Journal of Psychiatric Research 51 (2014) 21–29

Contents lists available at ScienceDirect

Journal of Psychiatric Research

journal homepage: www.elsevier.com/locate/psychires

Impact of patient selection and study characteristics on signal detection in placebo-controlled trials with antidepressants

Michele Mancini [a,*], Alan G. Wade [b], Giulio Perugi [c], Alan Lenox-Smith [d], Alexander Schacht [e]

[a] Eli Lilly Italia S.p.A., Neuroscience, Via A. Gramsci 731/733, 50019 Sesto Fiorentino, FI, Italy
[b] CPS Research, Glasgow, UK
[c] University of Pisa, Institute of Psychiatry, Pisa, Italy
[d] Lilly UK, Medical Department, Basingstoke, UK
[e] Lilly Deutschland GmbH, Global Biostatistics, Bad Homburg, Germany

ARTICLE INFO

Article history:
Received 20 August 2013
Received in revised form
6 December 2013
Accepted 3 January 2014

Keywords:
Effect size (ES)
Major depressive disorder (MDD)
Hamilton Depression Rating scale (HAM-D)
Placebo
Signal detection
Duloxetine

ABSTRACT

An increasing rate of antidepressant trials fail due to large placebo responses. This analysis aimed to identify variables influencing signal detection in clinical trials of major depressive disorder. Patient-level data of randomized patients with a duloxetine dose ≥60 mg/day were obtained from Lilly. Total scores of the Hamilton Depression Rating scale (HAM-D) were used as efficacy endpoints. In total, 4661 patients from 14 studies were included in the analysis. The overall effect size (ES), based on the HAM-D total score at endpoint, between duloxetine and placebo was −0.272. Although no statistically significant interactions were found, the following results for factors influencing ES were seen: a very low ES (−0.157) in patients in the lowest baseline HAM-D category and in patients recruited in the last category of the recruitment period (−0.122). A higher ES in patients recruited in centers with a site-size at but not more than 2.5 times the average site-size for the study (−0.345). Study characteristics that resulted in low signal detection in our database were: <80% study completers, a HAM-D placebo response >5 points, a high variability of placebo response (SD > 7 points HAM-D), >6 post baseline visits per study, and use of an active control drug. Simpler trial designs, more homogeneous and mid-sized study sites, a primary analysis based on a higher cutoff blinded to investigators to avoid the influence of score inflation in mild patients and, if possible, studies without an active control group could lead to a better signal detection of antidepressive efficacy.

1. Objectives of the study and background

Placebo-controlled studies of antidepressants often demonstrate a large placebo response resulting in relatively small effect size (ES) between placebo and active drug in patients suffering from major depressive disorder (MDD). This results in reduced signal detection, i.e. the failure to confirm a treatment effect. It has been shown that this effect has been increasing over time (Khin et al., 2011; Rutherford and Roose, 2013), and around 50% of the clinical trials with active antidepressants fail to show any significant difference from placebo (Khan et al., 2003a, 2003b; Gelenberg et al., 2008; Turner et al., 2008; Khin et al., 2011). The inevitable outcome is delay and increasing inefficiency in the development of new antidepressants. Several authors have tried to identify factors and patient characteristics as predictors for increasing signal detection in clinical trials with antidepressants. Khan et al. (2004) noted that the severity of depressive symptoms at baseline, flexible dosing, fewer treatment arms, and a lower percentage of female patients were significantly associated with successful trials. Khin et al. (2011) confirmed the decline of treatment effects in MDD studies over a 25 years period, suggesting that the baseline disease severity of patients included in clinical trials seems to be a more important factor in study outcome than other factors. Kirsch et al. (2008) and Fournier et al. (2010) found that the magnitude of benefit of antidepressant medication compared with placebo increases with the

* Corresponding author. Tel.: +39 0554257534, +39 3454330339 (mobile); fax: +39 0554257348.
E-mail address: mancini_michele@lilly.com (M. Mancini).

0022-3956
http://dx.doi.org/10.1016/j.jpsychires.2014.01.001

Table 1
Design characteristics of the 14 studies from the duloxetine integrated database included in the Effect Size analysis.

Study ID	Protocol title	Study phase	Year of protocol	Publications	Length of acute phase (weeks)[a]	Number of post-baseline visits	Placebo controlled	Active controlled	Number of treatment arms	Patients assigned to placebo (%)	Ratio of completers	Ratio of placebo responders[b]	Number of countries in study
HMAQA	Duloxetine Versus Placebo in the Treatment of Major Depression	II	1998	Goldstein et al., 2002	8	8	yes	yes	3	40.46	65.32	48.57	1
HMAQB	Duloxetine Versus Placebo in the Treatment of Major Depression	II	1998	Nemeroff et al., 2002, Mallinckrodt et al., 2003	8	8	yes	yes	3	38.66	63.92	44	1
HMATA	Duloxetine Versus Placebo and Paroxetine in the Acute Treatment of Major Depression	III	2000	Nemeroff et al., 2002 Mallinckrodt et al., 2003	8	5	yes	yes	4	25.42	68.64	43.33	1
HMATB	Duloxetine Versus Placebo and Paroxetine in the Acute Treatment of Major Depression	III	2000	Goldstein et al., 2004	8	5	yes	yes	4	25.21	59.21	42.7	1
HMAYA	Duloxetine Versus Placebo and Paroxetine in the Treatment of Major Depression	III	2000	Detke et al., 2004	8	5	yes	yes	4	25.34	86.65	69.89	6
HMAYB	Duloxetine Versus Placebo and Paroxetine in the Treatment of Major Depression	III	2000	Perahia et al., 2006	8	5	yes	yes	4	25.26	91.07	78.79	7
HMBHA	Duloxetine Once-Daily Dosing Versus Placebo in the Acute Treatment of Major Depression	III	2000	Detke et al., 2002a	9	6	yes	no	2	49.8	67.76	36.07	1
HMBHB	Duloxetine Once-Daily Dosing Versus Placebo in the Acute Treatment of Major Depression	III	2000	Detke et al., 2002b	9	6	yes	no	2	52.06	62.92	48.2	1
HMBV	Duloxetine Versus Placebo in the Treatment of Elderly Patients With Major Depressive Disorder	IV	2002	Raskin et al., 2008a, Wise et al., 2007, Raskin et al., 2008b, Raskin et al. 2007 [c]	8	4	yes	no	2	33.44	77.81	35.58	2
HMCB	Duloxetine Once-Daily Dosing Versus Placebo in Patients with Major Depression and Pain	IIIb	2001	Brannan et al., 2005	7	5	yes	no	2	50	65.6	51.77	1
HMCR	Duloxetine Versus Escitalopram and Placebo in the Treatment of Patients With Major Depression	IIIb	2003	Nierenberg et al., 2007	8	6	yes	yes	3	20.03	71.93	46.72	1
HMFA	Cymbalta Vs Placebo in Long-Term Treatment Late-Life MDD	IV	2006	Robinson et al., 2014	12	5	yes	no	2	32.7	69.46	55.37	4
HMFS	MDD efficacy in depressive sympt improvements and usual fx	IV	2007	Oakes et al., 2012	36	11	yes	no	2	33.03	71.89	63.07	2
HQAC	Validation of Daily Telephone Self-Assessment in the Study of Antidepressant Treatment Outcome	II	2002	Mundt et al., 2007b	4	4	yes	no	3	50	94.29	42.86	1

[a] Without placebo lead-in phase.
[b] Response was defined as 30% reduction from baseline (LOCF) on the HAM-D 17 total score at endpoint.
[c] HMBV primary manuscript.

severity of symptoms at baseline. It is not clear, however, if the increasing difference between the effect of antidepressant drugs and placebo in more severe patients is due to an increasing efficacy of the medication or to a declining placebo effect (Fournier et al., 2010), or to a methodological bias underlying such findings (Höschl, 2008; Turner and Rosenthal, 2008; Fountoulakis and Möeller, 2011; Schacht et al., 2013). Furthermore, various approaches have been used to try to identify the reasons for these unduly large placebo responses, which have increased further in more recent studies with several antidepressants (Khan et al., 2010).

Apart from the fact that patients selected for clinical trials suffering from psychiatric disorders, such as MDD, are generally prone to large placebo effects, many aspects of the design of antidepressant trials may have a significant impact on treatment outcomes; thus in recent years, discussion has focused on the design features of current MDD studies that might artificially inflate placebo responses. For example, Mundt et al. (2007a) pointed out that, further to the strong placebo response, methodological factors such as patient selection and competitive enrollment targets, as well as a tendency to inflate baseline scores to enroll patients, contribute to increasing rates of failed antidepressant trials. Therefore, other authors have tried to identify methods to improve signal detection (Gelenberg et al., 2008). For example, Mallinckrodt et al. (2007) suggested that the use of mixed model repeated measure (MMRM) analysis and some Hamilton Depression Rating scale (HAM-D) subscales among other factors could minimize the high rate of false positive and false negative findings in trials with antidepressants. Very recently, Gueorguieva et al. (2011) have provided inspiring and convincing proof that some patients receiving medication during an antidepressant clinical trial do even worse than patients receiving placebo. The hidden bipolarity of some patients included in MDD clinical trials, or early discontinuation from the drug in some enrolled patients, could partially explain such findings. Furthermore, these authors challenged the belief that placebo response is associated predominantly with rapid and transient clinical improvement, proving instead that placebo response differences are a dimensional rather than a categorical characteristic. Based on these unresolved questions and on the extensive debate present in the literature and among clinicians, the aim of the presented analysis was to comprehensively evaluate design aspects of all acute placebo-controlled studies in the duloxetine integrated database and their influence on ES compared to placebo. Study design features, including the region where the study was performed, patient characteristics, and technical aspects of study management, such as the time-point of recruitment relative to the study progress (i.e. early or late during the study), size of the study site (i.e. the number of patients recruited), and the variability of patient parameters at baseline were evaluated.

This analysis should enable us to identify specific parameters that could lead to a smaller placebo response, a larger ES, better signal detection, and consequently to a higher success rate for placebo-controlled clinical trials with antidepressants.

2. Materials and methods

This analysis was based on individual patient data from a database including all patients participating in clinical trials with duloxetine used as an antidepressant performed by Eli Lilly & Co. All studies were performed according to the appropriate version of the Declaration of Helsinki and the study designs were approved by the appropriate ethics committees. All participants provided informed consent after being informed about the nature of the procedures of the respective study. Results of the individual studies were previously reported as detailed in Table 1. From this overall database, studies were included in the present meta-analysis if they met the following criteria:

- Studies investigating efficacy and safety of duloxetine in MDD
- Randomized controlled studies (placebo with or without active comparator control)
- Studies conducted in patients with acute depressive symptoms (i.e. data from patients in relapse prevention studies were excluded)
- At least 1 duloxetine arm with a dose $\geq$60 mg/day
- Studies with symptoms measured using HAM-D as endpoint.

Table 1 presents the list of studies included in the analysis with their key study design features and literature references.

2.1. Statistical analyses

The analysis population used for the statistical analyses comprised all randomized patients excluding patients in duloxetine treatment arms that received a dose <60 mg/day. This population will be referred to as all randomized patients.

Study overview and study design parameters were summarized by study. Size of sites (i.e. the number of patients) was analyzed as absolute size and relative size (compared to the median site size of all sites within the study) by region, by study, and by country. In addition, patient demographic, disease, pre-treatment, and baseline characteristics were analyzed. All analyses for study and patient characteristics were done using descriptive statistics and were based on all randomized patients.

Efficacy endpoints were analyzed using an Analysis of Covariance (ANCOVA; last observation carried forward [LOCF]) based meta-analysis or with meta-regression methodology using information on a study-level and on site-level. The population used was all randomized patients without missing data. All p-values and confidence intervals are unadjusted for multiplicity and cannot be considered confirmatory.

The ANCOVA model was calculated for each group, including treatment, patient characteristics, and their interaction as fixed effects, and the baseline value of the efficacy endpoint as covariate. ES in each model was calculated for least square mean (LSMEAN) differences, divided by the standard deviation (SD) of the residuals provided by the model of this group. Overall LSMEAN estimates and ES were calculated as a weighted mean of the corresponding estimates within study, with weights based on study variance, assuming a fixed study effect.

The following patient level parameters were included in the ANCOVA meta-analysis models: HAM-D total score as efficacy endpoint and the following study and study site characteristics:

- Absolute size of site (number of patients per site)
- Relative size of site (number of patients per site divided by average number of patients per site in the respective study)
- Time since recruitment started (<50 days, 50–100 days, 100–150 days, >150 days)
- Psychiatrist/psychologist (yes/no, missing = no)
- Number of studies per site (1/>1)
- Academic site (yes/no)
- Cluster score measuring investigator enrollment history in previous trials taken from a clinical operations database.

Two additional ANCOVAs were calculated, with treatment and the baseline value of the efficacy endpoint as covariates, whereby one model was used per study (model 1) or per site (model 2). The resulting LSMEAN difference estimates were used as a basis for the meta-regression models. For each study/site characteristic of interest

and group comparison, a meta-regression model was used. For these analyses, the standard errors of LSMEAN estimates, as provided by the basis ANCOVA model, were used as weights. The only exceptions from this are the analyses of size of sites, where the standard error correlates with the variable of interest and therefore equal weights were applied.

A simple meta-regression was used for all categorical variables. For all continuous variables, a meta-regression model with polynomials of the covariate of interested was performed, as well as a meta-regression for categorization of the covariate of interested. The results from the meta-regression on the continuous variable were further presented graphically, once for all group comparisons (not shown), and once, in more detail, for the placebo vs. duloxetine comparison.

The following study or study site parameters were included in the ANCOVA meta-regression analysis models (i.e. based on study or site level) on the HAM-D total score:

- Size of site
- Relative size of site
- Time to the start of recruitment
- Time to recruitment (within site)
- Number of studies per site (1/>1)
- Academic site (yes/no)
- Cluster score
- Proportion of placebo patients
- Proportion of completers
- Proportion of 30% HAM-D placebo responders
- Change in HAM-D total score in placebo group
- SD of change in HAM-D total score in placebo group
- Proportion of treatment-emergent adverse events in placebo group
- Number of post-baseline visits
- Active controlled study (yes/no)
- Year of protocol
- Number of post-baseline visits per week
- Titration of duloxetine at study start (yes/no)
- Placebo lead-in phase (yes/no)
- Number of countries in study
- Fixed vs. flexible duloxetine dose
- Number of treatment arms
- SD of HAM-D total score at baseline

3. Results

Overall, data of 4661 individual patients collected in 14 studies (see Table 1) were used for the present analysis: 2385 patients (51.2%) received duloxetine, 1573 (33.7%) placebo, and 703 (15.1%) active comparator. Of all patients included in the present analysis, the largest proportion (776 patients; 16.6%) came from study HMFS, followed by study HMCR (684 patients; 14.7%). All other studies contributed less than 10% of the patients to the present analysis.

Table 2
Patient demographics, disease characteristics, and site size of the patients from the duloxetine integrated database included in the Effect Size analysis.

Parameter	Duloxetine ($N = 2385$)	Active comparator ($N = 703$)	Placebo ($N = 1573$)	Total ($N = 4661$)
Patient demographics				
Age [years] mean (SD)	48.3 (16.39)	42.9 (12.40)	46.8 (16.15)	47.0 (15.87)
Sex [female] *n* (%)	1549 (64.9)	457 (65.0)	1020 (64.8)	3026 (64.9)
Ethnicity *n* (%)				
White	1904 (79.8)	590 (83.9)	1253 (79.7)	3747 (80.4)
Black	206 (8.6)	52 (7.4)	141 (9.0)	399 (8.6)
Other	275 (11.5)	61 (8.7)	179 (11.4)	515 (11.0)
Geographical region *n* (%)				
US	1887 (79.1)	520 (74.0)	1323 (84.1)	3730 (80.0)
Europe	350 (14.7)	159 (22.6)	175 (11.1)	684 (14.7)
Other	148 (6.2)	24 (3.4)	75 (4.8)	247 (5.3)
Disease characteristics[a]				
Patients with ≥1 pre-existing condition	1746 (73.2)	486 (69.1)	1116 (70.9)	3348 (71.8)
Number of previous episodes, median (IQR)	2 (1; 4)	2 (0; 4)	2 (0; 4)	2 (1; 4)
Duration of current episode, [months] median (IQR)	4.6 (1.8; 12.0)	4.4 (1.8; 12.0)	5.5 (2.0; 12.0)	4.6 (1.8; 12.0)
Patients with ≥1 previous medication for MDD, *n* (%)	1316 (55.2)	381 (54.2)	795 (50.5)	2492 (53.5)
CGI-S at baseline, mean (SD)	4.29 (0.68)	4.15 (0.71)	4.26 (0.69)	4.26 (0.69)
HAM-D 17 total score at baseline, mean (SD)	20.2 (4.99)	18.7 (4.98)	20.1 (4.92)	19.9 (4.99)
Median (IQR)	220 (17.0; 3.0)	19 (15.0; 22.0)	20.0 (17.0; 23.0)	20.0 (17.0; 23.0)

	US ($N = 3730$)	Europe ($N = 684$)	Other ($N = 247$)	Total ($N = 4661$)
Size of site (patients/site)	Patients, *n* (%)			
<10	565 (15.1)	30 (4.4)	47 (19.0)	642 (13.8)
10–14	667 (17.9)	119 (17.4)	50 (20.2)	836 (17.9)
15–19	607 (16.3)	146 (21.3)	38 (15.4)	791 (17.0)
20–24	598 (16.0)	150 (21.9)	46 (18.6)	794 (17.0)
25–29	564 (15.1)	25 (3.7)	0 (0.0)	589 (12.6)
≥30	729 (19.5)	214 (31.3)	66 (26.7)	1009 (21.6)
Mean (SD)	22.3 (14.59)	25.5 (14.66)	20.0 (9.62)	22.7 (14.43)
Relative size of site	Patients, *n* (%)			
<0.5	246 (6.6)	23 (3.4)	18 (7.3)	287 (6.2)
0.5–<1	903 (24.2)	243 (35.5)	67 (27.1)	1213 (26.0)
1–<1.5	1041 (27.9)	204 (29.8)	84 (34.0)	1329 (28.5)
1.5–<2	840 (22.5)	63 (9.2)	43 (17.4)	946 (20.3)
2–<2.5	312 (8.4)	0 (0.0)	0 (0.0)	312 (6.7)
2.5–<3	112 (3.0)	151 (22.1)	0 (0.0)	263 (5.6)
≥3	276 (7.4)	0 (0.0)	35 (14.2)	311 (6.7)
Mean (SD)	1.481 (0.90)	1.385 (0.77)	1.463 (0.94)	1.466 (0.89)

[a] Some variables were not collected in all studies; CGI-S: clinical global impression – severity; HAM-D: Hamilton depression scale; IQR: interquartile range (i.e. 25–75%); MDD: major depressive disorder; *N*: number of patients; *n*: number of patients with respective result; SD: standard deviation; US: United States of America.

The studies were initiated between 1998 and 2007, and 11 studies had an acute treatment duration of 7–9 weeks. The majority of patients were recruited in the United States of America (US).

Table 2 presents an overview of patient demographics by treatment group and disease characteristics, and the size of study sites for the complete dataset (i.e. across studies).

About 2/3rd of the patients were female. Patients had a median age of 45.9 years (interquartile range 34.8–57.3 years), 80.4% (3747) of the patients were White. Over 28% of the patients had a pre-existing condition of interest (i.e. linked to depression or in the same spectrum), with migraine (6.1%), anxiety (4.6%), fatigue (2.7%), irritable bowel syndrome (2.4%), and myalgia (2.0%) being the most common. All other pre-existing conditions of interest were reported by <2.0% of the patients.

More than 50% of the patients had received at least 1 previous medication indicated for MDD, with fluoxetine (17.4%), sertraline (16.1%), paroxetine (14.8%), venlafaxine (8.2%), citalopram (6.9%), and escitalopram (6.2%) being the most common. All other medications were taken by <5% of the patients.

Mean disease severity at baseline, measured by the HAM-D 17 total score, was 19.9 (SD: 4.99), with an interquartile range of 17–23.

The overall ES between duloxetine and placebo, based on the HAM-D total score at endpoint (LOCF), was −0.272 (i.e. in favor of duloxetine) for all patients, and −0.357 excluding patients from study sites with <10 patients.

3.1. Analysis of factors influencing ES differences

An overview of the influence of study design parameters on the ES, based on the HAM-D total score, is presented in Table 3a.

Absolute study site size showed inconsistent results regarding an impact on duloxetine vs. placebo HAM-D total score ES differences. For a graphical representation based on a meta-regression see Online Fig. 1a.

Relative study site size did not show a significant interaction with the relative ES between placebo and duloxetine. However, a pattern can be seen indicating that patients from sites with on average or just above the average relative site size within the study showed the highest ES. For a graphical representation see Online Fig. 1b.

Further, time to recruitment (from study start) did not show a significant interaction with the relative ES between placebo and duloxetine, as measured by the HAM-D total score. But the last (Q4) time to recruitment groups, showed a considerably smaller ES. For a graphical representation see Online Fig. 1c.

As shown in Table 3a, the category of patients having an HAM-D score closer to the inclusion cutoff of the studies showed the smallest ES of −0.157, whereas for all other baseline HAM-D categories the ES ranged from −0.279 to −0.333. The differences found between categories can be considered relevant in terms of signal detection.

An overview of the influence of study design parameters on a by-study level on treatment differences in the change from baseline reporting LSMEAN estimates and 95% CI is presented in Table 3b.

Table 3a
Analysis of the interaction of study design parameters with the relative effect size of duloxetine vs. placebo. A: Results from meta-analysis adjusted by Study − Each covariate analyzed separately.

(*N* = number of patients)	Comparison of duloxetine vs. placebo			
	LSMEAN difference	Effect size	95% CI	*p*-Value
HAM-D total score at endpoint (LOCF)				
Dulox vs. Placebo	−1.721	−0.357	(−2.13, −1.31)	<0.0001
Absolute size of site, i.e. number of patients per site				
<10 (*N* = 599)	−1.547	−0.226	(−2.70, −0.39)	0.0085
10–14 (*N* = 740)	−2.103	−0.332	(−3.10, −1.11)	<0.0001
15–19 (*N* = 665)	−1.696	−0.256	(−2.76, −0.63)	0.0018
20–24 (*N* = 628)	−2.035	−0.305	(−3.13, −0.94)	0.0003
25–29 (*N* = 507)	−1.252	−0.180	(−2.53, 0.02)	0.0539
≥30 (*N* = 819)	−1.908	−0.295	(−2.87, −0.95)	0.0001
Relative size of site, i.e. number of patients per site, compared to the median site size of all sites within the study				
Dulox vs. Placebo				
<0.5 (*N* = 246)	−1.794	−0.242	(−3.61, 0.02)	0.0528
0.5–<1 (*N* = 1014)	−1.312	−0.213	(−2.16, −0.47)	0.0023
1–<1.5 (*N* = 1092)	−2.208	−0.331	(−3.03, −1.38)	<0.0001
1.5–<2 (*N* = 825)	−2.257	−0.345	(−3.25, −1.27)	<0.0001
2–<2.5 (*N* = 305)	−2.242	−0.316	(−3.95, −0.54)	0.0100
2.5–<3 (*N* = 211)	−0.716	−0.133	(−2.47, 1.03)	0.4225
≥3 (*N* = 265)	−1.213	−0.173	(−2.96, 0.53)	0.1730
Time to patient recruitment at each site [days]				
Dulox vs. Placebo				
≤50 days (*N* = 2490)	−1.901	−0.275	(−2.48, −1.32)	<0.0001
>50–≤100 days (*N* = 617)	−2.465	−0.381	(−3.54, −1.39)	<0.0001
>100–≤150 days (*N* = 282)	−1.807	−0.300	(−3.35, −0.26)	0.0220
>150 days (*N* = 569)	−0.691	−0.122	(−1.79, 0.40)	0.2162
Baseline HAM-D score				
≤18; mild (*N* = 1379)	−1.012	−0.157	(−1.75; −0.27)	0.0072
19–21 mild to moderate (*N* = 1031)	−2.213	−0.333	(−3.07; −1.35)	<0.0001
22–24 moderate (*N* = 860)	−1.884	−0.279	(−2.84; −0.93)	0.0001
≥25 severe (*N* = 688)	−2.249	−0.330	(−3.39; −1.11)	0.0001
Baseline SDS total score				
≤median (*N* = 695)	−2.032	−0.309	(−3.09; −0.97)	0.0002
>median (*N* = 750)	−1.416	−0.218	(−2.43; −0.41)	0.0060

The unadjusted mean HAM-D score at endpoint for duloxetine (*N* = 2385) was 10.639 (95% CI: 10.39, 10.89), for active comparator (*N* = 703) 10.445 (95% CI: 10.04; 10.85), and for placebo (*N* = 1573) 12.799 (95% CI: 12.49, 13.11).
CI: confidence interval; HAM-D: Hamilton depression scale; LOCF: last observation carried forward; LSMEAN: least square mean; SDS: Sheehan Disability Scale.

Table 3b
Analysis of the interaction of study design parameters with the relative effect size of duloxetine vs. placebo. B: Results from meta-regression across Studies – Each covariate analyzed separately.

(N = number of studies)	LSMEAN estimate	Standard error	95% CI	p-Value
Proportion of patients treated with placebo within study				
Dulox vs. Placebo				
>20%–40% ($N = 9$)	−1.669	0.366	(−2.495, −0.902)	0.0006
>40% ($N = 5$)	−2.124	0.551	(−3.324, −0.924)	0.0023
Proportion of completers within study				
Dulox vs. Placebo				
0%–80% ($N = 11$)	−1.743	0.345	(−2.494, −0.991)	0.0003
≥80% ($N = 3$)	−2.148	0.664	(−3.595, −0.702)	0.0071
Mean HAM-D placebo response within study				
Dulox vs. Placebo				
≤5 points ($N = 6$)	−2.526	0.411	(−3.421; −1.630)	<0.0001
>5 points ($N = 8$)	−1.354	0.339	(−2.093; −0.616)	0.0018
SD in HAM-D in placebo group within study				
Dulox vs. Placebo				
≤7 points ($N = 9$)	−2.309	0.317	(−3.000; −1.617)	<0.0001
>7 points ($N = 5$)	−1.015	0.413	(−1.915; 0.114)	0.0304
Number of post-baseline visits within study				
Dulox vs. Placebo				
4 or 5 visits ($N = 8$)	−1.719	0.420	(−2.644; −0.793)	0.0018
6 visits ($N = 3$)	−2.283	0.656	(−3.726; −0.840)	0.0051
>6 visits ($N = 3$)	−1.624	0.686	(−3.134; −0.114)	0.0374
Active controlled study				
Dulox vs. Placebo				
No ($N = 7$)	−1.930	0.431	(−2.870; −0.991)	0.0008
Yes ($N = 7$)	−1.723	0.441	(−2.683; −0.762)	0.0021

CI: confidence interval; HAM-D: Hamilton depression scale; LSMEAN: least square mean; SD: standard deviation.

The proportion of patients who received placebo did not show a statistically significant interaction as measured by the difference of the change in HAM-D total score. However, studies with >40% patients randomized to placebo had an LS mean difference of 2.12 vs. 1.67 points in the HAM-D score when compared to studies with a lower proportion of patients allocated to placebo. For a graphical representation see Online Fig. 1d.

Further study characteristics that resulted in low signal detection in our database were: <80% study completers, a HAM-D placebo response >5 points, a high variability of placebo response (SD > 7 points HAM-D), >6 post baseline visits per study, and use of an active control drug vs. a placebo controlled study (see Table 3b).

4. Discussion

We could not find single study design parameters or parameters of study conduct that clearly and statistically significantly interacted with the relative ES between placebo and duloxetine as an active treatment for depression. However, we were able to identify a group of study characteristics that showed a trend towards higher vs. lower ES in the trials included in our database.

Significant p-values for interaction tests are only an indication for a variable able to change the effect. However, non-significant p-values do not imply that the studied variable has no impact on the treatment effect. Given the sample sizes, especially for the meta-regression analyses where the number of studies is a relevant factor (i.e. not only number of patients), the descriptive nature of the interaction should not be neglected. For example, an increase of the ES by 0.1 would substantially reduce the sample size in future studies. The absolute reduction in patient numbers needed to achieve a given power depends on the area in which the ES is. An ES increase from 0.3 to 0.4 has a stronger effect on the sample size than an increase from 0.6 to 0.7. If variables like study design can help to reduce the study size, even small increases in the ES are relevant as it leads to less patients being exposed to placebo, more effective use of research budget; and faster implementation of the study among other consequences of smaller sample size. Thus, for an exploratory analysis of design aspects and conduct of clinical studies, this level of evidence is sufficient to discuss possible implications for future studies in MDD.

Factors that may be associated with ES can be classified into one of 3 groups, i.e. related to 'study design', 'patient population', or 'study implementation'. Aspects of study design which may affect ES include the proportion of patients receiving placebo relative to active treatment, and the number of post-baseline visits. Aspects related to patient populations which may affect ES include the severity of patients' depression and the geographical location of the study sites as well as size of study site (i.e. the number of patients) and time to patient recruitment are implementation factors.

All study design factors associated directly or indirectly with a strong placebo response resulted in decreased signal detection. This was found for absolute and relative size of placebo response, high variability of the placebo response, and for study designs with low proportions of patients randomized to placebo (e.g. active placebo-controlled studies with a 1:2 ratio [placebo:active treatment] had smaller ES than studies with a 1:1 ratio). In a standard design using an active and placebo-controlled study, the percentage of patients randomized to placebo is usually 33%, leading to an increased placebo effect due to patient and physician expectations, i.e. because of the greater likelihood of receiving active treatment (duloxetine or active comparator) than placebo. This could be overcome by randomizing 50% of the patients to placebo and 25% to each active treatment in order to achieve the same proportion of patients receiving placebo as in a standard 2-arm placebo-controlled study. In addition, the possibility of receiving placebo should be emphasized, using percentages, in the patient information sheet. It is worth to mention that Papakostas and Fava (2009) found that the percentage of patients randomized to placebo was the factor having the strongest influence on signal detection in MDD trials. This was also confirmed by Mallinckrodt et al. (2010) for Schizophrenia trials, where the signal detection tended to improve and the mean placebo improvement decreased as the proportion of patients randomized to placebo increased. Furthermore, consistently with what we suggest, in the case of an active and placebo controlled trial the results of Mallinckrodt et al. also indicated that the probability of getting at least one significant drug-placebo contrast within a multi-arm study was maximized by randomizing more patients to placebo than to the individual active treatment arms. In this case the more convenient allocation ratio seems to be an equal allocation between the placebo arm and the two active drugs arms combined (e.g. 2:1:1).

Another factor that might be linked with increased placebo response, and thus reduced signal detection, was a high number of post-baseline visits in the study. A similar finding was also reported by Dunlop et al. (2012), in a meta-analysis of clinical studies with venlafaxine and desvenlafaxine. The number of post-baseline visits depends on the length of the study. For instance, a 12-month study would, by definition, include more visits than a 6-week study. Nevertheless, we recommend that in future studies, visits should be short as possible, with intervals between visits that are as long as possible, whilst being compatible with safety and reasonably achieving the aims of the study.

We also found that relative site size seems to be more important than absolute site size, as an optimum curve showed that very large sites, i.e. recruiting high numbers of patients, as well as very small sites, were associated with a smaller ES relative to sites recruiting moderate numbers of patients. At larger study sites, reasons for the

impact on ES may be related to advertising for recruitment, which may encourage participation of large numbers of patients, and high numbers of raters. Conversely, very small sites may not recruit enough patients to allow regular and stable ratings. Thus, the critical issue may be the level of recruitment relative to the potential patient base, and the number of patients per rater, i.e. a high enough number of patients to ensure standardization of the approach and low enough that the rater is not pressurized by time constraints.

A further important aspect of study implementation was time to patient recruitment, which showed that signal detection was reduced when patients were recruited late during the course of a study. This applied both to back-up study sites that were only added later during the course of the study, and to patients recruited late at the initial study sites. Both of these scenarios might be a result of the pressure to recruit patients fast, thus including patients and centers that are not well suited to the study. Of course, centers that are included late would also have less time to recruit, and therefore may have an even stronger recruitment pressure. Also, the use of back-up centers might reduce signal detection due to differences, e.g. in the size of patient populations, between them and the initially selected centers. Thus, adding further centers to speed up recruitment may harm the scientific value of a study in MDD. In addition, it may be pertinent to provide notice of study closure well in advance to all sites, and from then on only allow patients to be recruited in numbers which reflect the recruitment rate prior to the closure announcement.

We assessed some variables like “mean HAM-D placebo response” and “SD in HAM-D in placebo group” that are usually associated with a low ES via the size of the placebo response. However the ES between drug and placebo could still be the same, if the response in the placebo and drug group would increase in similar proportions. For this to happen, the treatment effect would really need to be additive to the placebo effect. We notice a considerable association between large placebo responses and high SDs in the HAM-D in the placebo group, i.e. the ES is not only decreased by increasing the denominator but also by decreasing the numerator.

In terms of patient populations participating in the studies, we found a decrease in signal detection with patients who had baseline HAM-D scores that were closest to the cutoffs chosen for enrollment into the source trials. We divided the pooled population into categories based on quartiles according to baseline HAM-D of the complete integrated dataset. Mild baseline HAM-D values close ($\leq$3 points) to the cutoff values for study inclusion ($\leq$18) were found for 34.8% of all patients. These patients showed a much smaller ES of 0.157 relative to the other HAM-D baseline categories, in particular when compared to mild to moderate patients with a HAM-D of 19–21, which showed an ES of 0.333 i.e. the numerically largest ES of any baseline severity category. Therefore, we hypothesize that the smaller ES in patients with a baseline HAM-D $\leq$18 was probably due to the proximity to the HAM-D inclusion cutoff of the study rather than due to a specific HAM-D severity score that eliminates mild patients. The problem of baseline score inflation in antidepressant clinical trials has been identified in several reports (Kobak et al., 2007; Rutherford and Roose 2013; Mundt et al., 2007a). However, traditionally negative study results of patients close to the lower cutoff were interpreted as a decrease in signal detection with patients having mild depressive symptoms relative to patients having moderate or severe symptoms at baseline (Khan et al., 2002, 2004, 2005; Fournier et al., 2010; Kirsh et al., 2008; Rutherford and Roose, 2013). Instead our results might indicate that baseline score inflation affecting HAM-D baseline values that are only a few points higher than the study inclusion criteria may have influenced the results of those meta-analyses.

There has always been a discrepancy between clinical practice and the accepted evidence from randomized controlled trials as to whether or not antidepressant drugs are effective in patients with milder symptoms. Clinicians use antidepressants regularly in these patients, whereas researchers appear to show that the drugs only differentiate from placebo in moderate to severe patients. Our hypothesis is that regardless of the level of severity of patients admitted to a study, possibly due to baseline rating inflation, the lowest category of the study will be the least robust at differentiating active drug from placebo. Thus it may be that the reason for lack of evidence of efficacy in milder patients is related to their position in the lowest category of patients recruited to a study and not to the absolute severity of depression (Posternak et al., 2002; Simon, 2002; Zimmerman et al., 2002, 2005; Fountoulakis and Möller, 2012). As the effect of baseline inflation may be a methodological artifact of randomized clinical trials, the relevance of the small ES in mild patients in clinical trials with regard to the success of treatment of mild depressive patients in clinical practice is questionable. In particular in clinical trials, the high number of patients included with HAM-D baseline levels close to minimum severity, assuming that this is partially due to score inflation at study start, could drag down the overall ES of a study. This would impair signal detection in clinical trials and could account for the discrepancy regarding the perceived effectiveness of antidepressants in mild depressive patients in clinical reality vs. what is usually reported from clinical trials.

One way suggested to improve signal detection in clinical trials in antidepressants would be to use central investigators as raters for baseline HAM-D assessments that are blinded to the study entry criteria. There is, in fact, increasing evidence in the literature that the use of centralized raters could significantly impact the study sample requirement in MDD trials and lead to significantly less change in mood ratings among patients in the placebo arm (Kobak et al., 2010). The centralized rating could improve the reliability of clinical measurement and could theoretically lead to a better and more consistent interview quality, which should have a positive impact on signal detection (Kobak et al., 2005). We cannot deny on the other hand that the use of centralized raters would make the study execution more complex, therefore decreasing the success implementation of the study for other reasons. Another way suggested is to use one depression scale (e.g. HAM-D) for study entry decisions and a different scale (e.g. Montgomery Asperg Depression Rating Scale, MADRS) for the primary efficacy analysis. However, some of the studies in our analysis adopted this approach.

In our opinion, the most promising solution for the problem of baseline inflation could be to restrict the primary analysis of a clinical trial to a population with a higher cutoff (which is blinded to investigators) compared to the cutoff for study participation.

The country of study conduct was found to be another patient population-related factor potentially influencing signal detection. In particular, we found that centers located in the US showed higher proportions of those parameters that were found to have a numerically negative effect on ES compared to placebo (e.g. sites of very small and very large size, low baseline HAM-D score). However, the interpretation of these differences was confounded by the fact that most studies were either performed in one or the other location, and any differences might have been because of study design differences instead of country-related differences. However, some differences in patient recruitment were found, e.g. a larger proportion of patients were recruited in very small and very large centers, and had mild symptoms when recruited in studies conducted in the US (based on individual study results). It could be speculated that financial incentives offered in the US (e.g. guaranteed health insurance), and advertising for patients, may result in less severely affected patients seeking enrollment in studies. Our

results, however, might not always concur with other reports in the literature. In fact, in an extensive analysis, Khin et al. (2011) found a comparable drug-placebo difference in US and non-US MDD trials. Therefore, further research into this topic, including studies that were performed across Europe and the US, are needed.

The current analysis does have potential limitations. It was performed on an extensive database but all studies were done by the same sponsor and thus there is possibly some learning effect over time, which might have confounded or obscured some potential findings. Pooling of patient level data from various sponsors might overcome this problem. In addition, the active treatment in our analyses was mainly duloxetine, thus it not possible to assume that these results would generalize broadly across antidepressants, without data from such.

Furthermore, most studies in the database had a treatment duration of only 8 weeks, and therefore the results might be valid only for acute treatment studies of MDD and not in longer term (usually active controlled) or relapse prevention studies. However, analyses of the relapse prevention studies showed a better performance in signal detection for duloxetine compared to the present analysis (Hudson et al., 2007; Perahia et al., 2009).

5. Conclusions

Our results indicate that the use of simpler trials, with a low number of post-baseline visits, more homogeneous and mid-sized study sites, and potentially restricting the primary analysis of a clinical trial to a population with a higher cutoff (which is blinded to investigators) compared to the cutoff for study participation would lead to better signal detection of antidepressive efficacy.

Importantly, although active controlled studies are vital for comparative efficacy evaluations, adding an active control group to a placebo-controlled clinical trial of MDD may decrease the ability to separate the effect of active treatments from placebo. In addition, we suggest that although recruitment pressure may help to keep to timelines, it may negatively affect the scientific outcome of the study by simultaneously reducing the ES.

Role of the funding source

The studies were funded by Eli Lilly and Company.

Contributors

The studies were funded by Eli Lilly and Company. Data were analyzed by Accovion, Eschborn, Germany, under the oversight of Alexander Schacht. The manuscript was drafted by all authors. All authors made important contributions to the manuscript. All authors reviewed the final draft and provided important intellectual content.

Conflict of interest

MM, ALS, and AS are full-time employees of Eli Lilly & Co. MM and AS own Lilly stock. AGW and GP were involved in clinical trials conducted by Lilly and are members of a Lilly Advisory Board. AGW and GP received conference attendance support and conference support or received speaker's fee by Lilly.

Disclosures

The studies were funded by Eli Lilly and Company. Data were analyzed by Accovion, Eschborn, Germany, under the oversight of Alexander Schacht. The manuscript was drafted by all authors. MM, ALS, and AS are full-time employees of Eli Lilly &Co. MM and AS own Lilly stock. All authors made important contributions to the manuscript. All authors reviewed the final draft and provided important intellectual content. AGW and GP were involved in clinical trials conducted by Lilly and are members of a Lilly Advisory Board. AGW and GP received conference attendance support and conference support or received speaker's fee by Lilly.

Acknowledgments

First and foremost we would like to thank all patients for participating in these studies. Thomas Wagner and Mike Riley from Trilogy Writing & Consulting, Frankfurt/Main, Germany provided medical writing support for this manuscript.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jpsychires.2014.01.001.

References

Brannan SK, Mallinckrodt CH, Brown EB, Wohlreich MM, Watkin JG, Schatzberg AF. Duloxetine 60 mg once-daily in the treatment of painful physical symptoms in patients with major depressive disorder. J Psychiatric Res 2005;39:43–53.
Detke MJ, Lu Y, Goldstein DJ, Hayes JR, Demitrack MA. Duloxetine, 60 mg once daily, for major depressive disorder: a randomized double-blind placebo controlled trial. J Clin Psychiatry 2002a;63:308–15.
Detke MJ, Lu Y, Goldstein DJ, McNamara RK, Demitrack MA. Duloxetine 60 mg once daily dosing vs. placebo in the acute treatment of major depression. J Psychiatric Res 2002b;36:383–90.
Detke MJ, Wiltse CG, Mallinckrodt CH, McNamara RK, Demitrack MA, Bitter I. Duloxetine in the acute and long-term treatment of major depressive disorder: a placebo- and paroxetine-controlled trial. Eur Neuropsychopharmacol 2004;14:457–70.
Dunlop BW, Thase ME, Wun CC, Fayyad R, Guico-Pabia CJ, Musgnung J, et al. A meta-analysis of factors impacting detection of antidepressant efficacy in clinical trials: the importance of academic sites. Neuropsychopharmacology 2012;37:2830–6.
Fountoulakis KN, Möller HJ. Efficacy of antidepressants: a re-analysis and re-interpretation of the Kirsch data. Int J Neuropsychopharmacol 2011;14(3):405–12.
Fountoulakis KN, Möller HJ. Antidepressant drugs and the response in the placebo group: the real problem lies in our understanding of the issue. J Psychopharmacol 2012;26(5):744–50.
Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. J Am Med Assoc 2010;303:47–53.
Gelenberg AJ, Thase ME, Meyer RE, Goodwin FK, Katz MM, Kraemer HC, et al. The history and current state of antidepressant clinical trial design: a call to action for proof-of-concept studies. J Clin Psychiatry 2008;69:1513–28.
Goldstein DJ, Mallinckrodt C, Lu Y, Demitrack MA. Duloxetine in the treatment of major depressive disorder: a double-blind clinical trial. J Clin Psychiatry 2002;63:225–31.
Goldstein DJ, Lu Y, Detke MJ, Wiltse C, Mallinckrodt C, Demitrack MA. Duloxetine in the treatment of depression: a double-blind placebo-controlled comparison with paroxetine. J Clin Psychopharmacol 2004;24:389–99.
Gueorguieva R, Mallinckrodt C, Krystal JH. Trajectories of depression severity in clinical trials of duloxetine: insights into antidepressant and placebo responses. Arch Gen Psychiatry 2011;68:1227–37.
Höschl C. European perspectives in psychiatry: antidepressants as an example. Acta Psychiatr Scand 2008;118:89–90.
Hudson JI, Perahia DG, Gilaberte I, Wang F, Watkin JG, Detke MJ. Duloxetine in the treatment of major depressive disorder: an open-label study. BMC Psychiatry 2007;7:43.
Khan A, Leventhal RM, Khan SR, Brown WA. Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. J Clin Psychopharmacol 2002;22(1):40–5.
Khan A, Detke M, Khan SR, Mallinckrodt C. Placebo response and antidepressant clinical trial outcome. J Nerv Ment Dis 2003a;191:211–8.
Khan A, Khan SR, Walens G, Kolts RL, Giller EL. Frequency of positive studies among fixed and flexible dose antidepressant trials: an analysis of the food and drug administration summary basis of approval reports. Neuropsychopharmacology 2003b;28:552–7.
Khan A, Kolts RL, Thase ME, Krishnan KR, Brown W. Research design features and patient characteristics associated with the outcome of antidepressant clinical trials. Am J Psychiatry 2004;161:2045–9.

Khan A, Brodhead AE, Kolts RL, Brown WA. Severity of depressive symptoms and response to antidepressants and placebo in antidepressant trials. J Psychiatric Res 2005;39:145–50.
Khan A, Bhat A, Kolts R, Thase ME, Brown W. Why has the antidepressant-placebo difference in antidepressant clinical trials diminished over the past three decades? CNS Neurosci Ther 2010;16:217–26.
Khin NA, Chen YF, Yang Y, Yang P, Laughren TP. Exploratory analyses of efficacy data from major depressive disorder trials submitted to the US Food and Drug Administration in support of new drug applications. J Clin Psychiatry 2011;72:464–72.
Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. PLOS Med 2008;5:e45.
Kobak KA, Feiger AD, Lipsitz JD. Interview quality and signal detection in clinical trials. Am J Psychiatry 2005;162:628.
Kobak KA, Leuchter A, DeBrota D, Engelhardt N, Williams JB, Cook IA, et al. Site versus centralized raters in a clinical depression trial: impact on patient selection and placebo response. J Clin Psychopharmacol 2010;30:193–7.
Kobak KA, Kane JM, Thase ME, Nierenberg AA. Why do clinical trials fail? The problem of measurement error in clinical trials: time to test new paradigms? J Clin Psychopharmacol 2007;27:1–5.
Mallinckrodt CH, Goldstein DJ, Detke MJ, Lu Y, Watkin JG, Tran PV. Duloxetine: a new treatment for the emotional and physical symptoms of depression. Prim Care Companion – J Clin Psychiatry 2003;5:19–28.
Mallinckrodt CH, Meyers AL, Prakash A, Faries DE, Detke MJ. Simple options for improving signal detection in antidepressant clinical trials. Psychopharmacol Bull 2007;40:101–14.
Mallinckrodt CH, Zhang L, Prucka WR, Millen BA. Signal detection and placebo response in schizophrenia: parallels with depression. Psychopharmacol Bull 2010;43:53–72.
Mundt JC, Greist JH, Jefferson JW, Katzelnick DJ, DeBrota DJ, Chappell PB, et al. Is it easier to find what you are looking for if you think you know what it looks like? J Clin Psychopharmacol 2007a;27:121–5.
Mundt JC, Debrota DJ, Greist JH. Anchoring perceptions of clinical change on accurate recollection of the past: memory enhanced retrospective evaluation of treatment (MERET). Psychiatry 2007b;4:39–45.
Oakes TM, Myers AL, Marangell LB, Ahl J, Prakash A, Thase ME, et al. Assessment of depressive symptoms and functional outcomes in patients with major depressive disorder treated with duloxetine versus placebo: primary outcomes from two trials conducted under the same protocol. Hum Psychopharmacol Clin Exp 2012;27:47–56.
Nemeroff CB, Schatzberg AF, Goldstein DJ, Detke MJ, Mallinckrodt C, Lu Y, et al. Duloxetine for the treatment of major depressive disorder. Psychopharmacol Bull 2002;36:106–32.
Nierenberg AA, Greist JH, Mallinckrodt CH, Prakash A, Sambunaris A, Tollefson GD, et al. Duloxetine versus escitalopram and placebo in the treatment of patients with major depressive disorder: onset of antidepressant action, a non-inferiority study. Curr Med Res Opin 2007;23(2):401–16.
Papakostas GI, Fava M. Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD. Eur Neuropsychopharmacol 2009;19:34–40.
Perahia DG, Wang F, Mallinckrodt CH, Walker DJ, Detke MJ. Duloxetine in the treatment of major depressive disorder: a placebo- and paroxetine-controlled trial. Eur Psychiatry 2006;21:367–78.
Perahia DG, Maina G, Thase ME, Spann ME, Wang F, Walker DJ, et al. Duloxetine in the prevention of depressive recurrences: a randomized, double-blind, placebo-controlled trial. J Clin Psychiatry 2009;70:706–16.
Posternak MA, Zimmerman M, Keitner GI, Miller IW. A reevaluation of the exclusion criteria used in antidepressant efficacy trials. Am J Psychiatry 2002;159:191–200.
Raskin J, Wiltse CG, Siegal A, et al. Efficacy of duloxetine on cognition, depression, and pain in elderly patients with major depressive disorder: an 8-week, double-blind, placebo-controlled trial. Am J Psychiatry 2007;164:900–9.
Raskin J, Wiltse CG, Dinkel JJ, Walker DJ, Desaiah D, Katona C. Safety and tolerability of duloxetine at 60 mg once daily in elderly patients with major depressive disorder. J Clin Psychopharmacol 2008a;28:32–8.
Raskin J, Xu JY, Kajdasz DK. Time to response for duloxetine 60 mg once daily versus placebo in elderly patients with major depressive disorder. Int Psychogeriatrics 2008b;20(2):309–27.
Robinson M, Oakes TM, Raskin J, Liu P, Shoemakers S, Nelson JC. Acute and long-term treatment of late-life major depressive disorder: duloxetine versus placebo. Am J Geriatric Psychiatry 2014;22(1):34–45.
Rutherford BR, Roose SR. A model of placebo response in antidepressant clinical trials. Am J Psychiatry 2013;170(7):723–33.
Schacht A, Dyachkova Y, Walton RJ. Critical evaluation of mixed treatment comparison meta-analyses using examples assessing antidepressants and opioid detoxification treatments. Int J Methods Psychiatric Res 2013;22:166–74.
Simon GE. Evidence review: efficacy and effectiveness of antidepressant treatment in primary care. Gen Hosp Psychiatry 2002;24:213–24.
Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. New Engl J Med 2008;358:252–60.
Turner EH, Rosenthal R. Efficacy of antidepressants. Br Med J 2008;336(7643):516–7.
Wise TN, Wiltse CG, Losifescu DV, Sheridan M, Xu JY, Raskin J. The safety and tolerability of duloxetine in depressed elderly patients with and without medical comorbidity. Int J Clin Pract 2007;61:1283–93.
Zimmerman M, Posternak MA, Chelminski I. Symptom severity and exclusion from antidepressant efficacy trials. J Clin Psychopharmacol 2002;22:610–4.
Zimmerman M, Chelminski I, Posternak MA. Generalizability of antidepressant efficacy trials: differences between depressed psychiatric outpatients who would or would not qualify for an efficacy trial. Am J Psychiatry 2005;162:1370–2.