation, irritability and gastrointestinal problems. Due to their distressing and debilitating nature, it is also conceivable that withdrawal symptoms such as brain zaps, dizziness and flu-like symptoms can reactivate (or trigger) a depression episode when patients cannot cope with them.^{5,13} In any case, various events coded as depression relapse in the placebo arms would then be a consequence of the withdrawal syndrome following abrupt/rapid drug discontinuation, that is, a direct pharmacological effect (also referred to as oppositional tolerance).^{14,15} Accordingly, several studies indicate that there is an excess of ‘relapses’ in patients randomised to placebo during the first few weeks after drug discontinuation and that, after 12 weeks, antidepressants protect minimally better against relapse than inert placebo.^{16–18} These findings suggest that continuing antidepressant treatment largely prevents withdrawal reactions rather than genuine depression relapses and that the true prophylactic effects of continued antidepressant treatment remain uncertain.⁷ A recent Cochrane review of relapse prevention trials thus concluded: ‘We cannot make any firm conclusions about effects and safety of the approaches studied to date. The true effect and safety are likely to be substantially different from the data presented due to assessment of relapse of depression that is confounded by withdrawal symptoms’.⁹

Addressing counterarguments

Despite strong evidence that withdrawal syndromes frequently occur after abrupt/rapid discontinuation of antidepressants,¹⁹ authors of depression practice guidelines and drug regulators contend that withdrawal can be ruled out as an explanation for the observed differences in relapse rates between antidepressant and placebo arms shown in relapse prevention trials. For instance, the National Institute for Health and Care Excellence (NICE), in their current practice guideline for adult depression, relying on an influential meta-analysis by Geddes *et al.*,⁴ argued, ‘The authors found no evidence to support the contention that the risk of relapse after withdrawal from active treatment in the placebo group was due to a direct pharmacological effect (for example, ‘withdrawal’ or ‘rebound’) since there was not an excess of cases within a month of drug discontinuation’.² However, Geddes *et al.* did not report data that would support this statement and they stated explicitly that their analysis ‘cannot exclude the possibility that there is a related effect’.⁴

Similar to NICE, the United States (US) Food and Drug Administration (FDA), in their review of relapse prevention trials submitted to the agency for marketing approval, stated: ‘Our analysis does not support the suggestion that the higher relapse rate in placebo arms reflects antidepressant withdrawal symptoms rather than the actual recurrence or relapse of a depressive episode. If this were the case, an excess of relapse events in the first 1 or 2 weeks of treatment discontinuation would be expected. Our data shows, however, that 94% of all relapse events occurred after the first 2 weeks postrandomisation in both the placebo and drug arms. In addition, our efficacy analyses that censored patients who had a relapse event in the first 2 and in the first 4 weeks of the double-blind phase, respectively, yielded results similar to the analysis including all postrandomisation relapse events’.³ However, their data analysis and interpretation has two major limitations. First, the FDA’s decision to restrict relapse excess to the first 1 or 2 weeks post-randomisation seems inappropriate, since several trials included in their analysis required that depression symptoms must be present for at least two consecutive visits (which usually take place every week in the early randomised maintenance phase) and/or that DSM-IV criteria for major depression must be met (which require a minimal symptom duration of 2 weeks). Finding an excess of relapses in the first 1 or 2 weeks was thus impossible by design in various trials. Second, the FDA

restricted censoring to the first 4 weeks, which yielded an absolute placebo–antidepressant risk difference in relapse rates of 12%, as compared with 19% for the uncensored analysis. This corresponds to a relative risk reduction of 37%. Although the FDA asserted that ‘results were little changed’,³ their data are compatible with the notion of a considerably reduced prophylactic drug effect. Moreover, according to the survival curves shown, the antidepressant–placebo separation becomes particularly pronounced around week 6 but, unfortunately, they did not report an analysis censoring events that occurred in the first 6 weeks.

Aims of the present study

The aim of this study was to re-analyse the Kaplan–Meier survival curves shown in the FDA review of relapse prevention trials.³ If antidepressants truly prevent genuine depression relapses, we would expect that the drug–placebo separation increases steadily over the study observation period. This is because the risk of relapse increases steadily over time after remission. Such a linear risk increase over at least 5 years after remission is due to the episodic course of depression and was demonstrated in numerous epidemiological studies.^{20–22} Therefore, if continued antidepressant treatment truly has prophylactic effects then, in patients randomised to antidepressants, the linear risk increase over time should be lower than in patients randomised to placebo. This results in a constantly increasing drug–placebo separation, which was indeed shown in an agomelatine maintenance study, an antidepressant agent that, contrary to most other newer antidepressants,^{23–25} does not cause withdrawal symptoms upon discontinuation (note that agomelatine was not approved by the FDA for depression and thus is not included in the present analysis).²⁶ By contrast, if antidepressants have minimal prophylactic effects and continued treatment after remission largely prevents withdrawal reactions rather than genuine depression relapses, then we would expect a disproportionately large drug–placebo separation in the first few weeks of the randomised maintenance phase and a sharply diminishing drug–placebo separation over time after the early maintenance period.

Methods

Data source

The data were extracted from all relapse prevention trials with 13 approved antidepressants

submitted to the FDA between 1987 and 2012 ($n = 15$ trials) as reported in Borges *et al.*³ One trial (study B) could not be included in the analysis because no survival curves (Kaplan–Meier estimators) were presented. We digitised the survival curves (Kaplan–Meier estimators) of the remaining 14 relapse prevention trials reported in Borges *et al.* with the software WebPlotDigitizer,^{3,27} using the average windows algorithm and three pixels for ΔY and ΔX . Manual corrections were necessary when the placebo and antidepressant survival curves overlapped or for breaks in the dashed survival curves. Study D reported the survival curves for two different antidepressant doses, so we used the mean of these two curves. We reduced the digitised data-points by only selecting the maximum value of relapse rates for each consecutive week. The data was extracted by the second author and cross-validated by the first author. The raw data and statistical code are freely available online on the Open Science Framework (OSF; see <https://osf.io/btm5p/>).

Outcomes and statistical analysis

For each study, we calculated the drug–placebo separation at the end of the observation period. This corresponds to the difference in relapse rates between placebo and antidepressants at the last point of the Kaplan–Meier survival curve for which data was available for both the placebo and antidepressant arms. The placebo–drug difference at the end of the observation period was then used as the denominator to calculate the relative percentage difference in relapse rates for each consecutive week. For example, if a trial showed a placebo–antidepressant difference in relapse rates of 10% at week 3 and of 20% at the end of the observation period, then this corresponds to a 50% relative difference in the outcome variable at week 3 (as 50% of the placebo–antidepressant endpoint-difference in relapse rates was achieved at week 3). For the sake of simplicity, we also denote the relative placebo–drug difference in relapse rates as ‘percent of total drug effect’.

We also calculated a ratio of observed *versus* expected drug effect at any point of the maintenance phase, assuming antidepressants had a constant prophylactic effect over the mean observation period of 38.9 weeks (the mean length of the survival curves depicted in Borges *et al.*³). Consequently, under this assumption of a constant prophylactic effect, at week 3 (7.7% of mean observation period completed) we would expect

7.7% of the total drug effect and, at week 6 (15.4% of mean observation period completed), we would expect 15.4% of the total drug effect. The ratio between percent of total drug effect and percent of mean observation period completed thus gives an indicator of the excess effect at any time point. All statistical analyses were conducted with R version 3.6.3.²⁸

According to Swiss law, this study was exempt from ethical approval, because it was a secondary analysis of anonymised data available in the public domain.

Results

Characteristics of included studies

We included 14 studies with 3874 participants altogether (2131 in the antidepressant arms and 1743 in the placebo arms). The investigated drugs were not detailed in Borges *et al.*,³ but given that all trials were submitted to the FDA between 1987 and 2012, they most likely included all approved selective serotonin reuptake inhibitors, the serotonin-norepinephrine reuptake inhibitors and various atypical antidepressants (e.g. bupropion, vilazodone, vortioxetine). Likewise, the mode of antidepressant discontinuation was not mentioned but, based on previous systematic reviews, it is likely that active treatment was stopped either abruptly or rapidly (i.e. by tapering over maximally 2 weeks) in most trials.^{9,17,29} The double-blind randomised maintenance phase lasted 24–52 weeks, with a mean of 35.1 weeks (median = 36 weeks). The mean observation period based on the Kaplan–Meier survival curves was 38.9 weeks. Note that this observation period is slightly longer than the mean trial duration (35.1 weeks), because survival analysis can also make projections beyond the actual trial duration. Relapse criteria differed widely between trials, including varying cut-off scores on clinician depression rating scales [Hamilton Depression Rating Scale (HDRS-17) or Montgomery-Asberg Depression Rating Scale (MADRS)], cut-offs on the Clinical Global Impression Severity scale (CGI-S), investigator's judgement, insufficient response and combinations thereof. For more details, see Table 1.

Antidepressant–placebo separation at different stages of the maintenance phase

Across the 14 trials, the mean percent of the total drug effect evolved in an obviously nonlinear

way, as depicted by a very steep increase in the first 6 weeks, a flattening curve between week 6 and 24, and no further effect thereafter, even though the mean observation period lasted for another 15 weeks (see Figure 1). The separation between placebo and antidepressants, that is, the prevention of relapses with continued active treatment compared with discontinuation and switching to placebo, was thus fully established in the first 24 weeks of the randomised maintenance phase, with no apparent prophylactic effects after week 24.

The drug effect was most pronounced in the first few weeks after randomisation (Table 2). At week 3, a mean percent of 20.6% (95% CI = 10.9–30.3%) of the total drug effect was achieved. The corresponding values for weeks 6, 12 and 24, respectively, were 50.3% (37.3–63.3%), 69.0% (55.1–82.8%) and 101.0% (94.6–107.3%). This means that censoring relapses that occurred in the first 3, 6 and 12 weeks would reduce the total drug effect by 20.6%, 50.3% and 69.0%, respectively. As of week 24, there was no further placebo–antidepressant separation at all, that is, censoring events until week 24 would fully dissolve the total drug effect measured at the end of the observation period (reduction by 101.0%). The latter finding also held true when we restricted the analysis to the eight trials with a maintenance phase that lasted longer than 26 weeks (i.e. 36–52 weeks duration), which showed a placebo–antidepressant separation of 97.9% (88.0–107.8%) at week 24 relative to the endpoint difference.

To demonstrate the disproportionately large placebo–antidepressant separation (i.e. excess of relapses in the placebo arm) at the early stages of the maintenance phase, we also plotted the percent of the total drug effect at specific time points against the percent of total observation period completed at those time points (see Figure 2). This analysis confirms that the placebo–antidepressant separation was most pronounced around week 6. It is further shown that the excess effect diminished continuously after week 6. The ratios at selected time points are also given in Table 2. An analysis of each study individually confirmed an excess effect in the first half of the maintenance phase in every single study. A pooled analysis of these individual study results fully replicated our main results reported above (for details, see Supplemental Figure S3; <https://osf.io/btm5p/>).

Table 1. Description of the included studies.

Study	Sample size	Duration open-label phase (weeks)	Duration randomised phase (weeks)	Relapse criteria
A	AD: 145 Pbo: 79	8	44	CGI-S ≥ 4 or discontinuation due to lack of efficacy
C	AD: 152 Pbo: 74	6–8	24	MADRS ≥ 25 and investigator's judgement
D	AD: 105 Pbo: 42	6–8	24	MADRS ≥ 22 and investigator's judgement
E	AD: 202 Pbo: 96	12 (incl. 3 weeks stabilisation)	38	Meeting DSM-IV criteria for MDD for 2 weeks or HDRS-17 ≥ 14 for 3 weeks
F	AD: 61 Pbo: 64	16 (incl. ≥ 2 weeks stabilisation)	36	HDRS-17 ≥ 18 at two consecutive visits, or investigator's judgement
G	AD: 106 Pbo: 107	26 (incl. 17 weeks stabilisation)	52	CGI-S ≥ 4
H	AD: 154 Pbo: 138	8	26	Meeting DSM-IV criteria for MDD and CGI-S ≥ 4 at two consecutive visits, or final CGI ≥ 4 if withdrawn from study
I	AD: 76 Pbo: 80	8–12 (incl. 2 weeks stabilisation)	40	Investigator's judgement, HDRS-17 ≥ 18 at a single visit, HDRS-17 of 15–17 at two consecutive visits, suicide or suicide attempt
J	AD: 207 Pbo: 210	8 (incl. 3 weeks stabilisation)	44	Investigator's judgement
K	AD: 149 Pbo: 163	10 (incl. 2–3 weeks stabilisation)	52	HDRS-17 ≥ 14 , CGI-S ≥ 3 (with ≥ 2 points increase, and meeting DSM-IV criteria for MDD at two consecutive visits)
L	AD: 181 Pbo: 92	8	36	MADRS ≥ 22 or discontinuation due to insufficient response
M	AD: 132 Pbo: 137	12 (incl. 3 weeks stabilisation)	26	CGI-S of ≥ 2 points increase and meeting DSM-IV criteria for MDD at two consecutive visits
N	AD: 189 Pbo: 185	12	24	HDRS-17 ≥ 16 , CGI-I ≥ 6 , or discontinuation for insufficient response
O	AD: 272 Pbo: 276	20 (incl. 12 weeks stabilisation)	26	HDRS-17 ≥ 16 , discontinuation for insufficient response, hospitalisation for depression, suicide attempt, or suicide

AD, antidepressant; CGI-S, Clinical Global Impression Severity scale; DSM-IV, Diagnostic and Statistical Manual of Mental Disorders, 4th edition; HDRS-17, Hamilton Depression Rating Scale; incl., including; MADRS, Montgomery-Asberg Depression Rating Scale; MDD, major depressive disorder; Pbo, placebo.

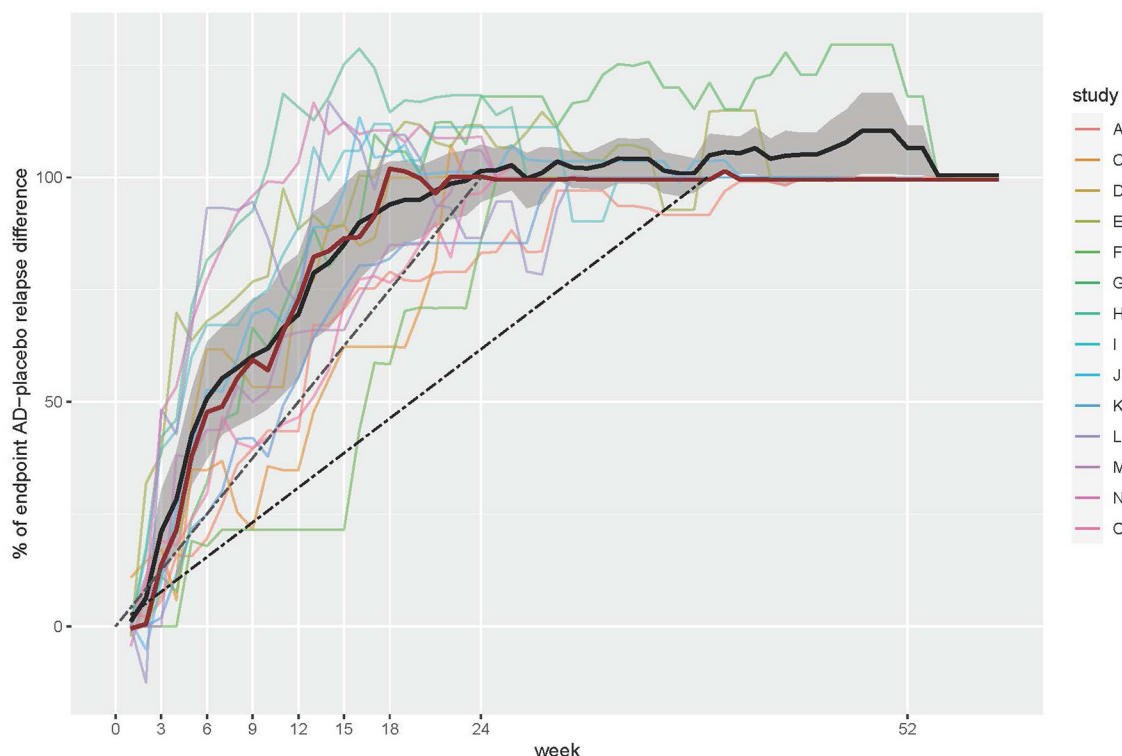


Figure 1. Percent of AD–placebo difference in relapse events at given time points, relative to the difference at the end of the observation period (percent of total drug effect). The bold black line denotes the mean value; the bold red line denotes the median value across studies. The grey area denotes the 95% CI for the mean. The dashed black line corresponds to the expected linear drug effect (constant prophylactic effect over time) based on the mean observation period [39 weeks]; the dashed grey line corresponds to the expected linear drug effect based on the minimum trial length [24 weeks]. AD, antidepressant; CI, confidence interval.

Table 2. Proportion of total drug effect at selected time points in relation to proportion of mean observation period completed (mean observation period = 38.9 weeks). The larger the ratio at given time points, the more the observed effect is in excess of an assumed constant prophylactic effect over the mean observation period.

Week	Percentage of total drug effect (95% CI)	Percentage of mean observation period completed	Ratio
3	20.6 (10.9–30.3)	7.7	2.7
6	50.3 (37.3–63.3)	15.4	3.3
12	69.0 (55.1–82.8)	30.9	2.2
24	101.0 (94.6–107.3)	61.8	1.6
39	104.4 (99.2–109.7)	100.4	1.0

CI, confidence interval.

Post hoc analyses

We conducted two sensitivity analyses. First, as suggested by one reviewer, using the mean may lead to biased results due to the different sample sizes of the trials. Therefore, we ran a sensitivity

analysis with weighted means, based on the number of patients at week zero provided in Figure 1 by Borges *et al.* (see also Table 1).³ The results with weighted means were virtually identical, therefore we present the unweighted analysis in

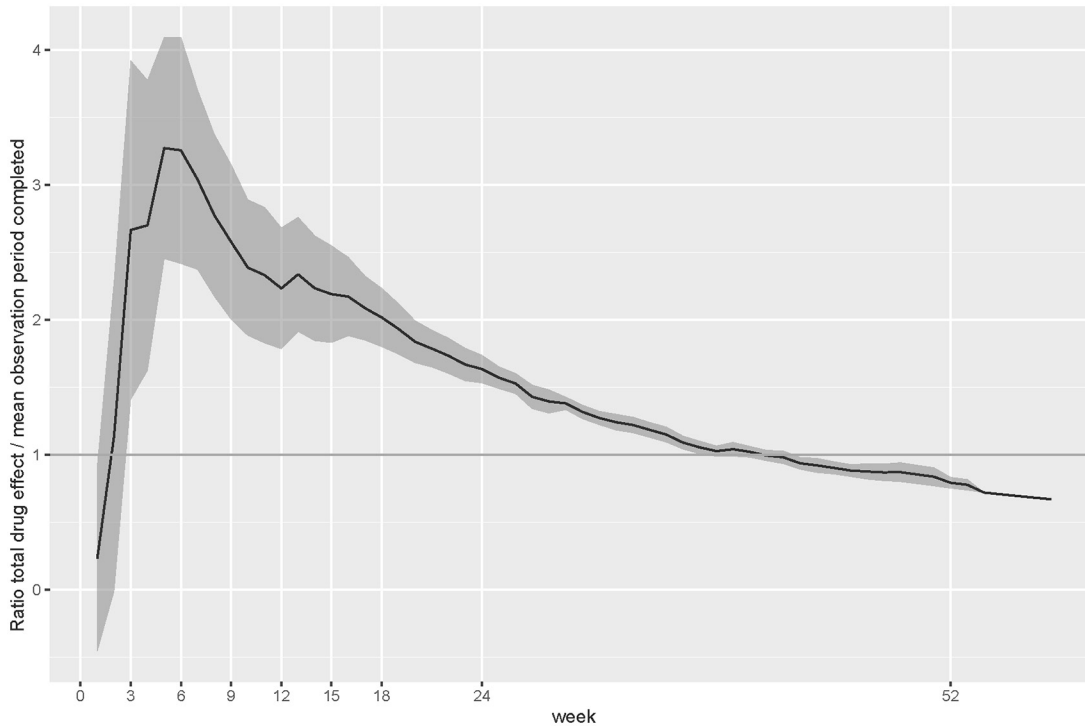


Figure 2. Proportion of total drug effect in relation to proportion of mean observation period completed as an indicator of excess antidepressant–placebo separation at any time point (the larger the ratio, the larger the excess effect). The bold black line represents the mean across all studies and the grey area the 95% CI for the mean.

CI, confidence interval.

the manuscript and provide the results for the weighted analysis in the online supplemental material (<https://osf.io/btm5p/>).

Second, we also restricted our analysis to trials with a stabilisation phase. In the eight trials with at least 2 weeks stabilisation before randomisation (studies E, F, G, I, J, K, M, O), the percent of total drug effect at weeks 3, 6 and 12 of the maintenance phase was 15.0% (3.8–26.2%), 42.0% (28.8–55.2%) and 65.1% (49.0–81.1%). The corresponding ratio for percent of total drug effect in relation to percent of mean observation period completed was 1.9 at week 3, 2.7 at week 6 and 2.1 at week 12. In the five trials with at least 3 weeks of stabilisation (studies E, G, J, M, O), the percent of total drug effect at weeks 3, 6 and 12 was 15.7% (3.1–28.3%), 45.2% (31.4–59.0%) and 70.9% (57.0–84.9%). The corresponding ratio for percent of total drug effect in relation to percent of mean observation period completed was 2.0 at week 3, 2.9 at week 6 and 2.3 at week 12. Trials with a stabilisation phase thus yielded results very similar to the main results reported above (for more details, see the online supplemental material;

<https://osf.io/btm5p/>). Likewise, there was no meaningful and consistent correlation between duration of the open-label pre-randomisation phase (range 6–26 weeks) and the antidepressant–placebo separation at weeks 6, 12 and 26, respectively, of the randomised maintenance phase (Spearman $r = -0.30$, $p = 0.30$; $r = 0.02$, $p = 0.95$ and $r = 0.26$, $p = 0.38$).

Discussion

Summary of findings

Our re-analysis of the Kaplan–Meier survival curves (time-to-event data) presented in the FDA review of antidepressant relapse prevention trials submitted to the agency between 1987 and 2012 showed that, over a mean observation period of 39 weeks, 21% the total drug effect (i.e. placebo–antidepressant difference in relapse rates at the end of the maintenance phase) was already achieved at week 3, and 50% (i.e. half) of the total drug effect at week 6. At week 12, on average, 69% of the total drug effect was achieved. After week 24, no further placebo–antidepressant separation was observed,

although, on average, the observation period lasted another 15 weeks. The result that there was no further placebo–antidepressant separation after week 24 also held true when we restricted our analyses to the eight trials that lasted longer than 26 weeks (i.e. between 36 and 52 weeks). These findings indicate that the benefits of continuing antidepressant treatment relative to abrupt/rapid discontinuation decline sharply after a few weeks. Censoring relapse events that occurred in the first 3, 6, 12 and 24 weeks reduced the total drug effect by 21%, 50%, 69% and 101%, respectively. Assuming antidepressants had a constant prophylactic effect over 39 weeks (the mean observation period based on the Kaplan–Meier survival curves), we further showed that, around week 6, the placebo–antidepressant separation was disproportionately large, that is, about 3 times larger than expected. Consistent with previous studies,^{8,16} our data thus provide empirical evidence for an excess of relapse events in the placebo arms in the early period of the maintenance phase, especially between weeks 3 and 6 after discontinuation of active treatment and switching to placebo.

Interpretation

The excess of relapse events shortly after discontinuation of active treatment could be due to a rapid return of depression symptoms in people randomised to placebo who were not in stable remission.⁶ However, as demonstrated by Borges *et al.*, with 19% versus 17%, the average antidepressant–placebo difference in relapse rates did not differ meaningfully between trials without a stabilisation phase (where re-emergence of suppressed but still active depression symptoms is likely after randomisation to placebo) and trials with a stabilisation phase (where most patients are probably in stable remission before randomisation, thus re-emergence of suppressed but still active depression symptoms is less likely).³ Likewise, when restricting our analysis to trials with a stabilisation phase, the results were very similar, thus confirming an excess of relapse events in the placebo arms in the first 6 weeks after discontinuation of active drug treatment despite of at least 3 weeks of stabilisation before randomisation. We neither found an association between duration of the open-label period (range 6–26 weeks) and the drug effect at weeks 6, 12 and 26 of the randomised maintenance phase. The notion of re-emergence of suppressed by still active depression symptoms in the placebo arms is thus not supported by our data.

An anonymous reviewer argued that the duration of open-label treatment phase and the number of previous episodes might account for the diminishing treatment effects over time, as patients randomised to placebo could experience symptom deterioration due to unblinding. However, as detailed above, we found no association between duration of the open-label phase and the antidepressant–placebo separation (drug effect) at weeks 6, 12 and 26 of the randomised maintenance phase. Given that Borges *et al.* did not report the number of previous depression episodes, we were not able to control for this factor.³ According to a previous meta-analysis by Kaymaz *et al.*,¹⁷ there is likewise no evidence that the duration of the open-label (i.e. pre-randomisation) phase affects the antidepressant–placebo separation at the end of the randomised maintenance phase, but in patients with recurrent depression the drug effect was significantly smaller. By contrast, a meta-analysis by Sim and colleagues found that a longer duration of the pre-randomisation phase, but not the number of previous depression episodes, did relate to a larger antidepressant–placebo separation.³⁰ The scientific literature is thus inconsistent.

Moreover, a major limitation is that the frequency and duration of pre-study treatment is a largely unknown confounder. The scarce information available indicates that most patients enrolled in relapse prevention trials had been repeatedly (or continuously) on antidepressants over many years, and this disregarded medication history likely influences (or biases) the results of these trials.^{5,7,13} That is, simply because a relapse prevention trial had a short open-label phase (with or without stabilisation), we cannot assume that the patients had not been exposed to long-term antidepressant treatment before the index episode. Unless future research accounts for pre-study medication history, we cannot know how previous antidepressant exposure affects the outcome of relapse prevention trials. According to Baldessarini and Tondo, ‘the benefits and adverse effects of prior treatment, even after it is discontinued, are likely to continue into the early days or even weeks of a new treatment trial – respectively, exaggerating the apparent benefits of a new treatment and either blunting (carry-over effect) or increasing (discontinuation effect) the impact of randomization to an inactive placebo, as well as misleadingly increasing ratings of adverse effects for a time’.⁵

Given that severe withdrawal reactions frequently meet common depression relapse criteria,^{11,12} a

plausible explanation for the disproportionately large placebo–antidepressant separation during the first few weeks of the maintenance phase are thus withdrawal reactions that eventually fulfil relapse criteria in patients randomised to placebo.^{6,7,9,16} Withdrawal symptoms can develop acutely within a few days after discontinuation, but, worthy of note, slow progression of symptoms and delayed onset are also possible.^{11,23,31} Therefore, it is not unusual that withdrawal syndromes fulfil relapse criteria only after a few weeks, especially when a drug taper was applied and after discontinuation of antidepressants with a long half-life.³² When we further take into account that many relapse prevention trials require that relapse criteria must be met at two consecutive visits or fulfil DSM-IV major depression criteria (i.e. at least two weeks symptom duration), it necessarily follows that an excess of relapse events in the placebo arms and thus a marked placebo–antidepressant separation will not be apparent in the first 2 weeks, but rather between weeks 3 and 6, as demonstrated by our data. These findings indicate that there is serious withdrawal confounding in relapse prevention trials,⁸ and that relapse prevention trials substantially overestimate antidepressants' true prophylactic long-term effects.^{33–35} A recent Cochrane review likewise concluded that the true prophylactic effects of antidepressants remain uncertain due to high risk of withdrawal bias in relapse prevention trials.⁹

By contrast, the NICE depression practice guideline asserts that there is no evidence that withdrawal reactions would account for the placebo–antidepressant separation in relapse prevention trials.² NICE reference this statement with an influential meta-analysis by Geddes *et al.*⁴ However, Geddes *et al.* did not empirically examine this fundamental issue.⁴ In the discussion they merely state, 'In this review, there did not seem to be an excess of cases within a month of drug discontinuation as is the case with lithium, but we cannot exclude the possibility that there is a related effect. If there is an effect, the effectiveness of continuation therapy could have been overestimated'. Thus, basically, they admit that withdrawal confounding cannot be excluded and that their estimates of prophylactic effects are possibly exaggerated. Moreover, whether there is an excess of relapses in the first month cannot be deduced from their study, because this would have required a survival analysis (time-to-event data). As the authors specify in their Methods, such data were not available: 'The published reports did not provide enough information to

allow a meta-analysis of time-to-relapse data. We therefore used tabulated data on total relapse rates at a specific point after randomisation'.⁴ Interestingly, the nearest timepoint after randomisation that they reported was 6 months, thus it is unclear how they could claim that 'there did not seem to be an excess of cases within a month of drug discontinuation'.

In their own review, the FDA did empirically examine potential withdrawal confounding. The authors concluded that there is no excess of relapses in the first 2 weeks in the placebo arms, since 94% of relapses occurred after week 2.³ However, an excess of relapses in the first 2 weeks is largely excluded due to the trial designs. As detailed above, many trials required that a certain threshold on a depression rating scale or the CGI must be met at two consecutive visits and/or that DSM-IV criteria for major depression must be satisfied (which require symptom duration of at least 2 weeks). In line with this, we found a disproportionately small placebo–antidepressant separation up to week 2, but a strong excess effect between weeks 3 and 6 (see Figure 2). The FDA also censored relapses that occurred in the first 2 and 4 weeks and concluded that these results did not differ meaningfully from the uncensored main analysis. However, as we detailed in the Introduction, the absolute placebo–antidepressant difference in relapse rates was 12% in the 4-week censored analysis as compared with 19% in the uncensored analysis. This corresponds to a relative reduction of the drug effect by 37%, which is not negligible. Contrary to their own interpretation, the results of the FDA analysis are thus compatible with the notion that censoring of relapses occurring in the first 4 weeks indeed leads to a substantially reduced drug effect. As shown by our analysis, this effect was even more pronounced when relapses occurring in the first 6 weeks were censored (drug effect reduced by 50%). Thus, it is obvious that censoring relapse events up to weeks 4 and 6 continuously and substantially reduces the total drug effect.

Limitations

The main limitation of our analysis is that the FDA did not specify which survival curves correspond to which drug, so there was no possibility to differ between antidepressant drugs or classes. Likewise, due to very limited information provided in the original FDA study, we were not able to assess the risk of bias in individual studies, to measure effect

variability among studies (i.e. heterogeneity), or to grade the certainty of evidence. The relapse criteria also differed widely between studies, and some included highly subjective criteria (i.e. investigator's judgement, insufficient response). Whether such opaque relapse criteria affected our results could not be determined. Given that every trial used a unique definition of relapse, there was no possibility to statistically adjust for it or to examine differences based on relapse definition. In addition, our analysis was restricted to trials submitted to the FDA for marketing approval. With 14 studies, our data set was thus smaller than in other reviews of relapse prevention trials. On the other hand, this dataset is arguably representative and less biased by selective reporting, which is a serious issue in antidepressant trials.^{36–38} Another limitation is that our analyses were based on data digitised from printed figures. This may introduce some imprecision. However, it is very unlikely that this has introduced systematic bias. We also acknowledge that we did not measure withdrawal syndromes and thus cannot directly assess the association between withdrawal reactions and relapse events. The best approach to measure withdrawal confounding would be to directly assess withdrawal symptoms, but this outcome is almost categorically neglected in relapse prevention trials.⁹ Unless future studies routinely include such a measure, a stringent test (and quantification) of withdrawal bias in relapse prevention trials remains elusive. Finally, we previously wrote about the disproportionately large drug–placebo separation shortly after randomisation depicted in the FDA review,⁷ but we did not preregister a study protocol. The analysis may thus be conceived of as exploratory rather than confirmatory.

Conclusion

Our analysis of time-to-event data showed that there is a disproportionately strong drug effect between weeks 3 and 6 in antidepressant relapse prevention trials with discontinuation design due to an excess of relapse events in patients randomised to placebo. After week 6, the drug effect declines sharply and approaches zero at week 24. Based on these findings we suggest that the excessive drug effect observed between weeks 3 and 6 was likely driven by withdrawal syndromes in patients randomised to placebo. Withdrawal syndromes following abrupt/rapid discontinuation of antidepressants are scientifically well established in double-blind randomised placebo-controlled trials.^{10–12,39} There is also robust scientific evidence that antidepressant

withdrawal comprises various symptoms that qualify as depression relapse, including low mood, suicidality, irritability, agitation, anxiety and sleep problems.^{23,25,40} The most plausible explanation therefore is that genuine depression relapses and withdrawal reactions are confounded in the early stages of the double-blind randomised maintenance phase.⁷ However, as our analysis did not assess the presence of withdrawal symptoms, we cannot firmly exclude alternative explanations. To provide a stringent test and quantification of withdrawal confounding, relapse prevention trials should assess and report the frequency and severity of withdrawal symptoms in both antidepressant and placebo arms. Despite the various limitations detailed above, in line with the authors of a recent Cochrane review and other experts, we conclude that the results of relapse prevention trials cannot provide valid estimates of long-term prophylactic (relapse-preventive) effects.^{7,8,16,41} The true prophylactic effects of antidepressants are likely much smaller than the antidepressant–placebo difference in relapse rates suggests.

Conflict of interest statement

The authors declare that there is no conflict of interest.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Michael P. Hengartner  <https://orcid.org/0000-0002-2956-2969>

Supplemental material

The supplemental material is present in the below link. <https://osf.io/btm5p/>

References

1. American Psychiatric Association. *Practice guideline for the treatment of patients with major depressive disorder*. Washington, DC: American Psychiatric Association, 2010.
2. National Institute for Health and Care Excellence. Depression in adults: recognition and management, nice.org.uk/guidance/cg90 (2009, accessed 17 October 2017).
3. Borges S, Chen YF, Laughren TP, *et al*. Review of maintenance trials for major depressive disorder: a 25-year perspective from the US Food

- and Drug Administration. *J Clin Psychiatry* 2014; 75: 205–214.
4. Geddes JR, Carney SM, Davies C, *et al.* Relapse prevention with antidepressant drug treatment in depressive disorders: a systematic review. *Lancet* 2003; 361: 653–661.
 5. Baldessarini RJ and Tondo L. Effects of treatment discontinuation in clinical psychopharmacology. *Psychother Psychosom* 2019; 88: 65–70.
 6. El-Mallakh RS and Briscoe B. Studies of long-term use of antidepressants: how should the data from them be interpreted? *CNS Drugs* 2012; 26: 97–109.
 7. Hengartner MP. How effective are antidepressants for depression over the long term? A critical review of relapse prevention trials and the issue of withdrawal confounding. *Ther Adv Psychopharmacol* 2020; 10: 2045125320921694.
 8. Recalt AM and Cohen D. Withdrawal confounding in randomized controlled trials of antipsychotic, antidepressant, and stimulant drugs, 2000–2017. *Psychother Psychosom* 2019; 88: 105–113.
 9. Van Leeuwen E, van Driel ML, Horowitz MA, *et al.* Approaches for discontinuation versus continuation of long-term antidepressant use for depressive and anxiety disorders in adults. *Cochrane Database Syst Rev* 2021; 4: CD013495.
 10. Baldwin DS, Montgomery SA, Nil R, *et al.* Discontinuation symptoms in depression and anxiety disorders. *Int J Neuropsychopharmacol* 2007; 10: 73–84.
 11. Michelson D, Fava M, Amsterdam J, *et al.* Interruption of selective serotonin reuptake inhibitor treatment. Double-blind, placebo-controlled trial. *Br J Psychiatry* 2000; 176: 363–368.
 12. Rosenbaum JF, Fava M, Hoog SL, *et al.* Selective serotonin reuptake inhibitor discontinuation syndrome: a randomized clinical trial. *Biol Psychiatry* 1998; 44: 77–87.
 13. Healy D. Treatment-induced stress syndromes. *Med Hypotheses* 2010; 74: 764–768.
 14. Andrews PW, Kornstein SG, Halberstadt LJ, *et al.* Blue again: perturbational effects of antidepressants suggest monoaminergic homeostasis in major depression. *Front Psychol* 2011; 2: 159.
 15. Fava GA. May antidepressant drugs worsen the conditions they are supposed to treat? The clinical foundations of the oppositional model of tolerance. *Ther Adv Psychopharmacol* 2020; 10: 2045125320970325.
 16. Greenhouse JB, Stangl D, Kupfer DJ, *et al.* Methodologic issues in maintenance therapy clinical trials. *Arch Gen Psychiatry* 1991; 48: 313–318.
 17. Kaymaz N, van Os J, Loonen AJ, *et al.* Evidence that patients with single versus recurrent depressive episodes are differentially sensitive to treatment discontinuation: a meta-analysis of placebo-controlled randomized trials. *J Clin Psychiatry* 2008; 69: 1423–1436.
 18. Viguera AC, Baldessarini RJ and Friedberg J. Discontinuing antidepressant treatment in major depression. *Harv Rev Psychiatry* 1998; 5: 293–306.
 19. Hengartner MP, Davies J and Read J. Antidepressant withdrawal - the tide is finally turning. *Epidemiol Psychiatr Sci* 2019; 29: e52.
 20. Eaton WW, Shao H, Nestadt G, *et al.* Population-based study of first onset and chronicity in major depressive disorder. *Arch Gen Psychiatry* 2008; 65: 513–520.
 21. Hardeveld F, Spijker J, De Graaf R, *et al.* Recurrence of major depressive disorder across different treatment settings: results from the NESDA study. *J Affect Disord* 2013; 147: 225–231.
 22. Ten Have M, de Graaf R, van Dorsselaer S, *et al.* Recurrence and chronicity of major depressive disorder and their risk indicators in a population cohort. *Acta Psychiatr Scand* 2018; 137: 503–515.
 23. Fava GA, Gatti A, Belaise C, *et al.* Withdrawal symptoms after selective serotonin reuptake inhibitor discontinuation: a systematic review. *Psychother Psychosom* 2015; 84: 72–81.
 24. Fava GA, Benasi G, Lucente M, *et al.* Withdrawal symptoms after serotonin-noradrenaline reuptake inhibitor discontinuation: systematic review. *Psychother Psychosom* 2018; 87: 195–203.
 25. Jha MK, Rush AJ and Trivedi MH. When discontinuing SSRI antidepressants is a challenge: management tips. *Am J Psychiatry* 2018; 175: 1176–1184.
 26. Goodwin GM, Emsley R, Rembry S, *et al.*; Agomelatine Study Group. Agomelatine prevents relapse in patients with major depressive disorder without evidence of a discontinuation syndrome: a 24-week randomized, double-blind, placebo-controlled trial. *J Clin Psychiatry* 2009; 70: 1128–1137.
 27. Rohatgi A. WebPlotDigitizer, <https://automeris.io/WebPlotDigitizer> (accessed 20 December 2020).

28. R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2020.
29. Cohen D and Recalt A. Discontinuing psychotropic drugs from participants in randomized controlled trials: a systematic review. *Psychother Psychosom* 2019; 88: 96–104.
30. Sim K, Lau WK, Sim J, *et al*. Prevention of relapse and recurrence in adults with major depressive disorder: systematic review and meta-analyses of controlled trials. *Int J Neuropsychopharmacol* 2015; 19: pyv076.
31. Chouinard G and Chouinard VA. New classification of selective serotonin reuptake inhibitor withdrawal. *Psychother Psychosom* 2015; 84: 63–71.
32. Horowitz MA and Taylor D. Tapering of SSRI treatment to mitigate withdrawal symptoms. *Lancet Psychiatry* 2019; 6: 538–546.
33. Ghaemi SN. Why antidepressants are not antidepressants: STEP-BD, STAR*D, and the return of neurotic depression. *Bipolar Disord* 2008; 10: 957–968.
34. Hengartner MP. Methodological flaws, conflicts of interest, and scientific fallacies: implications for the evaluation of antidepressants' efficacy and harm. *Front Psychiatry* 2017; 8: 275.
35. Pigott HE, Leventhal AM, Alter GS, *et al*. Efficacy and effectiveness of antidepressants: current status of research. *Psychother Psychosom* 2010; 79: 267–279.
36. de Vries YA, Roest AM, de Jonge P, *et al*. The cumulative effect of reporting and citation biases on the apparent efficacy of treatments: the case of depression. *Psychol Med* 2018; 48: 2453–2455.
37. Melander H, Ahlqvist-Rastad J, Meijer G, *et al*. Evidence b(i)ased medicine—selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* 2003; 326: 1171–1173.
38. Turner EH, Matthews AM, Linardatos E, *et al*. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008; 358: 252–260.
39. Tint A, Haddad PM and Anderson IM. The effect of rate of antidepressant tapering on the incidence of discontinuation symptoms: a randomised study. *J Psychopharmacol* 2008; 22: 330–332.
40. Haddad PM and Anderson IM. Recognising and managing antidepressant discontinuation symptoms. *Adv Psychiatr Treat* 2007; 13: 447–457.
41. Ghaemi SN and Selker HP. Maintenance efficacy designs in psychiatry: randomized discontinuation trials - enriched but not better. *J Clin Transl Sci* 2017; 1: 198–204.