REVIEW

Declining efficacy in controlled trials of antidepressants: effects of placebo dropout

Stein Schalkwijk[1,2,3], Juan Undurraga[2,4,6], Leonardo Tondo[1,2,5] and Ross J. Baldessarini[1,2]

[1] *Department of Psychiatry, Harvard Medical School, Boston, MA, USA*

[2] *International Consortium for Research on Mood & Psychotic Disorders, McLean Hospital, Belmont, MA, USA*

[3] *Department of Pharmaceutics, Utrecht Institute for Pharmaceutical Sciences (UIPS), Utrecht University, Netherlands*

[4] *Bipolar Disorders Program, Institute of Neuroscience, Hospital Clínic Barcelona, IDIBAPS, CIBERSAM, University of Barcelona, Spain*

[5] *Lucio Bini Mood Disorders Centre, Cagliari and Rome, Italy*

[6] *Department of Psychiatry, Clinica Alemena, Universidad del Desarrollo, Santiago, Chile*

Abstract

Drug-placebo differences (effect-sizes) in controlled trials of antidepressants for major depressive episodes have declined for several decades, in association with selectively increasing clinical improvement associated with placebo-treatment. As these trends require adequate explanation, we tested the hypothesis that decreasing trial-dropout rates may be an important contributor. We gathered reports of peer-reviewed, placebo-controlled trials of antidepressants (1980–2011) by computerized literature searching, and applied meta-analysis, meta-regression and multiple linear regression methods to evaluate associations of dropout rates and other factors of interest, to reporting year and reported efficacy [standardized mean drug-placebo difference (SMD) as Hedges' *g*-statistic]. In 56 trials meeting inclusion and exclusion criteria, we confirmed significant overall efficacy of antidepressants but declining drug-placebo contrasts over the past three decades. Among other changes, there was a corresponding increase in placebo-associated improvement with a decline in placebo-dropout rate, mainly for lack of efficacy. These effects were found only when last-observation-carried-forward (LOCF) analyses were used. Other trial-design and subject factors, including drug-responses and drug-dropout rates, were much less associated with efficacy. We propose that declining placebo-dropout rates ascribed to inefficacy combined with use of LOCF analyses led to increasing improvement in placebo-arms that contributed to declining antidepressant–placebo contrasts in controlled treatment trials since the 1980s.

Introduction

Antidepressants are currently a standard treatment for major depressive disorder, supported by comparisons to placebo treatment under blinded conditions in randomized controlled trials (RCTs) (Baldessarini, 2013). Although development, testing and licensing of new antidepressants has continued apace over the past several decades, several reviews have identified a decline in antidepressant-placebo differences (effect-size) in antidepressant trials, with a substantial proportion of trials failing statistically to distinguish responses to treatment with antidepressants *vs.* placebo (Walsh et al., 2002; Khan et al., 2003a, 2010; Undurraga and Baldessarini, 2012; Baldessarini, 2013). These trends have been associated with selective increases in clinical responses associated with placebos and stable responses to drug-treatment, with corresponding decreases in drug-placebo

Address for correspondence: Dr R. J. Baldessarini, Mailman Research Centre, McLean Hospital, 115 Mill Street, Belmont, MA 02478-9106, USA.
Tel.: 1-617-855-3203 *Fax*: 1-617-855-3479
Email: rbaldessarini@mclean.harvard.edu

differences (Walsh et al., 2002; Khan et al., 2003a, 2010; Undurraga and Baldessarini, 2012; Undurraga et al., 2013).

Effect-sizes based on drug-placebo differences for specific drugs usually are pooled and compared in comprehensive, systematic reviews employing meta-analysis (Hedges, 1981; Borenstein et al., 2009). Trials considered for such reviews differ in design and conduct as well as in characteristics of participants and study-sites, assessment methods, duration, dosing of drugs and outcome measures. Such heterogeneity may contribute to differences in results among trials of even nominally identical treatments, greatly complicating efforts to determine best estimates of effect-sizes and to compare specific treatments by relative efficacy. Attempts have been made to identify methodological features and patient characteristics that contribute to heterogeneity and may account for inter-trial variability and the steady decline of effect-size in antidepressant trials over the years. Proposed factors include *trial-duration* (Khan et al., 2010; Henkel et al., 2012), *trial complexity* associated with the number of participating sites and subjects (Khan et al., 2004; Henkel et al., 2012; Undurraga and Baldessarini, 2012)

minimum required depression scores or other indications of *baseline illness-severity* for inclusion (Khan et al., 2002, 2007; Kirsch et al., 2008), pre-trial placebo *washout of previous drugs* (Lee et al., 2004), depression *rating-scale* employed (Khan et al., 2010), number of treatment arms or *probability of randomization to placebo* (Khan et al., 2004; Papakostas and Fava, 2009), *frequency of assessments* (Iovieno et al., 2011), and *flexible vs. fixed dosing* (Khan et al., 2003b). Although many of these factors may well contribute to variance of trial-outcomes, it is our impression that the decrease of effect-size across years, in particular, remains to be explained adequately.

A particularly striking change in antidepressant trials has been a steady and selective increase in reported responses among subjects randomized to placebo (Walsh et al., 2002; Undurraga and Baldessarini, 2012). This increase might be related to changing characteristics of patient samples or trial methods. A possible, specific contributor may be change in dropout rates over the past decades. Individual trials usually analyse the data on an intention-to-treat (ITT) basis, so that virtually all randomized subjects are included in analyses of symptomatic change between trial intake and protocol endpoint (Tierney and Stewart, 2005). However, subjects tend to leave trials prematurely for various reasons that typically are dissimilar in drug (usually poor tolerability) and placebo arms (typically, lack of efficacy) of trials (Little et al., 2012). A particularly prevalent way to conduct an ITT analysis is to apply the last observation carried forward (LOCF) method. This method can be problematic, especially when there are systematic differences in the timing and reasons for dropping out in different trial-arms (Hennen, 2003; Lane, 2008; Wittes, 2009; Little et al., 2012). Dropouts combined with LOCF analyses can distort trial outcomes in trials of antipsychotic drugs (Kemmler et al., 2005; Rabinowitz and Davidov, 2008; Hutton et al., 2012), but their effects are less well studied in trials of antidepressants.

Given these considerations, we hypothesized that changes in the proportion of dropouts in antidepressant trials, particularly in placebo-arms, in combination with LOCF methods for ITT endpoint analysis, may contribute significantly to declining antidepressant-placebo differences observed in meta-analyses of antidepressant trials over the past several decades. This hypothesis is encouraged by the observation that the LOCF method may underestimate placebo-associated responses in antidepressant trials, due to leaving trials early with less chance of spontaneous clinical improvement (Gomeni et al., 2009). That is, more and longer retention in placebo-treatment should lead to greater spontaneous improvement, given the time-limited nature of most acute episodes of major depressive disorder. In turn, increased improvements with placebo would yield smaller drug-placebo contrasts, especially if drug-associated retention were longer. To test this hypothesis, we quantified changes of dropout rates in both placebo and antidepressant arms of a large, representative sample of peer-reviewed, placebo-controlled trials of antidepressants from the 1980s to the present, and assessed reported changes in improvements in depressive symptoms between antidepressant and placebo groups.

Method

Search strategy

We systematically sought findings from RCTs of antidepressant drugs in major depression in peer-reviewed reports since 1980. The search methods were detailed previously (Undurraga and Baldessarini, 2012), and consisted of a computerized literature searching of Medline, CINAHL, Cochrane and PsycINFO medical research literature databases. Search-terms included individual generic names of all clinically employed antidepressants with regulatory approval in the US, plus fluvoxamine, for treatment of acute episodes of major depression, terms for pharmacological classes of antidepressants (serotonin-reuptake inhibitors, tricyclic antidepressants and others), as well as 'major depressive disorder' and 'depression.' In addition, reference citations in reports and reviews on antidepressant efficacy were hand-searched for additional, relevant reports.

Eligibility and study selection

Data considered were limited to those provided in peer-reviewed reports of randomized, placebo-controlled, monotherapy RCTs in acute episodes of adult unipolar major depressive disorder (or with <10% bipolar depressed subjects). Studies with up to 10% bipolar subjects were considered because in earlier studies (mainly in the 1980s) inclusion required 'major depression' without necessarily excluding cases of bipolar depression. All subjects were diagnosed by standard international criteria (DSM-III, III-R, or -IV, ICD-9 or -10 or RDC), as reported between 1980 and 2011. Since our aim was to test for trends over time in comparable trials, and not to quantify efficacy, we did not attempt to acquire data from reports not based on peer review or not published, particularly since access to such findings, especially from earlier trials, is likely to be incomplete. In addition, we did not include drugs with questionable efficacy (e.g. hypericum, gepirone) or without regulatory approval in the US (e.g., agomelatine, reboxetine) (Baldessarini 2013). Since meta-analyses of antidepressant trials identifying a decrease in effect-sizes over the years (Walsh et al., 2002; Khan et al., 2010; Undurraga and Baldessarini, 2012) included outcomes based almost entirely on intention-to-treat (ITT) analyses, these were therefore an inclusion criterion for this study. ITT analysis seeks to include all subjects randomized to all trial treatment-arms, and typically considers outcomes among all subjects with at least one post-randomization assessment (Tierney and Stewart, 2005). We defined ITT as all subjects with at least one

post-baseline assessment at least one week after randomization. Dropout rates from placebo and antidepressant trial-arms, and their proposed causes, also were recorded. Abstracts of initially identified reports were screened for possible relevance, verified by independent review of full-texts by two investigators who selected trials included for analysis, based on their meeting inclusion/exclusion criteria, by consensus.

Outcome measures

As an outcome measure of treatment-response, we employed the standardized mean difference [(SMD) as Hedges' *g*-statistic] of scores on standard depressive symptom rating-scales between intake and exit points in each trial for paired-comparisons of symptomatic changes associated with antidepressant and placebo (Hedges, 1981; Borenstein et al., 2009). Rating scales considered were the 17- or 21-item versions of the Hamilton Depression Rating Scale (HDRS; Hamilton, 1967), or the Montgomery-Åsberg Depression Rating Scale (MADRS; Montgomery and Åsberg, 1979).

Retrieved data

Abstracted data from drug and placebo arms of trials included efficacy, with computed SMD, total dropout rates, as well as rates ascribed to lack of efficacy, intolerability or adverse events, or miscellaneous other factors. In order to test for independent associations of dropout rates with effect-size (SMD), we also considered several methodological factors reported to be associated with trial outcomes or otherwise of plausible interest, as summarized in the introduction. These included planned trial-duration, number of participating subjects and collaborating sites, baseline illness-severity, symptom rating-scale employed, total daily antidepressant dose (as imipramine-equivalent mg/d [Baldessarini, 2013]), as well as average age and proportion of male or female participants. Because the number of studies reporting required information needed was limited, we restricted the number of trial features and patient characteristics included in reported analyses.

Data analyses

To manage data from trials involving multiple comparisons, trials with more than one antidepressant-arm were pooled into a single drug-arm by weighted-averaging (Caldwell et al., 2005). When arms with different doses of the same drug were employed, a single arm was made by averaging the depression ratings and changes and assuming a mean dose of drug (both weighted by subject-count). Linear regression was used to quantify temporal trends in dropout rates and other factors over the past three decades. We employed random-effects meta-analytic modelling, with SMD (*g*-statistic) as the outcome measure, to test for drug-placebo differences in clinical change (Hedges, 1981; Borenstein et al., 2009).

Changes in depression ratings with drug or placebo treatment were standardized by subject-counts, and variance was standard deviation (SD) reported, calculated or imputed from pooled SD from all trials (Furukawa et al., 2006). Dropout rates were tested for independent association with SMD using meta-regression analysis, including reporting year as well as other factors with at least suggestive associations ($p \leqslant 0.10$). Egger's test was used to test for effects of small studies or possible publication bias (Egger et al., 1997).

Endpoint-analysis based only on trial protocol-completers is biased by omitting subjects dropping out of trials prematurely, and potentially also by imputing missing values at later times by LOCF analysis. To evaluate the influence of these methods, we included a *post-hoc* analysis using trials that reported both completer and LOCF analyses.

Averaged data are reported as means±SD, as medians with interquartile range (IQR; especially for non-normally distributed measures), or as measures with 95% confidence intervals (CI), as noted. Analyses were performed with commercial software (Stata.10®, StataCorp, College Station, TX; Statview 5.0®, SAS Institute, Cary, NC).

Results

Electronic and hand searching yielded 2437 potentially relevant reports appearing between 1980 and 2011. Based on reviewing titles and abstracts, 179 reports initially appeared to be potentially eligible for analysis. Subsequent exclusions (123/179) were for the following reasons: [a] 17 studies involved <20 patients/arm; [b] 17 included >10% of subjects with diagnoses other than non-bipolar major depressive episode; [c] 11 involved special populations (such as juvenile or geriatric patients); [d] 4 did not include a placebo control arm; [e] 19 did not allow an ITT analysis; [f] 21 did not provide information required to calculate SMD; [f] 12 represented subpopulations of trials already included; and [g] another 22 did not provide adequate information on dropouts or reasons for dropout. Overall 56 trials, corresponding to 175 treatment-arms, met all study inclusion/exclusion criteria and were analysed. In 24 of the 56 trials, two antidepressant arms were condensed into one. Salient characteristics and references of the 56 trials are summarized in an Appendix (Supplementary Table S1).

Studies analysed included a total of 17189 consenting, adult, depressed subjects randomized to either an active antidepressant (*n*=11349) or a placebo (*n*=5840). Most trials (87.0%) involved only outpatients, and 95.6% had pharmaceutical sponsorship. Three trials (5.4% of trials, 0.2% of subjects) included <10% bipolar depressed patients. All trials employed LOCF methods to deal with missing data. Owing to different trial-arm counts per trial, the overall median chance of randomization to placebo was 0.33 (range: 0.14–0.50), and median trial-duration was 8.0 (4.0–13) weeks, following 7.0

Table 1. Preliminary analysis of factors associated with reporting year and effect-size in controlled trials of antidepressants

a. Linear regression for factors of interest *vs.* year of reporting (*n*=56)

Factors	Slope (*β*) [95% CI]	*t*-score	*p*-value
Subject age (years)	+0.067 [−0.048 to +0.183]	1.17	0.25
Women/trial (%)	+0.281 [−0.013 to +0.574]	1.92	0.06
Sample size (*N*)	+8.00 [+1.97 to +14.0]	2.66	0.01
Trial duration (weeks)	+0.110 [+0.052 to +0.169]	3.78	0.0004
Dose (IMI-eq mg/day)	+0.824 [−2.41 to +4.06]	0.51	0.61
Dropout rate (%)			
Overall	−1.23 [−1.56 to −0.897]	7.41	<0.0001
Drug	−1.01 [−1.36 to −0.652]	5.68	<0.0001
Placebo	−1.54 [−1.95 to −1.12]	7.37	<0.0001

b. Meta-regression analysis for factors associated with effect-size (*n*=56)

Factors	Slope (*β*) [95% CI]	*t*-score	*p*-value
Total dropout rate (%)			
Antidepressants	−0.004 [−0.01 to +0.001]	1.47	0.15
Placebo	+0.007 [+0.002 to +0.012]	2.89	0.006
Reporting year	−0.009 [−0.016 to −0.002]	2.68	0.01
Trial duration (weeks)	+0.006 [−0.017 to +0.029]	0.54	0.59
Sample size (*N*)	<−0.001 [−0.0002 to +0.0002]	0.17	0.87

(4–16) days of preliminary placebo-washout of previous treatments in the 82.1% of trials reporting such information. Drugs considered (*n*=14), ranked by frequency of testing, were: imipramine (18.0%), venlafaxine (15.0%), fluoxetine (12.3%), desvenlafaxine (10.0%), paroxetine (10.0%), duloxetine (7.50%), amitriptyline (5.00%), escitalopram (5.00%), sertraline (5.00%), bupropion (3.75%), citalopram (3.75%), fluvoxamine (1.25%), mirtazapine (1.25%), and trazodone (1.25%). Some trials with multiple arms also involved agents such as hypericum and gepirone not approved for depression and not considered in the present study.

Trials analysed included a median (range) size of 290 (52–717) subjects from a median of 12 (1–69) collaborating sites; 63% (37–85%) of subjects were women, and median age was 40.6 (34.4–56.0) years. Clinical ratings involved the 17- or 21-item versions of the HDRS (51.8 and 33.9%) or the MADRS (14.3%), with a median of 3 (2–5) assessments and depression-ratings/month (6 [4–10] total). Median baseline-depression scores (based on majority-use of the 17-item HDRS only, for comparability) were 23.1 (17.7–26.4) among subjects randomized to drugs and 23.4 (17.2–26.6) with placebo. Baseline depression-severity did not change significantly over the years of reporting, controlling for depression rating scales employed (slope [*β*]=−0.065 [−0.147 to +0.018]; *t*=1.57, *p*=0.12).

Trends in efficacy

Based on random-effects meta-analysis, as expected (Undurraga and Baldessarini, 2012), there was an overall, highly significant, drug-placebo difference in improvement of depression ratings across the 56 trials of 14 antidepressant drugs. Pooled SMD (95% CI) was 0.372 ([0.328–0.415]; *z*=16.8, *p*<0.0001). We also confirmed a strong decline in effect-size (SMD) over years of reporting (slope [*β*]=−0.014 [CI: −0.019 to 0.010]; *t*=6.55, *p*<0.0001, Table 1*a*). This decline was associated with little change in antidepressant-associated responses over the era (*β*=+0.008 [−0.002 to +0.017], *t*=1.65, *p*=0.10), but with major and consistent increases in responses in placebo-arms (*β*=+0.024 [+0.014 to +0.034], *t*=4.81, *p*<0.0001). These trends were similar and significant across drug-types (TCAs, SRIs and SNRIs), with decreasing effect sizes, stable drug-associated effects and increasing placebo-associated effects over time (data not shown). Effects of other factors associated with SMD are summarized in Table 1*b*: the most strongly associated factor was placebo dropout for lack of efficacy, and later reporting years had significantly lower SMD, whereas trial duration and size were not significantly associated with effect-size.

Trends in dropout rates

In addition to increases in trial-size and duration over the years, there were particularly large decreases in dropout rates over years (Table 1*a*). Dropout rates for all causes declined highly significantly across reporting years (overall slope function, *β*=−1.23 [95%CI: −1.56 to −0.897]; *t*=7.41, *p*<0.0001). Dropout rates in the first decade averaged 39.8±11.1% with antidepressants, and 47.7±13.2% with placebo, compared to 24.3±8.27% and 24.0±9.04%, respectively, in the final decade sampled. These rates
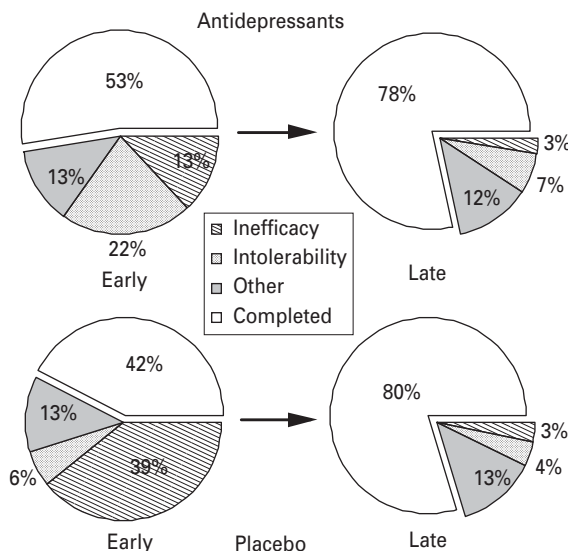
Fig. 1. Pie-chart distribution of outcomes in placebo-controlled trials of antidepressants for antidepressant arms (upper charts) or placebo arms (lower charts), based on linear regression modelling for all 56 trials, and reporting estimated values for the early (1985) and late (2009) observed years. Outcomes: *completed trials* as scheduled (white, separated sections), dropout for intolerability (textured sections), dropout for apparent lack of efficacy (striped sections), dropout for other reasons (grey sections), with corresponding mean rates for representative years: Early (left side) *vs.* Late (right side) indicated as percentages.

declined significantly and similarly in antidepressant arms ($\beta=-1.01$ [−1.36 to −0.652]), but even more in placebo-arms ($\beta=-1.54$ [−1.95 to −1.12]) of trials. The proportion of dropouts from placebo-arms was correlated with that from the drug-arm ($\beta=1.03$ [95%CI: 0.81 to −1.26]; $t=9.31$, $p<0.0001$). However, dropout rates varied markedly between drug and placebo arms by cited causes, including lack of efficacy, intolerability or adverse effects and miscellaneous other reasons (Fig. 1). In antidepressant-arms, dropouts for intolerability declined most (slope [β]=−0.630 [−0.909 to −0.352], and less for lack of efficacy ($\beta=-0.434$ [CI: −0.642 to −0.236]), both $p<0.0001$, whereas dropouts for other reasons remained stable ($\beta=-0.014$ [−0.243 to +0.216]) over the years. In contrast, in placebo-arms, dropouts for lack of efficacy declined markedly over years ($\beta=-1.50$ [−1.88 to −0.1.11]; $p<0.0001$), whereas dropouts for apparent intolerability declined non-significantly ($\beta=-0.081$ [−0.192 to +0.030]), and dropouts for other reasons remained stable over years ($\beta=+0.019$ [−0.189 to +0.228]).

Relationship of dropout rate to antidepressant efficacy

We used meta-regression modelling to test for association of particular aspects of dropout rates with effect-size (SMD). Dropout in placebo-arms associated with lack of efficacy was highly significantly and independently associated with effect-size, controlling for year, whereas placebo-dropout associated with intolerability, as well as antidepressant-associated dropout ascribed to either intolerability or lack of efficacy were not independently associated with effect-size (Table 2*a*). Moreover, dropout in placebo arms for apparent lack of efficacy remained highly significantly associated with effect-size, controlling for other factors, including reporting year, trial-size and nominal trial duration (Table 2*b*).

Bivariate relationships of placebo-dropout rates ascribed to lack of efficacy with reporting year and with effect-size are illustrated in Fig. 2. The placebo-dropout rates for inefficacy were strongly associated with effect-size in that lower dropout rates for inefficacy were associated with much smaller drug-placebo contrasts as reflected in lower SMD values (Fig. 2*a*), and placebo dropouts declined markedly over time (Fig. 2*b*).

Secondary analyses

We also analysed outcomes of 21 placebo-controlled RCTs that reported outcomes based on a completer analysis as well as an ITT analysis with LOCF imputing. Based on ITT analysis, there was a significant decrease of effect-size (SMD) over the years ($\beta=-0.009$ [−0.018 to 0.001], $r=0.22$, $t=2.34$, $p=0.030$, Fig. 3*c*), corresponding to a significant increase in placebo-associated improvement (SMI-placebo; $\beta=+0.023$; [0.011 to 0.036]; $p=0.001$; Fig. 3*a*). Again, improvements associated with antidepressant treatment did not change over time. In outcomes based on completer analysis from the same 21 trials, no significant trends over years were seen for effect-size ($p=0.187$, Fig. 3*d*), placebo-associated improvements ($p=0.264$; Fig. 3*b*), or antidepressant-associated improvements ($p=0.636$). Moreover, in this subgroup of trials, ITT based analysis showed a significant association between dropout for lack of efficacy from the placebo arm and SMD ($\beta=+0.007$; [0.001 to 0.014]; $p=0.043$), as described above. However, analysis based on completers did not show such an association between dropout from placebo arms for inefficacy and SMD ($\beta=+0.005$; [−0.005 to 0.015]; $p=0.321$).

Egger's test for publication bias indicated some small-study effects, although the bias coefficient was not significant (1.21 [CI: −0.061 to 2.47]; $t=1.91$, $p=0.062$). We also tested for trends of reporting dropout rates and related causes, but the proportion of reports including data on dropout rates ($n=56$) among 106 trial reports previously included in a review (Undurraga and Baldessarini, 2012) did not change over the years considered ($\beta=+0.0008$ [−0.009 to +0.010]).

Discussion

We assessed selected factors that might contribute to the marked decline in apparent efficacy found in meta-analyses of antidepressant trials in recent decades,

Table 2. Meta-regression analyses of factors associated with efficacy of antidepressants

a. Association of reasons for dropouts with SMD, controlling for year (*n*=56)

Factors	Slope (*β*) [95% CI]	*t*-score	*p*-value
Antidepressant dropout rate (%)			
Lack of efficacy	−0.007 [−0.018 to +0.003]	1.43	0.15
Intolerability	−0.001 [−0.007 to +0.005]	0.32	0.75
Other	−0.004 [−0.014 to +0.005]	0.92	0.36
Placebo dropout rate (%)			
Lack of efficacy	+0.009 [+0.004 to +0.015]	3.24	0.002
Intolerability	−0.003 [−0.018 to +0.012]	0.35	0.73
Other	+0.004 [−0.005 to +0.013]	0.84	0.40
Reporting year	−0.005 [−0.013 to +0.002]	1.48	0.14

b. Selected factors associated with SMD (*n*=56)

Factors	Slope (*β*) [95% CI]	*t*-score	*p*-value
Placebo dropout for inefficacy	+0.007 [+0.003 to +0.011]	3.32	0.002
Reporting year	−0.006 [−0.014 to +0.001]	1.81	0.08
Trial duration (weeks)	+0.010 [−0.013 to +0.032]	0.90	0.37
Subjects/trial (*N*)	+0.00003 [+0.00017 to +0.0002]	0.34	0.74

guided by findings in previous reports (Walsh et al., 2002; Khan et al., 2004, 2010; Undurraga and Baldessarini, 2012). Outcome was based on standardized mean difference (SMD) in improvements of depression ratings in antidepressant *vs.* placebo arms in peer-reviewed, randomized, placebo-controlled antidepressant trials reported over the past three decades. Like effect-size, dropout rates have declined strikingly over the era considered (Figs. 1 and 2). This decrease was significant in both drug- and placebo-arms, but the decrease in placebo-dropout per year was more than twice greater and was selectively and robustly associated with declining efficacy (SMD) in the trials examined. Furthermore, of all reasons for leaving trials prematurely, only declining dropout for lack of efficacy of placebo-treatment was independently and significantly associated with decreasing SMD.

The observed major decline of dropout rates in antidepressant RCTs is potentially important. Losses of over 20% of subjects pose serious threats to the validity of results obtained in trials (Schulz and Grimes, 2002), and have exceeded 50% in trials of some psychotropic agents (Hutton et al., 2012). In the present trials, dropout rates in the first decade averaged 39.8±11.1% with antidepressants, and 47.7±13.2% with placebo, compared to 24.3±8.27 and 24.0±9.04%, respectively, in the final decade sampled. Dropout rates have been associated with various characteristics of trials, including their design (target drug *vs.* active comparator or *vs.* placebo), type of drug, geographic area or culture and year of publication (Vázquez et al., 2011; Kemmler et al., 2005).

The major decrease over years of dropout rates in both placebo- and drug-arms of antidepressant RCTs is not readily explained. We did not find evidence that the
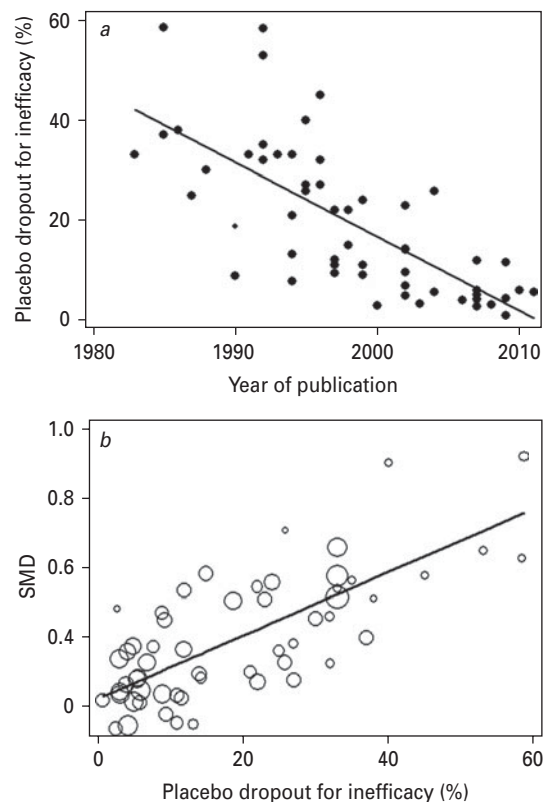

Fig. 2. Correlations: (*a*) Placebo dropout rate (%) ascribed to lack of efficacy *vs.* year of reporting, based on linear regression (*β*=−1.50 [−1.88 to −1.11]; *r*=0.728, *t*=7.83, *p*<0.0001); (*b*) SMD *vs.* placebo dropout for lack of efficacy based on bivariate meta-regression (*β*=+0.009 [0.007 to 0.012]; *t*=7.11, *p*<0.0001), with symbol-size reflecting the weight of each study based on trial-size (*N*) and standard error of mean effect-size (SMD as Hedges' *g*-statistic).
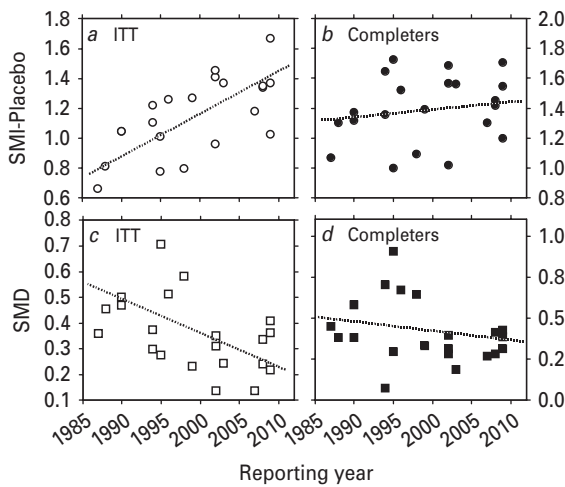
Fig. 3. Correlations: (*a*) Placebo standardized mean improvement (SMI-Placebo; mean/SD) *vs.* year of reporting, based on *ITT* analyses (*β*=−0.008 [0.011 to 0.036], *r*=0.66, *t*= 3.83, *p*=0.001). (*b*) Placebo standardized mean improvement (SMI-Placebo) *vs.* year of reporting, based on *completer* analyses (*β*=0.023 [−0.006 to 0.022]; *r*=0.26, *t*=1.15, *p*=0.30). (*c*) SMD over time based on *completer* analyses (*β*=−0.008 [−0.020 to 0.004], *r*=0.09, *t*=1.4, *p*=0.20). D. SMD over time based on *ITT* analyses (*β*=−0.009 [−0.018 to 0.001], *r*=0.22, *t*=2.34, *p*=0.030).

apparent nature or severity of the depressive illnesses studied changed appreciably over the decades. Notably, there was little change in the probability of having had previous episodes of depression, in the estimated duration of the index depressive episodes, or in initial severity of symptom ratings at trial entry. Indeed, baseline depression ratings as well as reported demographic characteristics (sex, age) of subjects have remained stable and well matched in drug and placebo arms of antidepressant trials since 1980. Nevertheless, most trial reports provide very limited clinical information about characteristics of patients or their illnesses, leaving open the possibility that some changes over years in patient or illness characteristics were not identified or not reported. The decrease in dropouts associated with adverse effects in drug-arms may plausibly be ascribed to increased tolerability of most modern antidepressants compared to older agents (Baldessarini, 2013). However, we identified a within-drug (paroxetine and fluoxetine) decline of dropouts for intolerability over reporting years (data not shown), suggesting that mechanisms other than pharmacological actions were involved. Longer retention in both drug- and placebo-arms may reflect increasing efforts to limit missing data by encouraging subjects to remain in trials (Amery and Dony, 1975; Murray and Findlay, 1988; Streiner, 2002; Wittes, 2009). Dropouts also may be affected by the growing shift from large trials based in multiple, and often dissimilar, academic clinics to contract-organization directed trials based in clinician offices, typically with relatively small numbers of subjects-per-site (Hecker et al., 2003).

Such trends may be reflected in marked increases in subjects/trial and sites/trial in modern trials for antidepressants and other psychotropic agents [Undurraga and Baldessarini, 2012; Yildiz et al., 2011a, b). We also found that subjects/site declined markedly in the present trials sample. Smaller samples may increase individual attention and encouragement to remain in trials.

The observed, independent association of placebo dropout, especially for lack of efficacy, with SMD needs to be addressed. Ideally, dropouts from trials should be absent if fair and unbiased conclusions are to be drawn; however, in most clinical trials, premature dropouts and missing data are virtually unavoidable (Wittes, 2009). To deal with missing data because of dropouts, many trials employ the final observation as the endpoint for each subject (last observation carried forward [LOCF] method of endpoint-analysis). The LOCF imputation method assumes that dropout is not related to treatment or to outcome, and that a subject who discontinues treatment would have retained constant clinical status from the time of dropout to the planned endpoint (Wittes, 2009). This assumption is highly improbable with depression, especially for dropouts due to lack of efficacy, which are clearly related to symptom severity. Placebo-treated patients who drop out for lack of perceived benefit may be more depressed, with higher symptom rating scores at the time of dropout than drug-treated subjects. In contrast, changes in dropout from drug-arms (mainly for intolerance) appeared to be less associated with SMD, was ascribed largely to adverse effects or drug-intolerance and were less associated with SMD than were declining dropouts with placebo, mostly for lack of benefit. It would be of interest to know the relative latency to dropout with placebo *vs.* active drug-treatment. However, trial reports rarely provided precise retention times in each trial arm. Dropout for intolerance may occur across all illness severities, whereas dropout after randomization to placebo is more likely with greater illness-severity; such differences might contribute to greater effects of longer retention in placebo-arms on reported effect-size. Furthermore, intolerance associated with antidepressant treatment might compromise blinding and enhance expectations of treatment and so encourage reporting of lower symptom severity in antidepressant arms (Greenberg et al., 1994; Baethge et al., 2013).

Subjects in antidepressant trials are, on average, likely to improve spontaneously over time, due to the natural, time-limited course of acute depressive episodes, with regression-to-the-mean effects that would tend to diminish drug-placebo contrasts over time. Therefore, dropouts from placebo-arms for lack of efficacy may lead to underestimating placebo-associated responses, whereas dropouts for intolerance may have less influence on drug-associated responses. We propose that such disparities in reasons for dropping out may result in bias that can lead to the impression that drug-placebo differences are

greater than they would be if subjects were consistently observed to the planned trial endpoint. Over the past three decades a decrease in rates of placebo-dropouts for lack of efficacy in antidepressant trials would correspond with a decline in carrying forward relatively high symptom scores as a surrogate outcome measure following dropout. Such temporal trends may have contributed to an increase of mean placebo-associated responses and decline in drug-placebo differences, as we found. Such effects are particularly likely when trial outcomes data are based on ITT analysis employing LOCF as a way of attempting to deal with missing values (Hennen, 2003).

In a secondary analysis, we compared trends of effect-size based on ITT analysis (based on LOCF) *vs.* completer analysis in studies reporting both analyses (Fig. 3). We found a significant increase in measured improvements in placebo arms of these trials using ITT-based data, but not when completer data were used. That is, ITT-based analyses selectively yielded declining effect-size over the years. In addition, dropout for lack of efficacy was not associated with completer-based placebo-associated responses, whereas in the same set of trials it was significantly associated with ITT-based placebo-associated responses. These considerations indicate that changes in effect-size over time in antidepressant–placebo comparisons depend on the methods of analysis used. Although the proportion of completers in the placebo arms of trials increased (Fig. 1), on average they did not show significantly greater improvement over the past three decades (Fig. 3b). When missing data were carried forward in LOCF-based ITT analyses, subjects randomized to placebo-treatment did seem to do better over the past three decades, whereas in completer-based analyses they did not, and their outcomes approached the completer-based mean (Fig. 3).

We do not suggest that completer analysis is a superior method of analysis of trials data, but considered it to compare trends when carrying forward of missing data was allowed or not. Limitations of completer analyses include loss of power and probable erosion of initial randomization, with potentially unrepresentative samples of remaining subjects who complete trials (Tierney and Stewart, 2005). Although ITT analysis has often been preferred (Tierney and Stewart, 2005), LOCF methods are far from an ideal way to manage missing data, because, as found in the present analysis, the assumption that patients remain clinically stable from the moment of dropout to the trial protocol endpoint usually is not met, resulting in risk of imputing greater morbidity after earlier dropout. In order to overcome this problem, other likelihood-based analytical methods to account for missing data such as mixed-effect modelling of repeated measures (MMRM) have been recommended, or subjects have been followed-up to assess outcomes even after they have dropped out of trials (Moher et al., 2001; Hennen, 2003; Lane, 2008; Rabinowitz and Davidov, 2008;

Siddiqui et al., 2009; Little et al., 2012). However, such methods have become even partially accepted only recently, and many meta-analyses of antidepressant efficacy are based on trials using ITT analysis based on LOCF methods and this choice may lead to identifying trends of declining efficacy and overestimating pooled effect-sizes or drug-placebo contrasts.

This study has several noteworthy limitations. One is that we considered findings in peer-reviewed, archival reports. Although we did not find evidence of publication bias, this selection option may under-represent trials with relatively low drug-placebo contrasts and high placebo-associated responses (Turner et al., 2008; Baldessarini, 2013). However, obtaining unpublished data from anti-depressant trials completed as long ago as the 1980s is virtually impossible and might well risk introducing time-based bias by selective inclusion of unpublished data from recent years. In addition, reports that did not provide dropout rates were not included. This limitation led to exclusion of 27.9% of otherwise eligible trials, although there was no change over reporting years in the proportion of trials reporting on dropouts within the present sample.

A major limitation that confounds efforts to analyse findings from controlled trials generally is that most reports do not provide specific information concerning individual or average retention times in each trial-arm and their association with morbidity ratings, except at endpoints arising from ITT and LOCF methods of analysis. This paucity of information limited our ability to evaluate effects of dropout based on time-in-trial so as to refine interpretation of declining dropout rates. This lack of information also limited our ability to include or conduct MMRM based analyses, for comparison with LOCF based analyses. Also, some trials were included that involved small proportions of bipolar depressed patients, who may be less responsive to antidepressants (Pacchiarotti et al., 2013). However, the proportion of subjects (0.2%) with bipolar depression was small and unlikely to bias the present analyses. Finally, we did not consider all factors previously associated with declining effect-sizes observed on antidepressant trials as reviewed above (Introduction), because of uneven reporting of such factors and limited statistical power with which to test for independent association of effect-size with a large number of potentially relevant factors.

In conclusion, among various factors that may impact measurements of antidepressant efficacy in randomized, placebo-controlled trials, decreases in dropout rates associated with placebo-arms of trials were particularly notable and probably contributed to a decrease of drug-placebo contrasts (effect-size) over the years since 1980. We propose that lowering dropout rates in placebo-arms of trials, specifically for lack of efficacy, tends to reduce drug-placebo differences found with ITT- and LOCF-based analyses that might be avoided by more routine use of alternative, likelihood-based, outcome analyses

such as mixed-effect modelling. Decreases in dropout rates probably have contributed to the decline in apparent efficacy of antidepressants over the years, as identified in this and previous meta-analyses. In short, placebo dropout rates complicate meta-analyses of trials and should be considered among factors influencing estimates of effect-size in future reviews.

Disclosures

None.

Supplementary material

For supplementary material accompanying this paper, visit http://dx.doi.org/10.1017/S1461145714000224

References

Amery W, Dony J (1975) A clinical trial design avoiding undue placebo treatment. J Clin Pharmacol 15:674–679.

Baethge C, Assall OP, Baldessarini RJ (2013) Systematic review of blinding assessment in randomized controlled trials in schizophrenia and affective disorders 2000–2010. Psychother Psychosom 82:152–160.

Baldessarini RJ (2013) Chemotherapy in Psychiatry, 3rd edn. New York: Springer.

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR (2009) Effect size and precision. Introduction to Meta-Analysis. New York: John Wiley.

Caldwell DM, Ades AE, Higgins JP (2005) Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ 331:897–900.

Egger M, Davey Smith G, Schneider M, Minder C (1997) Bias in meta-analysis detected by a simple, graphical test. BMJ 315:629–634.

Furukawa TA, Barbui C, Cipriani A, Brambilla P, Watanabe N (2006) Imputing missing standard deviations in meta-analyses can provide accurate results. J Clin Epidemiol 59:7–10.

Gomeni R, Lavergne A, Merlo-Pich E (2009) Modelling placebo response in depression trials using a longitudinal model with informative dropout. Eur J Pharm Sci 36:4–10.

Greenberg RP, Bornstein RF, Zborowski MJ, Fisher S, Greenberg MD (1994) Meta-analysis of fluoxetine outcome in the treatment of depression. J Nerv Ment Dis 182:547–551.

Hamilton M (1967) Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 6:278–296.

Hecker SJ, Preston C, Foote M (2003) Production of high-quality marketing applications: strategies for biotechnology companies working with contract research organizations. Biotechnol Annu Rev 9:269–277.

Hedges LV (1981) Distribution theory for Glass's estimator of effect size and related estimators. J Educ Behav Stat 6:107–128.

Henkel V, Casaulta F, Seemuller F, Krahenbuhl S, Obermeier M, Husler J, Moller HJ (2012) Study design features affecting outcome in antidepressant trials. J Affect Disord 141:160–167.

Hennen J (2003) Statistical methods for longitudinal research on bipolar disorders. Bipolar Disord 5:156–168.

Hutton P, Morrison AP, Yung AR, Taylor PJ, French P, Dunn G (2012) Effects of dropout on efficacy estimates in five Cochrane reviews of popular antipsychotics for schizophrenia. Acta Psychiatr Scand 126:1–11.

Iovieno N, Tedeschini E, Levkovitz Y, Ameral VE, Papakostas GI (2011) Does the frequency of follow-up assessments affect clinical trial outcome? Meta-analysis and meta-regression of placebo-controlled randomized trials. Int J Neuropsychopharmacol 14:1–8.

Kemmler G, Hummer M, Widschwendter C, Fleischhacker WW (2005) Dropout rates in placebo-controlled and active-control clinical trials of antipsychotic drugs: a meta-analysis. Arch Gen Psychiatry 62:1305–1312.

Khan A, Leventhal RM, Khan SR, Brown WA (2002) Severity of depression and response to antidepressants and placebo: analysis of the Food and Drug Administration database. J Clin Psychopharmacol 22:40–45.

Khan A, Detke M, Khan SR, Mallinckrodt C (2003a) Placebo response and antidepressant clinical trial outcome. J Nerv Ment Dis 191:211–218.

Khan A, Khan SR, Walens G, Kolts R, Giller EL (2003b) Frequency of positive studies among fixed and flexible dose antidepressant clinical trials: analysis of the Food and Drug Administration summary basis of approval reports. Neuropsychopharmacology 28:552–557.

Khan A, Kolts RL, Thase ME, Krishnan KR, Brown W (2004) Research design features and patient characteristics associated with the outcome of antidepressant clinical trials. Am J Psychiatry 161:2045–2049.

Khan A, Schwartz K, Kolts RL, Ridgway D, Lineberry C (2007) Relationship between depression severity entry criteria and antidepressant clinical trial outcomes. Biol Psychiatry 62:65–71.

Khan A, Bhat A, Kolts R, Thase ME, Brown W (2010) Why has the antidepressant-placebo difference in antidepressant clinical trials diminished over the past three decades? CNS Neurosci Ther 16:217–226.

Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT (2008) Initial severity and antidepressant benefits: meta-analysis of data submitted to the Food and Drug Administration. PLoS Med 5:e45.

Lane P (2008) Handling drop-out in longitudinal clinical trials: a comparison of the LOCF and MMRM approaches. Pharm Stat 7:93–106.

Lee S, Walker JR, Jakul L, Sexton K (2004) Does elimination of placebo responders in a placebo run-in increase the treatment effect in randomized clinical trials? A meta-analytic evaluation. Depress Anxiety 19:10–19.

Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, Frangakis C, Hogan JW, Molenberghs G, Murphy SA, Neaton JD, Rotnitzky A, Scharfstein D, Shih WJ, Siegel JP, Stern H (2012) Prevention and treatment of missing data in clinical trials. N Engl J Med 367:1355–1360.

Moher D, Schulz KF, Altman DG (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. BMC Med Res Methodol 1:2.

Montgomery SA, Åsberg M (1979) A new depression scale designed to be sensitive to change. Br J Psychiatry 134:382–389.

Murray GD, Findlay JG (1988) Correcting for the bias caused by drop-outs in hypertension trials. Stat Med 7:941–946.

Pacchiarotti I et al. (2013) International Society for Bipolar Disorders (ISBD) task-force report on antidepressant use in bipolar disorders. Am J Psychiatry 170:1249–1262.

Papakostas GI, Fava M (2009) Does the probability of receiving placebo influence clinical trial outcome? Meta-regression of double-blind, randomized clinical trials in major depressive disorder. Eur Neuropsychopharmacol 19:34–40.

Rabinowitz J, Davidov O (2008) The association of dropout and outcome in trials of antipsychotic medication and its implications for dealing with missing data. Schizophrenia Bull 34:286–291.

Schulz KF, Grimes DA (2002) Sample size slippages in randomized trials: exclusions and the lost and wayward. Lancet 359:781–785.

Siddiqui O, Hung HM, O'Neill R (2009) MMRM vs. LOCF: comprehensive comparison based on simulation study and 25 NDA datasets. J Biopharm Stat 19:227–246.

Streiner DL (2002) The case of the missing data: methods of dealing with dropouts and other research vagaries. Can J Psychiatry 47:68–75.

Tierney JF, Stewart LA (2005) Investigating patient exclusion bias in meta-analysis. Int J Epidemiol 34:79–87.

Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R (2008) Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med 358:252–560.

Undurraga J, Baldessarini RJ (2012) Randomized, placebo-controlled trials of antidepressants for acute major depression: thirty-year meta-analytic review. Neuropsychopharmacology 37:851–864.

Undurraga J, Tondo L, Schalkwijk S, Vieta E, Baldessarini R (2013) Re-analysis of the earliest controlled trials of imipramine. J Affect Disord 147:451–454.

Vázquez G, Baldessarini RJ, Yildiz A, Tamayo J, Tondo L, Salvatore P (2011) Multisite international collaborative clinical trials in mania. Int J Neuropsychopharmacol 14:1013–1016.

Walsh BT, Seidman SN, Sysko R, Gould M (2002) Placebo response in studies of major depression: variable, substantial, and growing. JAMA 287:1840–1847.

Wittes J (2009) Missing in action: preventing missing outcome data in randomized clinical trials. J Biopharm Stat 19:957–968.

Yildiz A, Vieta E, Leucht S, Baldessarini RJ (2011a) Efficacy of antimanic treatments: meta-analysis of randomized, controlled trials. Neuropsychopharmacology 36:375–389.

Yildiz A, Vieta E, Tohen M, Baldessarini RJ (2011b) Factors modifying drug and placebo responses in randomized trials for bipolar mania. Int J Neuropsychopharmacol 14:863–875.