

# Re-evaluation of the efficacy and tolerability of venlafaxine vs SSRI: meta-analysis

S. Weinmann · T. Becker · M. Koesters

Received: 3 April 2007 / Accepted: 2 October 2007 / Published online: 23 October 2007  
© Springer-Verlag 2007

## Abstract

**Rationale** A number of reviews have claimed that the selective serotonin and noradrenalin re-uptake inhibitor venlafaxine is more effective than selective serotonin re-uptake inhibitors (SSRIs) in achieving remission and symptom reduction in major depression.

**Objectives** The aim of this study was to systematically review studies on the efficacy of venlafaxine vs SSRI and to evaluate the influence of methodological issues on the effect sizes.

**Materials and methods** Following a systematic literature search, we pooled data on depression scores, response, remission and dropout rates. We also performed sub-group analyses.

**Results** Seventeen studies were included. We found no significant superiority in remission rates (risk ratio [RR]=1.07, 95% confidence intervals [95%CI]=0.99 to 1.15, numbers needed to treat [NNT]=34) and a small superiority in response rates (RR=1.06, 95%CI=1.01 to 1.12, NNT=27) over SSRIs. There was a small advantage to venlafaxine in change scores (effect size=-0.09, 95%CI=-0.16 to -0.02,  $p=0.013$ ), which did not reach significance when post-treatment scores were used (effect size=-0.06, 95%CI=-0.13 to 0.00). Discontinuation rates due to adverse events were 45% higher in the venlafaxine group. The main reasons for the differences between this analysis and previous reviews were the exclusion of studies with methodological

limitations, avoiding to pool selectively reported study results and exclusion of studies available as abstracts only.

**Conclusions** Our analysis does not support a clinically significant superiority of venlafaxine over SSRIs. Differences between our study and previous reviews were not accounted for by technical aspects of data synthesis, but rather by study selection and choice of outcome parameters.

**Keywords** Venlafaxine · Anti-depressant · Meta-analysis · Systematic review

## Introduction

Major depression has been recognised as a high priority in primary health care with treatment rates having increased (Kessler et al. 2005). However, many patients do not achieve remission, although there is a broad consensus that full remission and recovery should be the primary aims in treating depression (Paykel 2002; Solomon et al. 2000). Remission and recovery as optimal treatment outcomes are defined in most studies as an asymptomatic status with no or only minimal symptoms for a specified period after a depressive episode has resolved (Frank et al. 1991).

During recent years, several new anti-depressants have been developed with the aim of improving the efficacy and safety of pharmacological treatment. There has been many debates if available anti-depressants differ in the rates of remission or in the reduction of symptoms (Hansen et al. 2005; MacGillivray et al. 2003). Despite the considerable number of clinical trials, evidence of differential efficacy is sparse. The selective serotonin and noradrenalin re-uptake inhibitor (SNRI) venlafaxine has been claimed to be more effective than selective serotonin re-uptake inhibitors (SSRIs) particularly with respect to remission rates

**Electronic supplementary material** The online version of this article (doi:10.1007/s00213-007-0975-9) contains supplementary material, which is available to authorized users.

S. Weinmann (✉) · T. Becker · M. Koesters  
University of Ulm,  
Ulm, Germany  
e-mail: stefan.weinmann@gmx.de

(Rudolph 2002; Shelton et al. 2005; Smith et al. 2002; Thase et al. 2001). The proposed mechanism for the differences in efficacy is the dual action of venlafaxine, suggesting an additive benefit of re-uptake inhibition of serotonin and noradrenalin. The superiority of dual acting substances was supported by clinical data, e.g. in RCTs of clomipramine vs SSRIs such as citalopram or paroxetine (Danish University Antidepressant Group 1986, 1990).

However, most guidelines do not recommend venlafaxine as a first-line option (American Psychiatric Association 2000; Fochtmann et al. 2006; National Institute for Health and Clinical Excellence 2004). The discrepancy between national guideline recommendations and the reviews mentioned above may be explained by the differences in the interpretation of the study evidence, methodological choices in systematic review or guideline development or additional consideration of side effects (Cipriani et al. 2007). Despite being intended to provide the best summary of the available evidence, there may be many limitations and shortcomings of meta-analyses in depression research beyond the quality of the included primary studies (Anderson 2001). Recently, guidelines for quantitative reviews of anti-depressants have been published addressing a variety of methodological issues (Lieberman et al. 2005).

To synthesise the study evidence on the comparative efficacy of venlafaxine vs other anti-depressants, we performed a complete up-to-date systematic review and meta-analysis on this substance and explored the reasons for the differences in effect sizes between this and other reviews in the field. We restricted this analysis to the comparison of venlafaxine against SSRIs. This comparison is most interesting from a pharmacological and a clinical viewpoint, as SSRIs are now the anti-depressants with the highest prescription rates in developed countries. We then compared our findings with previous relevant reviews.

## Materials and methods

### Systematic review

This review was part of a larger review of venlafaxine vs other anti-depressants. We searched the Medline, EMBASE, PsycINFO, PSYINDEX, Cochrane Central Register of Controlled Trials, study registers (<http://www.clinicalstudyresults.org>; <http://www.clinicaltrials.gov>) and the manufacturer's database (<http://www.wyeth.com>) for primary studies between 1966 and January 2006. The literature search was updated in March 2007. Furthermore, in addition to the abovementioned databases, the Cochrane Database of Systematic Reviews, DARE and guideline databases (<http://www.guidelines.gov>; <http://www.g-i-n.net>) were screened for systematic reviews to identify further primary studies. We

used a sensitive search strategy based on a combination of text and index terms as a modification of the search strategy of the Cochrane Depression and Anxiety group (available from the authors).

### Study selection

In this review, we included double-blind randomised controlled trials in which venlafaxine was compared to citalopram, escitalopram, fluoxetine, fluvoxamine, paroxetine or sertraline with or without a placebo control. We firstly searched all publications comparing venlafaxine with other anti-depressants. In the next step, we chose only those studies where the substance was compared to an SSRI. We did not include conference abstracts unless we could obtain a full-text publication. We included only those studies in which depression was diagnosed according to ICD, DSM or Research Diagnostic Criteria (RDC). To assure the comparability of the studies, we included only studies with at least 6 weeks of double-blind treatment. Long-term studies with more than 6 months duration were excluded because we focused on short-term effectiveness and not on relapse and recurrence prevention. Similar to the criteria used in the British NICE guideline (National Institute for Health and Clinical Excellence 2004), we excluded studies in which more than 20% of participants had a primary diagnosis of dysthymia or more than 15% had a primary diagnosis of bipolar disorder. Studies in geriatric populations were included taking into account that there is no clear evidence of a differential anti-depressant efficacy in general between younger and older persons (Roose and Schatzberg 2005).

We only included studies which used the standard instruments Hamilton Depression Rating Scale (HAM-D) or Montgomery Asberg Depression Rating Scale (MADRS) as outcome parameters. Studies only reporting global ratings such as the Clinical Global Impression Scale (CGI) or Patient Global Improvement (PGI) were excluded because judgements are left to clinicians' discretion with poorly specified global ratings not necessarily reflecting clinical relevance. We did not employ pre-defined remission and response definitions. Although, according to Frank et al. (1991), the HAM-D score below 8 is a very common remission criterion, many other definitions are used (Keller 2003), and there is still an ongoing debate on specific cutoff points (Zimmerman et al. 2005). However, we performed a sensitivity analysis including only studies that used the following standards: remission defined as HAM-D-17 (17-item version)  $\leq 7$  or MADRS  $\leq 10$ ; response defined as a reduction of at least 50% in HAM-D or MADRS scores. Most studies used HAM-D or MADRS scores as primary endpoint. However, a small number of studies reported only composite endpoints derived at by combining several scales (e.g. response was defined in one study as the reduction in

the HAM-D or MADRS score of at least 50% and a CGI improvement score of 1 or 2). These composite scores were included when no single score was available because the literature suggests a relatively unimportant influence of choosing composite primary outcomes (Freemantle and Calvert 2007).

Publications were screened in two steps. First, all abstracts from the primary search were screened for potentially relevant articles. For these publications, full-text versions were obtained and screened using our inclusion and exclusion criteria. A total of 981 publications comparing venlafaxine to other anti-depressants were identified with our search strategy (see the [Electronic supplementary material](#)). One hundred fifty-one publications were screened in full text.

#### Data collection and analysis

Two authors extracted data independently. We assessed study and publication characteristics such as primary and secondary outcome measures, blinding, allocation concealment, length of follow-up, handling of missing data, statistical methods (last observation carried forward [LOCF] vs observed case [OC] analysis) and study population characteristics such as sociodemographic and baseline psychopathological parameters.

To evaluate the adequacy of the intention to treat (ITT) analysis, three categories were defined. Data analysis was judged as “adequate ITT analysis” when all randomised patients were evaluated within their allocation group using the LOCF or repeated measures analysis, or when only those randomised patients with one post-baseline measurement were included with no more than 10% of randomised patients being excluded from the analysis. Furthermore, there had to be a description of how patients lost to follow-up had been dealt with in the evaluation. Data analysis was judged as “acceptable ITT analysis” when LOCF or repeated measures analyses were performed with no more than 10% of patients lost to follow-up, no more than 30% patients having dropped out and adequate descriptions of the reasons for dropout were provided. Per protocol, OC or completer analysis, or analyses with less than 90% of randomised patients were classified as “inadequate ITT analysis”. Studies with “inadequate ITT analysis” according to our judgment were excluded.

#### Effect size calculation

Standardised effect sizes were calculated as the standardised mean differences from the change scores of the 17- or 21-item version of the HAM-D or the MADRS. In addition, post-treatment scores were extracted for sensitivity analyses. The standardised mean difference (SMD) is

the improvement score of the venlafaxine group minus the improvement score of the SSRI group divided by the pooled standard deviation of the improvement score. For dichotomous data, we calculated risk ratios (RR) and numbers needed to treat (NNT) or numbers needed to harm (NNH). Statistical significance was evaluated using the Mantel–Haenszel statistic with 95% confidence intervals (95%CI). The numbers of participants with response or remission were calculated including dropouts within last observation carried forward analyses (LOCF).

Missing standard deviations (SDs) were imputed for each trial by calculating the pooled SD from all other studies that used the same depression scale in the same version. Borrowing SDs from other studies by this imputation method has been shown to be safe and appropriate, and the risk of bias is low (Furukawa et al. 2006; Thiessen Philbrook et al. 2007). HAM-D values had to be estimated from figures in publications for some studies.

Primarily, we used a fixed effects model for pooled analyses as two drug classes were evaluated, but as part of heterogeneity analysis, we compared them with random effects models. The random effects model assumes a different underlying effect for each study and takes this into consideration as an additional source of variation (Egger et al. 2001). The fixed effects model considers this variability as exclusively due to random variation, and individual studies are simply weighted by their precision. To quantify heterogeneity between studies, we used the  $I^2$  statistic (Higgins et al. 2003). Low heterogeneity is associated with an  $I^2$  value of 25%, whilst high heterogeneity is associated with an  $I^2$  of 75% (Higgins et al. 2003). Statistical significance of heterogeneity was tested with chi square tests. We defined a substantial heterogeneity between study results as  $I^2 > 50%$  or  $p < 0.2$ . As a measure of tolerability data, RRs were calculated for the proportion of patients failing to complete the study (overall dropout rate) and for dropouts due to adverse events.

#### Sensitivity analyses

We performed a variety of sensitivity analyses to determine whether results were altered by excluding certain studies or by using different type of data or statistical methods. We prospectively defined several sub-groups for analysis: (1) age of participants (non-geriatric vs geriatric) with a cutoff of 65 years, (2) type of trial (active control-only trial vs active- and placebo-controlled trial), (3) adequate vs acceptable ITT analysis, (4) use of change scores vs post-treatment scores in standardised mean difference analysis, (5) inpatients vs outpatients, (6) fixed dose venlafaxine vs flexible dose venlafaxine and (7) venlafaxine as experimental vs control substance. In addition, we added further

sensitivity analyses after completing the efficacy results. We evaluated studies using or not using standard response or remission criteria (see the “Study Selection” section) and those with or without data derived from figures. Furthermore, we evaluated the influence of the dosage on the effect size by means of meta-regression. The mean daily dosage at endpoint was only reported in some of the flexible dosage studies, therefore, we calculated two meta-regression analyses separately with mean daily dosage and maximum allowed dosage as predictors. A further regression analysis was performed for the influence of the year of publication on effect size. A funnel plot was used to evaluate possible publication bias.

We searched other reviews with quantitative data synthesis and re-evaluated the recent comprehensive analysis of Smith et al. (2002) using our own study inclusion criteria, but limiting the studies to the period up to 2000 when that review had been undertaken. For both evaluations, we included full-text publications only. In case of Smith et al. (2002) having included only abstracts, we searched for the full-text publications of these abstracts. We then compared our own results with those of this review and other previous analyses. We discussed the reasons for the differences in effect size estimations and conclusions.

Data were analysed using Comprehensive Meta Analysis Version 2.2 (Borenstein et al. 2005).

## Results

### Study characteristics

We identified 28 studies comparing venlafaxine with other anti-depressants as potentially appropriate. Seventeen studies had a SSRI control group and were included in this meta-analysis (Allard et al. 2004; Alves et al. 1999; Bielski et al. 2004; Clerc et al. 1994; Costa e Silva 1998; Dierick et al. 1996; McPartlin et al. 1998; Mehtonen et al. 2000; Montgomery et al. 2004; Nemeroff and Thase 2007; Rudolph and Feiger 1999; Schatzberg and Roose 2006; Shelton et al. 2006; Silverstone and Ravindran 1999; Sir et al. 2005; Tylee et al. 1997; Tzanakaki et al. 2000) (see the [Electronic supplementary material](#)). We included 15 studies with adult non-geriatric patients and 2 studies with geriatric patients of at least 65 years of age. Among the studies of non-geriatric populations, three were active- and placebo-controlled trials with 3 treatment arms and 12 were active control-only trials with 2 treatment arms. Among the studies in old age, one study was active- and placebo-controlled and one was active control-only. Study duration was between 6 and 24 weeks with most studies (14 trials) having a double-blind study duration of 6–8 weeks. Trial size varied between 68 and 382 patients. One study

compared venlafaxine with citalopram, two with escitalopram, ten with fluoxetine, one with paroxetine and three with sertraline. Baseline HAM-D varied considerably between studies. Whereas in the non-geriatric study group in one study, the mean baseline HAM-D (17-item version) was 20.4 in the venlafaxine group and 19.9 in the escitalopram group (Montgomery et al. 2004), patients in another study had a baseline HAM-D (17-item version) of 29.1 (venlafaxine) and 29.7 (fluoxetine) (Clerc et al. 1994; see the [Electronic supplementary material](#)).

All studies used an ITT analysis using the LOCF method. However, only seven studies were judged to have an adequate ITT analysis (Allard et al. 2004; McPartlin et al. 1998; Montgomery et al. 2004; Nemeroff and Thase 2007; Rudolph and Feiger 1999; Schatzberg and Roose 2006; Sir et al. 2005). The randomisation method was adequately reported in only four studies (Rudolph and Feiger 1999; Sir et al. 2005; Tylee et al. 1997; Schatzberg and Roose 2006), whereas only one study reported methods to assure adequate allocation concealment beyond the blinding of patients and physicians (Schatzberg and Roose 2006). Funnel plots did not show signs of publication bias.

### Remission rates

Fifteen studies could be included in the pooled remission analysis. In two studies (Tylee et al. 1997; Montgomery et al. 2004), only MADRS data was available, whereas in 12 studies, the HAM-D was used alone or as combined outcome parameter to calculate the remission rates. Remission definitions differed considerably between studies.

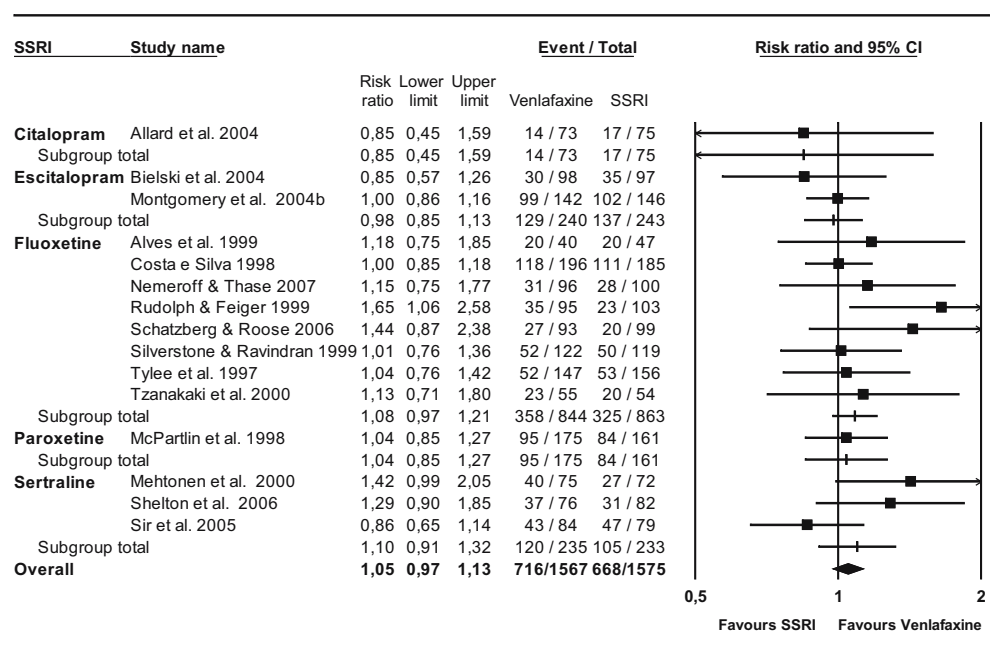
Figure 1 shows the final pooled remission rates. The RR for remission was 1.07 (95%CI=0.99 to 1.15, NNT=34). There was no statistically significant difference between venlafaxine and the SSRI group. Heterogeneity between studies was low and non-significant ( $I^2=18%$ ,  $p=0.248$ ). Therefore, using a random effects model changed the results only marginally (RR=1.04, 95%CI=0.96 to 1.13).

### Response rates

For response analysis, we could include all 17 studies. For all but two studies (Allard et al. 2004; Montgomery et al. 2004), HAM-D scores were used to calculate the response rates. Response was defined as a reduction of at least 50% in HAM-D score or as a reduction of at least 50% in HAM-D or MADRS score and a CGI improvement score of 1 or 2.

The RR for response was 1.06 (95%CI=1.01 to 1.12, NNT=27) with venlafaxine being marginally superior to the SSRI group (Fig. 2). Heterogeneity was moderate ( $I^2=32%$ ) and significant according to our definition ( $p=0.099$ ). Using the more conservative random effects approach, point

**Fig. 1** Mantel–Haenszel RR for remission; fixed effects model



estimates changed only marginally (RR=1.05, 95%CI= 1.00–1.10) but resulted in non-significance ( $p=0.053$ ).

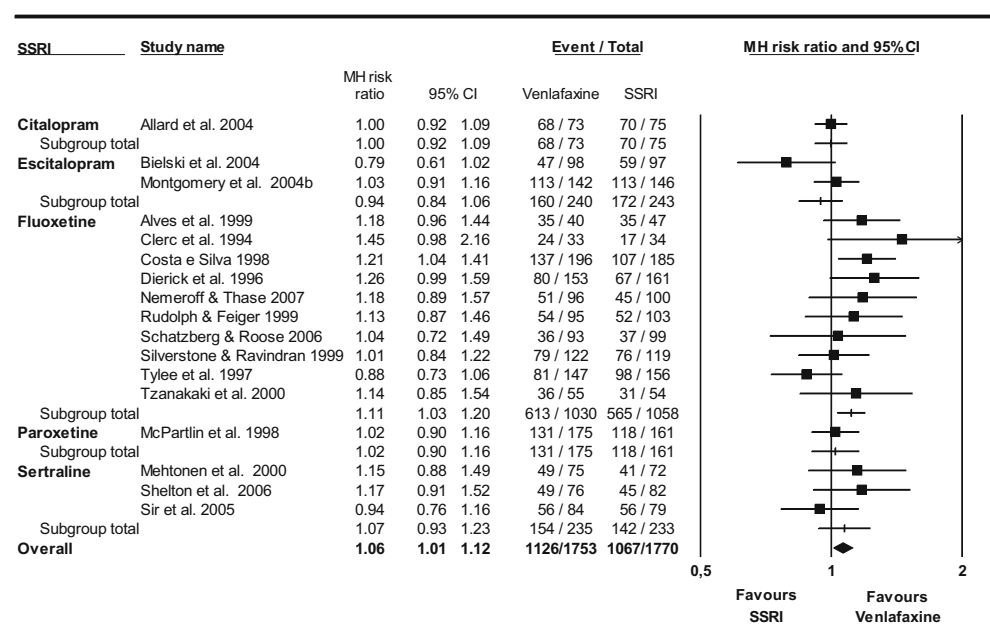
Effect size: standardised mean difference

All but one study provided data to calculate the standardised mean differences (SMD) in HAM-D or MADRS change scores. The SMD was  $-0.09$  (95%CI= $-0.16$  to  $-0.02$ ,  $p=0.013$ ) in favour of venlafaxine (Fig. 3). Heterogeneity was

moderate ( $I^2=39\%$ ;  $p=0.063$ ), but statistical significance did not change when a random effects model was applied (SMD= $0.10$ , 95%CI= $-0.19$  to  $-0.02$ ,  $p=0.02$ ).

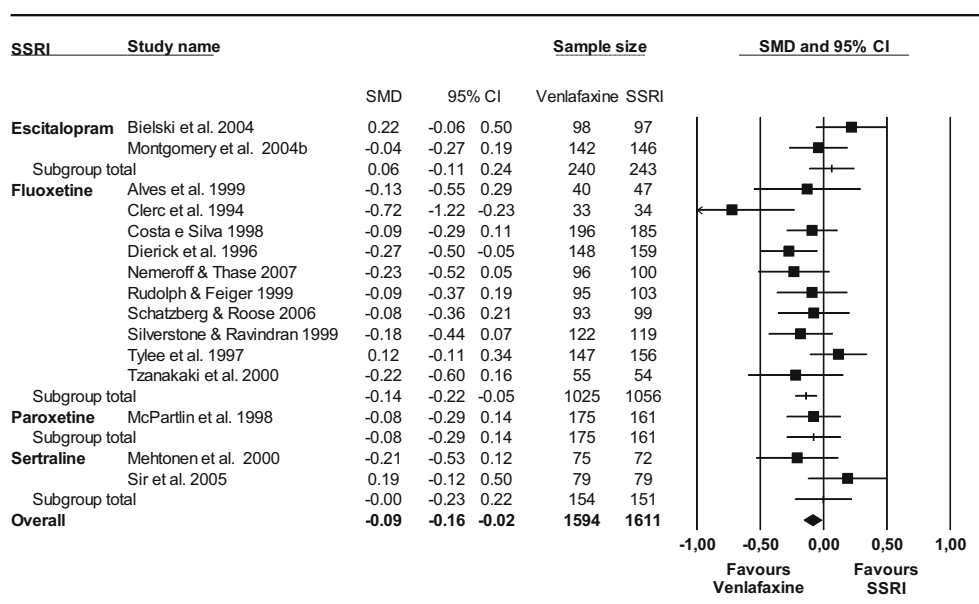
There was no difference in the overall symptom score between venlafaxine and SSRI when post-treatment scores were used (SMD= $-0.06$ , 95%CI= $-0.13$  to  $0.00$ ). The effect was homogeneous ( $I^2=0\%$ ,  $p=0.582$ ); therefore, using a random effects model yielded nearly the same result (SMD= $-0.06$ , 95%CI= $-0.13$  to  $-0.00$ ,  $p=0.07$ ).

**Fig. 2** Mantel–Haenszel RR for response; fixed effects model





**Fig. 3** Standardised mean difference of change scores; fixed effects model



**Tolerability**

For the analysis of total treatment discontinuation rates, we had to exclude the study of Allard et al. (2004) because dropout rates were not reported separately for study arms. However, all studies could be included in the analysis of the rates of dropout due to adverse effects. The total rate of treatment discontinuation did not differ between venlafaxine and SSRIs (RR=1.05, 95%CI=0.93 to 1.2, NNH=100). However, there were significantly more dropouts due to adverse effects in the venlafaxine group with a RR of 1.38 (95%CI=1.08 to 1.77, NNH=32) in favour of SSRIs (see the [Electronic supplementary material](#)).

**Sensitivity analyses**

Including only studies with adequate ITT analysis according to our criteria did not change the results substantially (Allard et al. 2004; McPartlin et al. 1998; Montgomery et al. 2004; Nemeroff and Thase 2007; Rudolph and Feiger 1999; Schatzberg and Roose 2006; Sir et al. 2005). In those studies, none of the outcome parameters showed an advantage of venlafaxine (Table 1). In studies with flexible venlafaxine dosage regimes, remission and response rates and effect sizes were slightly higher compared to the SSRI group, although results have to be interpreted with caution due to the low number of studies with fixed doses. None of the sub-group analyses showed a statistically significant group difference (Table 1). However, when venlafaxine was used as experimental substance, an advantage could be seen in all outcome parameters, whereas none of the analyses showed a difference between venlafaxine and SSRI when venlafaxine was used as comparator. It has to be taken into account that confounding effects (e.g. publication year)

could not be evaluated due to the low number of studies. There were no significant differences in remission or response rates between those studies using pre-defined standard response and remission scores and those using different cutoff points or composite endpoints or other. Furthermore, there were no significant differences in the SMD change scores between those studies with fully reported data (SMD=-0.06, 95%CI=-0.14 to -0.03) and those studies with data derived only from figures (SMD=-0.15, 95%CI=-0.26 to -0.03).

Meta-regression showed neither an influence of mean daily dosage ( $p=0.732$ ), publication year ( $p=0.742$ ) nor an influence of the maximum allowed dosage ( $p=0.690$ ) on remission rates or any other outcome (see the [Electronic supplementary material](#) as example).

A re-analysis of the Smith et al. (2002) data applying our inclusion and exclusion criteria resulted in eight studies with an overall RR for remission of 1.11 (95%CI=1.01 to 1.23) or an odds ratio (OR) of 1.22 (95%CI=1.01 to 1.47) in the SSRI controlled venlafaxine studies. Our results differed from those of the analysis by Smith et al. (2002) where 16 studies with an SSRI comparison group had been included and an OR of 1.43 (95%CI=1.21 to 1.71) favouring venlafaxine vs SSRI had been reported. A re-analysis of the response data of 11 studies included in Smith et al. (2002) meeting our inclusion criteria resulted in an OR of 1.26 (95%CI=1.07 to 1.49) favouring venlafaxine matching exactly the point estimate by Smith et al. 2002 (OR=1.26, 95%CI=1.02 to 1.58). Re-calculating the post-treatment score data of the Smith et al. (2002) analysis using our set of inclusion criteria resulted in no statistical superiority of venlafaxine. The overall SMD of the ten studies was -0.08 (95%CI=-0.17 to 0.00) vs SSRI. This result contrasted to the pooled effect size of -0.17 (95%CI=

**Table 1** Results of sub-group analyses

Sub-group	Remission		Response		Change score		Post-treatment score	
	<i>k</i>	RR (95%CI)	<i>k</i>	RR (95%CI)	<i>k</i>	SMD (95%CI)	<i>k</i>	SMD (95%CI)
Age								
Non-geriatric	12	1.06 (0.98 to 1.15)	15	1.07 (1.01 to 1.13)	15	-0.09 (-0.16 to -0.02)	15	-0.06 (-0.13 to 0.01)
Geriatric	2	1.16 (0.79 to 1.72)	2	1.01 (0.88 to 1.16)	1	-0.08 (-0.36 to 0.21)	1	-0.10 (-0.38 to 0.18)
Type of trial								
Active	11	1.04 (0.96 to 1.12)	14	1.06 (1.01 to 1.12)	12	-0.08 (-0.16 to 0.00)	13	-0.05 (-0.13 to 0.02)
Active+placebo	3	1.26 (1.01 to 1.57)	3	1.05 (0.9 to 1.22)	3	-0.12 (-0.28 to 0.03)	3	-0.11 (-0.27 to 0.04)
ITT analysis								
Adequate	7	1.07 (0.96 to 1.19)	7	1.04 (0.97 to 1.11)	6	-0.06 (-0.17 to 0.04)	7	-0.06 (-0.16 to 0.05)
Acceptable	8	1.07 (0.96 to 1.19)	10	1.08 (1.01 to 1.16)	9	-0.11 (-0.20 to -0.01)	10	-0.07 (-0.16 to 0.02)
Setting								
Inpatients	1	1.13 (0.71 to 1.80)	2	1.25 (0.98 to 1.59)	2	-0.40 (-0.70 to -0.11)	2	-0.27 (-0.57 to 0.02)
Outpatients	13	1.07 (0.99 to 1.15)	15	1.05 (1.00 to 1.11)	13	-0.07 (-0.14 to 0.00)	14	-0.05 (-0.12 to 0.02)
Dose								
Fixed	3	1.00 (0.86 to 1.17)	4	0.95 (0.86 to 1.05)	4	0.01 (-0.12 to 0.14)	4	-0.02 (-0.15 to 0.11)
Flexible	11	1.09 (1.00 to 1.19)	13	1.10 (1.04 to 1.17)	11	-0.13 (-0.21 to -0.04)	12	-0.08 (-0.16 to 0.00)
Venlafaxine group								
Comparator	4	0.98 (0.87 to 1.11)	4	0.98 (0.89 to 1.08)	3	0.10 (-0.06 to 0.25)	4	0.03 (-0.11 to 0.16)
Experimental	10	1.11 (1.01 to 1.22)	13	1.09 (1.03 to 1.16)	12	-0.13 (-0.21 to -0.06)	12	-0.09 (-0.17 to -0.02)
Data source								
Exact numbers	14	0.03 (0.00 to 0.06)	13	0.04 (0.01 to 0.08)	9	-0.06 (-0.14 to 0.03)	10	-0.05 (-0.14 to 0.03)
Figures	1	0.01 (-0.12 to 0.13)	4	0.02 (-0.04 to 0.09)	6	-0.15 (-0.26 to -0.03)	6	-0.09 (-0.21 to 0.03)
Standard remission definition	9	1.07 (0.97 to 1.19)						
Standard response definition			14	1.06 (1.01 to 1.12)				

*k*: number of studies

-0.27 to -0.08) favouring venlafaxine vs SSRI in Smith et al. (2002).

## Discussion

In a systematic review of randomised controlled studies, we found no benefits in remission rates and a small superiority in response rates of venlafaxine over SSRIs. There were no differences in effect size comparing the HAM-D and MADRS post-treatment scores. Study quality of most included trials was at best moderate with detailed information lacking concerning blinding of patients, physicians and outcome raters, allocation concealment and statistical analysis. The results of our pooled analysis suggest no solid evidence of a superiority of venlafaxine over SSRI. This corresponds with the NICE depression guideline, which concluded that venlafaxine was no more effective in treating depression than other anti-depressants (National Institute for Health and Clinical Excellence 2004, p. 220), but is in contrast with other published reviews in the field.

We could not identify a Cochrane Review comparing venlafaxine with SSRIs. In one Cochrane Review, venlafaxine was counted as SSRI, and only studies comparing venlafaxine

and older anti-depressants were evaluated (Geddes et al. 2006). We found seven reviews with formal meta-analysis of randomised controlled trials with venlafaxine (Rudolph 2002; Shelton et al. 2005; Smith et al. 2002; Thase et al. 2001; Einarson et al. 1999; Machado et al. 2006; Hansen et al. 2005); three of which did not perform a systematic literature search (Rudolph 2002; Shelton et al. 2005; Thase et al. 2001). Results of non-systematic reviews should therefore be regarded with caution. An unbiased sampling of studies with pre-defined inclusion and exclusion criteria on the basis of a systematic literature search in relevant electronic databases is a necessary prerequisite of a high-quality meta-analysis. One review included only six studies comparing fluoxetine vs venlafaxine, and it remains unclear why many other relevant studies were not considered (Hansen et al. 2005). In one meta-analysis, studies with venlafaxine extended release (XR), SSRI or tri-cyclic anti-depressant (TCA) response data were searched and compared (Einarson et al. 1999). However, response data from placebo-controlled venlafaxine studies and active control studies were evaluated jointly. In a review with narrow inclusion criteria performed to compare remission, dropout and adverse drug reaction rates of SNRI, SSRI and TCA (Machado et al. 2006), only four venlafaxine studies were included. In this analysis, remission rates were not taken

from similar active-controlled studies, but pooled indirectly using non-contemporary control groups.

The systematic review by Smith et al. (2002) included 16 studies for the remission analysis, 17 studies for the response analysis and 19 studies for the effect size analysis of venlafaxine vs SSRI. However, one study (Rudolph et al. 1998) included by Smith et al. (2002) was only placebo-controlled and, therefore, did not fulfil the inclusion criteria. A second study (Alves et al. 1999) only presented remission rates at 3 weeks and not at the final endpoint of 12 weeks. Therefore, in our analysis, only the response rate data of this study was considered. Furthermore, in two studies included only in Smith et al. (2002), a fair number of patients with dysthymia (Ballus et al. 2000) or patients with a bipolar disorder (Zanardi et al. 2000) participated, thus limiting the relevance for major depression. With one exception (Zanardi et al. 2000), the ORs for remission in studies included by Smith et al. (2002), but excluded from our analysis, were above 1.40. On the other hand, remission rates in studies included in both reviews were below 1.40 with three exceptions (Schatzberg and Roose 2006; Rudolph and Feiger 1999; Mehtonen et al. 2000). This trend towards higher ORs or larger effects sizes in studies included only by Smith et al. (2002) was true also for the response analysis and the SMD effect size analysis. In some studies included in both reviews, there was more than one remission definition. Although it was stated that preference had been given to the HAM-D scale, this was not done consistently. One publication reports remission rates of 60.2% in both the venlafaxine and the fluoxetine groups (Costa e Silva 1998). The OR for remission of 1.15 in favour of venlafaxine calculated for this study in Smith et al. (2002) seems to be due to having used not the HAM-D scale but combined outcome parameters. In the case of one study, Smith et al. (2002) used an OR of 0.54 instead of the published value of 1.76 for the response analysis (Dierick et al. 1996). For one study (Mehtonen et al. 2000), OC data were used resulting in higher ORs for remission. In one study included by Smith et al. (2002), neither DSM, ICD nor RDC was used to assure the depression diagnosis (Geerts et al. 1999). Furthermore, we did not include data of Salinas and Venlafaxine-XR- Study-Group (1997) because it was only available as an abstract.

As Kavirajan (2004) mentions, the data on over half of the patients included in the Smith et al. (2002) analysis were taken from studies that had not been published as articles in peer-reviewed journals or were available as abstracts only. Excluding abstracts without full-text publications contributed to the differences in results, as abstract-only publications tended to have larger effect sizes than full-text studies in the Smith et al. (2002) analysis.

The manufacturer of venlafaxine provided funding to five of the previous reviews (Rudolph 2002; Shelton et al.

2005; Smith et al. 2002; Thase et al. 2001; Einarson et al. 1999). One possible advantage of this source of funding may be that authors have the opportunity to access unpublished data. In fact, the so-called file-drawer effect referring to researchers' tendency to preferably publish positive findings and not to submit "negative trials" may be a major source of publication bias leading to the potential inflation of effect sizes (Khan et al. 2002). This must be weighed against the potential lack of transparency (Jorgensen et al. 2006).

Excluding industry data prevented a full re-analysis of the analysis by Smith et al. (2002) and may be seen as a limitation. Instead, we used strict inclusion criteria to re-evaluate the frequently cited meta-analysis of Smith et al. (2002) to detect sources of variation. Results showed that the differences were not due to the set of newer studies published after 2000. However, our sub-group analysis demonstrated that venlafaxine showed some differential efficacy when used as an experimental or as a comparator drug. Although there were no significant difference between the sub-groups, remission rates and effect sizes were higher in the venlafaxine groups compared to other anti-depressants when venlafaxine was used as the experimental substance. These results may be interpreted, as previously shown for fluoxetine (Barbui et al. 2004), as 'wish bias', suggesting a certain role for this kind of sub-group analysis in systematic reviews of anti-depressants. The results of the regression analyses did not show a dose-response relationship. However, due to the great number of flexible-dose studies, differential efficacy associated with dual action starting at doses above 150 mg could not be established.

In this analysis, SSRIs are treated as a drug class. However, there may be differences in efficacy or effectiveness among them, which may get lost in the overall comparison to venlafaxine due to the small number of head-to-head studies. For example, there is some published evidence for a greater efficacy of escitalopram vs other SSRIs (Kennedy et al. 2006). Furthermore, the overall result could obscure a clinically important difference in depression severity sub-groups, as we could not perform a sensitivity analysis on baseline depression severity. It is well known that at milder degrees of depression it may be hard to show a drug-placebo difference, which may be even more difficult between two active treatment arms. Effects may become clearer in the trials that include patients with more severe depression (Khan et al. 2002).

Excluding studies that do not meet certain quality standards from a systematic review is a controversial issue. There is evidence that randomisation, adequate allocation concealment and adequate ITT analysis are important to prevent bias (Moher et al. 1998). However, there is no consensus on how to integrate methodological quality of



RCTs in systematic reviews (Moja et al. 2005). We decided not to include studies with major methodological problems. Excluding studies diminishes statistical power. However, this must be weighed against the potential bias introduced by these studies.

Tolerability data showed no differences between venlafaxine and SSRIs in dropout rates. However, the higher rate of study withdrawal due to adverse events adds further to the argument that venlafaxine may have risks for many patients possibly outweighing this substance's claimed benefits (Cipriani et al. 2007). More large-scale studies may be necessary to identify patient predictors for a differential benefit compared to other substances.

In conclusion, our review does not support the notion of a clinical superiority or an improved trade-off between efficacy and side effects of the specific dual re-uptake inhibition antidepressant venlafaxine over SSRIs. Among the most important reasons for the differences in remission and response rates between our analysis and previous reviews were:

- exclusion of studies with low methodological quality;
- avoidance of pooling selectively reported study results or data from endpoints not prospectively defined and
- exclusion of abstract-only studies.

Manufacturer sponsoring may have been one reason for the differences in the results.

**Acknowledgements** All authors have been involved in a health technology assessment (HTA) project on the efficacy of selective serotonin and noradrenalin re-uptake inhibitors in major depression, which was funded by the German Institute for Quality and Efficiency in Health Care (IQWiG). However, apart from the literature search, this paper is entirely independent of the HTA project and based solely on published data. It may differ in its methods and conclusions and has therefore neither prejudicial quality nor does it anticipate any results of the HTA report. The paper does not necessarily reflect the opinion of the IQWiG.

The present study received no further external funding.

We thank two anonymous reviewers for valuable comments.

## References

- Allard P, Gram L, Timdahl K, Behnke K, Hanson M, Sogaard J (2004) Efficacy and tolerability of venlafaxine in geriatric outpatients with major depression: a double-blind, randomised 6-month comparative trial with citalopram. *Int J Geriatr Psychiatry* 19:1123–1130
- Alves C, Cachola I, Brandao J (1999) Efficacy and tolerability of venlafaxine and fluoxetine in outpatients with major depression. *Prim Care Companion J Clin Psychiatry* 5:57–63
- American Psychiatric Association (2000) Practice guideline for the treatment of patients with major depressive disorder (revision). *Am J Psychiatry* 157:1–45
- Anderson IM (2001) Meta-analytical studies on new antidepressants. *Br Med Bull* 57:161–178
- Ballus C, Quiros G, De Flores T, de la Torre J, Palao D, Rojo L, Gutierrez M, Casais L, Riesgo Y (2000) The efficacy and tolerability of venlafaxine and paroxetine in outpatients with depressive disorder or dysthymia. *Int Clin Psychopharmacol* 15:43–48
- Barbui C, Cipriani A, Brambilla P, Hotopf M (2004) “Wish bias” in antidepressant drug trials? *J Clin Psychopharmacol* 24:126–130
- Bielski RJ, Ventura D, Chang CC (2004) A double-blind comparison of escitalopram and venlafaxine extended release in the treatment of major depressive disorder. *J Clin Psychiatry* 65:1190–1196
- Borenstein M, Hedges L, Higgins J, Rothstein H (2005) *Comprehensive meta-analysis*, version 2. Biostat, Englewood, NJ
- Cipriani A, Geddes J, Barbui C (2007) Venlafaxine for major depression. More evidence that risks outweigh benefits for most patients? *BMJ* 334:215–216
- Clerc GE, Ruimy P, Verdeau-Pailles J (1994) A double-blind comparison of venlafaxine and fluoxetine in patients hospitalized for major depression and melancholia. *Int Clin Psychopharmacol* 9:139–143
- Costa e Silva J (1998) Randomized, double-blind comparison of venlafaxine and fluoxetine in outpatients with major depression. *J Clin Psychiatry* 59:352–357
- Danish University Antidepressant Group (1986) Citalopram: clinical effect profile in comparison clomipramine: a controlled multicenter study. *Psychopharmacology* 90:131–138
- Danish University Antidepressant Group (1990) Paroxetine: a selective serotonin reuptake inhibitor showing better tolerance, but weaker antidepressant effect than clomipramine in a controlled multicenter study. *J Affect Disord* 18:289–299
- Dierick M, Ravizza L, Realini R, Martin A (1996) A double-blind comparison of venlafaxine and fluoxetine for treatment of major depression in outpatients. *Prog Neuropsychopharmacol Biol Psychiatry* 20:57–71
- Egger M, Smith GD, Altman DG (2001) *Systematic reviews in health care: meta-analysis in context*. BMJ Books, London
- Einarson TR, Arikian, SR, Casciano J (1999) Comparison of extended-release venlafaxine, selective serotonin reuptake inhibitors, and tricyclic antidepressants in the treatment of depression: a meta-analysis of randomized controlled trials. *Clin Ther* 21:296–308
- Fochtman LJ, Gelenberg A, Fochtman LJ (2006) Guideline watch: practice guideline for the treatment of patients with major depressive disorder, 2nd edition. Available at: [http://www.psych.org/psych\\_pract/treatg/pg/MDD.watch.pdf](http://www.psych.org/psych_pract/treatg/pg/MDD.watch.pdf). Accessed October 6, 2006
- Frank E, Prien RF, Jarrett RB (1991) Conceptualization and rationale for consensus definitions of terms in major depressive disorder. Remission, recovery, relapse, and recurrence. *Arch Gen Psychiatry* 48:851–855
- Freemantle N, Calvert M (2007) Weighing the pros and cons for composite outcomes in clinical trials. *J Clin Epidemiol* 60:658–659
- Furukawa TA, Barbui C, Cipriani A, Bambilla P, Watanabe N (2006) Imputing missing standard deviations in meta-analyses can provide accurate results. *J Clin Epidemiol* 59:7–10
- Geddes JR, Freemantle N, Mason J (2006) Selective serotonin reuptake inhibitors (SSRIs) versus other antidepressants for depression (Cochrane Review). In: *The Cochrane Library*, Issue 3, 2006. Update Software, Oxford
- Geerts S, Lacante P, Mignon A (1999) Venlafaxine versus fluoxetine in the treatment of depressed patients with concomitant anxiety. *Eur Neuropsychopharmacol* 9:S224
- Hansen RA, Gartlehner G, Lohr KN (2005) Efficacy and safety of second-generation antidepressants in the treatment of major depressive disorder. *Ann Intern Med* 143:415–426
- Higgins JPT, Thompson SG, Deeks JJ (2003) Measuring inconsistency in meta-analyses. *BMJ* 327:557–560
- Jorgensen AW, Hilden J, Gotzsche PC (2006) Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: systematic review. *BMJ* 333:782

- Kavirajan H (2004) Venlafaxine and SSRI remission data revisited. *Br J Psychiatry* 184:452–453
- Keller MB (2003) Past, present, and future directions for defining optimal treatment outcome in depression: remission and beyond. *JAMA* 289:3152–3160
- Kennedy SH, Andersen HF, Lam RW (2006) Efficacy of escitalopram in the treatment of major depressive disorder compared with conventional selective serotonin reuptake inhibitors and venlafaxine XR: a meta-analysis. *J Psychiatry Neurosci* 31:122–131
- Kessler RC, Demler O, Frank RG (2005) Prevalence and treatment of mental disorders, 1990 to 2003. *N Engl J Med* 352:2515–2523
- Khan A, Leventhal RM, Khan SR, Brown WA (2002) Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. *J Clin Psychopharmacol* 22:40–45
- Lieberman JA, Greenhouse J, Hamer RM, Krishnan KR, Nemeroff CB, Sheehan DV, Thase ME, Keller MB (2005) Comparing the effects of antidepressants: consensus guidelines for evaluating quantitative reviews of antidepressant efficacy. *Neuropsychopharmacology* 30:445–460
- MacGillivray S, Arroll B, Hatcher S, Ogston S, Reid I, Sullivan F, Williams B, Crombie I (2003) Efficacy and tolerability of selective serotonin reuptake inhibitors compared with tricyclic antidepressants in depression treated in primary care: systematic review and meta-analysis. *BMJ* 326:1014
- Machado M, Iskedjian M, Riuz I, Einarson TR (2006) Remission, dropouts, and adverse drug reaction rates in major depressive disorder: a meta-analysis of head-to-head trials. *Curr Med Res Opin* 22:1825–1837
- McPartlin GM, Reynolds A, Anderson C, Casoy J (1998) A comparison of once-daily venlafaxine XR and paroxetine in depressed outpatients treated in general practice. *Prim Care Companion J Clin Psychiatry* 4:127–132
- Mehtonen O-P, Sogaard J, Roponen P, Behnke K (2000) Randomized, double-blind comparison of venlafaxine and sertraline in outpatients with major depressive disorder. *J Clin Psychiatry* 61:95–100
- Moher D, Jones A, Cook DJ (1998) Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analysis? *Lancet* 352:609–613
- Moja LP, Telaro E, D'Amico R (2005) Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality cross sectional study. *BMJ* 330:1053
- Montgomery SA, Husson AKT, Bothmer J (2004) A randomised study comparing escitalopram with venlafaxine XR in primary care patients with major depressive disorder. *Neuropsychobiology* 50:57–64
- National Institute for Health and Clinical Excellence (2004) Management of depression in primary and secondary care. National Clinical Practice Guideline Number 23. Available at: <http://www.nice.org.uk/CG023NICEguideline>. Accessed October 13, 2006
- Nemeroff CB, Thase ME (2007) A double-blind, placebo-controlled comparison of venlafaxine and fluoxetine treatment in depressed outpatients. *J Psychiatr Res* 41:351–359
- Paykel ES (2002) Achieving gains beyond response. *Acta Psychiatr Scand Suppl* 106:12–17
- Roose SP, Schatzberg AF (2005) The efficacy of antidepressants in the treatment of late-life depression. *J Clin Psychopharmacol* 25: S1–S7
- Rudolph RL (2002) Achieving remission from depression with venlafaxine and venlafaxine extended release: a literature review of comparative studies with selective serotonin reuptake inhibitors. *Acta Psychiatr Scand Suppl* 106:24–30
- Rudolph RL, Feiger AD (1999) A double-blind, randomized, placebo-controlled trial of once-daily venlafaxine extended release (XR) and fluoxetine for the treatment of depression. *J Affect Disord* 56:171–181
- Rudolph RL, Fabre LF, Feighner JP, Rickels K, Entsuah R, Derivan AT (1998) A randomized, placebo-controlled, dose-response trial of venlafaxine hydrochloride in the treatment of major depression. *J Clin Psychiatry* 59:116–122
- Salinas E, Venlafaxine-XR-Study-Group (1997) Once-daily extended release (XR) venlafaxine versus paroxetine in outpatients with major depression. *Biol Psychiatry* 42(Suppl 1):297S
- Schatzberg A, Roose S (2006) A double-blind, placebo-controlled study of venlafaxine and fluoxetine in geriatric outpatients with major depression. *Am J Geriatr Psychiatry* 14:361–370
- Shelton C, Entsuah R, Padmanabhan SK (2005) Venlafaxine XR demonstrates higher rates of sustained remission compared to fluoxetine, paroxetine or placebo. *Int Clin Psychopharmacol* 20:233–238
- Shelton RC, Haman KL, Rapaport MH, Kiev A, Smith WT, Hirschfeld RM, Lydiard RB, Zajecka JM, Dunner DL (2006) Randomized, double-blind, active-control study of sertraline versus venlafaxine XR in major depressive disorder. *J Clin Psychiatry* 67:1674–1681
- Silverstone PH, Ravindran A (1999) Once-daily venlafaxine extended release (XR) compared with fluoxetine in outpatients with depression and anxiety. *J Clin Psychiatry* 60:22–28
- Sir A, D'Souza RF, Uguz S, George T, Vahip S, Hopwood M (2005) Randomized trial of sertraline versus venlafaxine XR in major depression: efficacy and discontinuation symptoms. *J Clin Psychiatry* 66:1312–1320
- Smith D, Dempster C, Glanville J, Freemantle N, Anderson I (2002) Efficacy and tolerability of venlafaxine compared with selective serotonin reuptake inhibitors and other antidepressants: a meta-analysis. *Br J Psychiatry* 180:396–404
- Solomon DA, Keller MB, Leon AC, Mueller TI, Lavori PW, Shea MT (2000) Multiple recurrences of major depressive disorder. *Am J Psychiatry* 157:229–233
- Thase ME, Entsuah AR, Rudolph RL (2001) Remission rates during treatment with venlafaxine or selective serotonin reuptake inhibitors. *Br J Psychiatry* 178:234–241
- Thiessen Philbrooke H, Barrowman N, Garg AX (2007) Imputing variance estimates do not alter the conclusions of a meta-analysis with continuous outcomes: a case study of changes in renal function after living kidney donation. *J Clin Epidemiol* 60:228–240
- Tylee A, Beaumont G, Bowden MW, Reynolds A (1997) A double-blind, randomized, 12-week comparison study of the safety and efficacy of venlafaxine and fluoxetine in moderate to severe major depression in general practice. *Prim Care Companion J Clin Psychiatry* 3:51–58
- Tzanakaki M, Guazzelli M, Nimatoudis I, Zissis NP, Smeraldi E, Rizzo F (2000) Increased remission rates with venlafaxine compared with fluoxetine in hospitalized patients with major depression and melancholia. *Int Clin Psychopharmacol* 15:29–34
- Zanardi R, Franchini L, Serretti A (2000) Venlafaxine versus fluvoxamine in the treatment of delusional depression: a pilot double-blind controlled study. *J Clin Psychiatry* 61:26–29
- Zimmerman M, Posternak MA, Chelminski I (2005) Is the cutoff to define remission on the Hamilton Rating Scale for Depression too high? *J Nerv Ment Dis* 193:170–175