Annals of Clinical Psychiatry, 19[4]:215–220, 2007

ISSN: 1040-1237 print / 1547-3325 online
DOI: 10.1080/10401230701653179

informa
healthcare

Interpreting Antidepressant Clinical Trials

NORMAN SUSSMAN, MD
NYU School of Medicine, New York, NY, USA

Background. *Psychiatrists and other clinicians make decisions about antidepressant medications with little understanding on how to interpret the research literature.*
Methods. *Pertinent clinical literature is reviewed.*
Results. *The author reviews levels of evidence, study design, statistical significance, p values, defining outcomes, drop outs, and basic analytic strategies such as last observation carried forward and mixed-effects model repeated measures. Several recent clinical trials are dissected to illustrate these concepts.*
Conclusions. *Clinicians need to develop greater sophistication at interpreting research findings. No single study is definitive, and comparative antidepressant trials suffer from low statistical power.*

Keywords Response, Remission, LOCF, MMRM, Statistical significance

INTRODUCTION

To date, no regulatory body in the United States, Canada, United Kingdom, or European Union has sanctioned a claim that more patients are likely to respond or remit to one drug versus another. Nevertheless, most practitioners hold to a belief that some drugs are clearly more effective than others. A variety of factors influence a clinician's perception of efficacy. For example, it may reflect positive anecdotal experience with a drug, informal discussions with colleagues, CME and non-CME presentations, posters, abstracts, pharmaceutical sales and marketing activities, or reading pertinent studies from medical literature. Of these influences, clinical experience coupled with the critical reading of research findings should be the main considerations for antidepressant selection and use. Yet, most practicing psychiatrists lack the training in basic research design and statistical methodology. Most clinicians, in any event, are too busy to critically analyze study results. Consequently, their understanding of merits and limitations of potential pharmacological interventions are likely to be shaped by the opinions of experts or clever industry sponsored educational/marketing initiatives.

Ironically, at all levels in the process of molding perceptions of an antidepressant, parties cite the results of randomized clinical trials (RCTs)—studies as the basis for their beliefs. RCTs, after all, ostensibly represent the gold standard of contemporary evidence based medicine (EBM). Yet given the complexity of clinical trial design and conduct, and the reliance on statistics to determine the clinical relevance of study findings, can be easily misled.

Address correspondence to Norman Sussman, NYU School of Medicine, 550 1st Avenue, New York, NY 10016. E-mail: sussman01@aol.com

Evidence and Levels of Evidence

What constitutes good evidence? (1). Evidence is used to prove or disprove something. An individual typically cites it as the grounds for belief. So, when something is offered as proof, when research data are presented as evidence of differences in treatment effects, the evidence should be plain or clear. The randomized, placebo-controlled trial is used to avoid confounding factors in clinical research, as well other types of information reflecting on a treatment. If properly designed and conducted, an RCT is likely to be free of bias and is thus especially useful for examination of small or moderate effects. Randomized clinical trials always have the following features: Neither subjects nor researchers have any influence as to who goes into which group and neither patients nor researchers know who is in which group. Blinding greatly boosts a study's quality. But blinding is not always possible—one example of this would be a study that compares patients who've had surgery with those who have not.

Some clinicians mistakenly believe that there is a regulatory standard for RCTs, but not all RCTs are identical. They

have their limitations: they differ in their outcome measures, duration, and dosing, among other things. The results can be influenced by whether mean dosing and dose titration are clinically meaningful. The single most important limiting factor in every RCT done to date is that the number of subjects enrolled has been too small to permit. Based on the lack of size, to date, no study contrasting one new antidepressant with another has enrolled a sufficient number of patients to definitively determine if one is better than the other (2).

Study Design and Statistics

It is beyond the scope of this article, or the expertise of the author, to provide a comprehensive review of study design and statistical analysis. Rather, it will cite some examples of ways in which the outcomes of studies can be influenced by such factors as dosing, titration schedules, definitions of success (e.g., response and remission), and the way in which those who drop out prematurely are accounted for in the efficacy analysis. Just as important in creating perceptions is the way in which study authors present their findings (as reflected in the discussion or conclusions section of a paper), or whether the findings are even published or presented.

Null Hypothesis and Statistical Significance

How does one test for differences in efficacy of antidepressant drugs? What does significant really mean? These are basic questions. To answer the first question, you need to understand that the general assumption in statistics is that drugs are considered equal until proven otherwise. This is the so-called null hypothesis. It assumes that any observed effect based on sample results is due to some sampling or experimental error. A study should be designed overcome the random factors that might produce the appearance of a difference. The hypothesis exists so that that it can be rejected as an explanation for the results of the experiment. And a finding of statistical difference between drugs being tested represents rejection of the null hypothesis. In other words, rejection of the null hypothesis is to infer "statistical significance."

P Value

Typically, if statistical tests indicate that the P value is at or below a level of 0.01 or 0.05, the observed treatment effect is statistically significant (and thus, the null hypothesis is rejected). Statistical significance is usually reflected in the P value. Think of the "P" as standing for the probability that a difference between sample means is not due to chance. So, P = 0.05 implies a 5% chance of observing the reported difference if the populations studied had identical outcomes. A random sampling from identical populations would produce a smaller difference between groups 95% of the time and a larger difference 5% of the time. A threshold p value (alpha) of 0.05 is usually defined prior to the experiment. Higher p values ("trends") are *not* significant. Lower p values do not prove that the null hypothesis is true, just that it cannot be rejected.

Something may be statistically significant, but not be clinically significant. Even if $p < 0.05$, that is, a statistically significant value, it may or may not be clinically important. A well-designed and conducted RCT may have internal validity (it measures a drug effect it is intended to measure), but it may not provide external validity, meaning that the observed results can be generalized to the broader community. There are other, more sensitive ways to examine if a finding is meaningful. These include looking at effect size, odds ratio and confidence intervals.

Defining Outcome: Response and Remission

Clearly, the most compelling claim for any drug—in any therapeutic area—is that of better treatment outcome. While there are many meaningful ways to measure the outcome of antidepressant functioning, areas such as functioning, sense of well being, and productivity, the gold standard outcome measure in clinical trials is remission. In recent years, remission, defined as a HAM-D of 7 or less (or MADRS of 10 or less), has begun to replace response, defined as a 50% or more decrease, as a more meaningful indicator of how well a drug works. There is no guarantee that these outcome definitions are in fact being used in a study. Designers of clinical trials can opt to set the bar higher—for example defining remission as a HAM-D of 6 or less or MADRS of 8 or less—or lower—such as a HAM-D of 8 or MADRS of 12 as cutoffs. Obviously, the less improvement you need to meet that goal, the higher the remission rates that will be reported.

There are many ways to define successful treatment. In research terms these are called outcome measures. Outcome may be determined by looking at symptoms, functioning, and quality of life. Beyond the obvious clinical benefit of achieving remission, another reason that the use of a more sensitive endpoint, such as remission rates, is important is because it provides a more stringent measure of antidepressant efficacy than response. Further elaborations of outcome include time to response/remission, and sustained response/remission. Sustained response/remission, for example would identify those who achieve remission at some point and remain there until the study endpoint. There may be more important outcome measures than change in score on a rating scale.

How to Account for Drop Outs

Most short-term clinical trials of antidepressant efficacy last from 6 to 12 weeks. Most of these studies last 8 weeks. One of the major decisions to be made is how to calculate the benefit of treatment among those who prematurely discontinue as subjects (drop outs) and those who participate for the entire

duration specified by the protocol (completers). Three basic approaches are used, each yielding potentially different results. Often, the analysis that is presented is the one that reflects most favorably on a sponsor's agent.

Last Observation Carried Forward (LOCF)

The most widely used approach is based on the intent to treat model and considers as "evaluable" all patients who complete at least one phase of the study (one day, one visit, one week, etc). In most studies that use LOCF, patients are counted in the study as long as they show up and are evaluated for at least one post-baseline visit. Once that subject drops out, their score on the last assessment is carried forward, and inserted at all remaining points as if they were still participating in the trial. To an outside observer, they appear as phantom subjects. The main effect of the LOCF approach is that it punishes a comparator with higher discontinuation rates, especially when the drop outs occur early in treatment. In effect, those patients never have a chance to improve. LOCF thus favors treatments with better compliance/tolerability and may underestimate the value of a treatment due to the impact early or bothersome ongoing side effects.

Observed Cases

The opposite approach is called an observed cases analysis. This method considers only patients who completed the trial Measures treatment effectiveness. This approach may overestimate usefulness of treatment because it does not consider patients who cannot tolerate treatment. In those instances where the time course of drop outs and total early discontinuation rates are the same, both LOCF and observed cases efficacy results should look the same.

Mixed-Effect Model Repeated Measures (MMRM)

In the likelihood-based repeated measures analyses (MMRM), missing data (drop outs) are estimated on the basis of observed data. In other words, comparisons are made in simulated data and in data from other patients in the randomized clinical trial. Estimates of treatment group differences in mean change from baseline to endpoint are then made (3). In most scenarios, the efficacy results using MMRM fall somewhere between those derived using the observed cases or LOCF analyses.

Selective Reporting and Publication (Cherry Picking Studies)

Nothing subverts informed clinical decision making more than the longstanding practice, across all medical disciplines, to only publish, or at least to make known, the results of positive trials. A disproportionate percentage published studies only report positive findings. This is particularly true in cases where the study is funded by a manufacturer of that didn't fare well. Clearly, if only positive results are published or presented at a meeting, it obviously creates the impression that a drug is more effective than it really is. Selective reporting of trials distorts the body of evidence available for clinical decision-making.

RECENT COMPARISON STUDIES

In order to illustrate they types of potential misuse of clinical trials, several recent marketing activities on behalf of several antidepressants are discussed. In each instance, the result of either study design, statistical analysis, or presentation of the data resulted in an obvious message: "our drug is highly effective, possibly more effective than the competition product."

Remission Rates during Treatment with Venlafaxine or SSRIs

A 2001 paper (2) reported and discussed the findings of a pooled analysis of pivotal trials comparing venlafaxine and the SSRIs (mainly fluoxetine). The authors reported that venlafaxine, a serotonin-norepinephrine reuptake inhibitor (SNRI), was associated with higher rates of remission than treatment with selective serotonin reuptake inhibitors (SSRIs). The remission rates during treatment with SSRIs or venlafaxine in eight comparable randomized, double-blind studies of major depressive disorder were pooled to compare remission rates (Hamilton Rating Scale for Depression score <=7) during treatment with venlafaxine (n = 851), SSRIs (fluoxetine, paroxetine, fluvoxamine; n = 748) or placebo (four studies; n = 446).

They found that the remission rates were as follows

- Venlafaxine, 45% (382/851);
- SSRIs, 35% (260/748);
- Placebo, 25% (110/446)

The difference between the SSRIs and venlafaxine represented about a 30% advantage in favor of the latter. The authors also reported that the difference between venlafaxine and the SSRIs was significant at week 2, whereas the difference between SSRIs and placebo reached significance at week 4.

The manufacturer used these data as part of highly effective sales and marketing initiative that helped to establish a reputation for venlafaxine as being more effective than the SSRIs. Although there were many limitations and caveats to these findings, the article was widely cited promotionally as evidence of the superiority of venlafaxine over SSRIs. It was an effective campaign, and led to an increase in sales of venlafaxine. However, on March 26, 2004 (Reuters) the FDA warned the manufacturer about the use of material that claimed Effexor it outperforms competing antidepressants. It stated that the use of a meta-analysis could not serve as the basis for a claim of proven superiority. The FDA letter said that the claim that Effexor is better "has not been demonstrated by substantial evidence or substantial clinical experience," therefore, promotional claims in advertising material were deemed misleading.

The success of the venlafaxine campaign nevertheless led to remission rates becoming a more meaningful measure of a drug's efficacy in the minds of most clinicians, and it also prompted other pharmaceutical companies to offer studies of their own to demonstrate that they were as good or better than venlafaxine in producing remission. The examples and discussion that follow is not intended to be a commentary on the intrinsic effectiveness of the drugs in question. Instead, the studies presented are offered as case examples of how the design, interpretation, and dissemination of data from these studies involved varying degrees of what can only be described as information management. Offered below are some examples of how in a competitive setting, RTC findings can be used to produce misleading impressions about what a study did, or did not find.

Citalopram versus Fluoxetine: A Double-Blind, Controlled, Multicentre, Phase III Trial in Patients with Unipolar Major Depression Treated in General Practice

One of the first attempts to challenge the claims of superior remission rates for venlafaxine can be seen in the form of the advertisement for the brand citalopram (4). It consisted of an 8-week "head-to-head clinical study" between "CELEXA 20 mg and fluoxetine 20 mg." Other information in small print provides the number of subjects, severity criteria for entry in to the study, a MADRS score of 22 or higher, as well as the baseline scores for each group. The most striking information in the advertisement however is not that "that 'at week 8, there was no statistical difference between CELEXA and fluoxetine,' but that 75% of CELEXA treated patients showed complete recovery at week 8."

In the broader universe of antidepressant trials, a 75% "complete recovery rate" is truly remarkable. In almost all pivotal trials, the remission rates for SSRIs were in the 35% range. Without additional details of how the study was designed, or how the analysis was done, it is impossible to judge the credibility of the findings being presented here to influence potential prescribers.

Two clues as to how the 75% remission rate was achieved can be explained purely in terms of decisions made before the first patient was enrolled.

The definition of remission in this study was a MADRS score of 12 or less. A review of most studies using the MARS scale as a rating scale shows that a cutoff of 10 is the most commonly used definition of remission. Some have even argued that this score may not be equivalent to the HAM-D cutoff of 7. Nevertheless, deciding to use a less ambitious target as the standard for remission would predictably inflate the number of those achieving that goal. In addition, the advertisement does indicate that both groups had a mean entry score of just over 29 on the MADRS. In this case, the difference between the definition of response (at least a 50% decrease in the score) and remission (a score of 12 or less) becomes narrowed to 2 or 3 points. For example, a typical patient who met criteria for response would be at 14 points at endpoint, just 2 points away from being a remitter.

Another aspect of the study design that inflated the score was the use of a 2-week exclusion from the LOCF analysis of dropouts. The standard in this type of ITT analysis is one week. Not having a placebo control group in a study with such high remission rates is typical.

Efficacy and Tolerability of Controlled-Release and Immediate-Release Paroxetine in the Treatment of Depression

In an attempt to position itself as having comparable remission rates, the marketers of Paxil CR launched the new formulation of paroxetine by prominently showing the results of a 12-week comparison of Paxil CR, Paxil, and placebo (5). The most striking aspect of their printed promotional materials was the remission rate of 57%. To the informed reader, it was obvious that the way the results were discussed by the authors served to enhance the apparent efficacy of Paxil CR and imply that it matched venlafaxine in rates of remission. Specifically, the discussion in the published paper said:

> . . . considerable attention has been awarded to the pooled analysis of antidepressant remission by Thase et al., who reported remission rates of 25% for placebo, 35% for SSRIs, and 45% for venlafaxine. In the present study, we observed remission rates of 45% for paroxetine CR compared with 34% for placebo using the LOCF analysis . . . (5)

What reviewers of this study overlooked or ignored were the following. The authors of the study do not discuss the original Thase et al. paper as a point of information nor as a summary of the extant remission literature. They present it in comparative terms, so that readers can draw the conclusion that rates of remission for Effexor XR in the Thase et al. study, and the rates of remission for Paxil CR in the Golden et al. study were identical. Apart from the fact that rates from different studies should not be compared, there were major differences between the studies in terms of study details that were not emphasized, or even mentioned.

1. The paroxetine study was 12 weeks long, compared to the 6- to 8-week studies in the venlafaxine pooled analysis. So, the patients in the paroxetine study had an additional 4 to 6 weeks to achieve remission.
2. The final remission rates in both studies for the pooled venlafaxine population and those treated with Paxil CR were 45%. The authors write that "in the present study, we observed remission rates of 45% for paroxetine CR compared with 34% for placebo . . ." (5). Not specifically mentioned, however, is that the venlafaxine-placebo difference is 20%, while the paroxetine CR-placebo difference was 11%.

Only in the fine print do you read about the LOCF analysis. The graphic displays the paroxetine CR-friendly observed

cases analysis. Finally, there was no paroxetine-placebo difference at the endpoint. So this clinical trial, and the way its results were publicized, serves to illustrate that it is impossible to draw meaningful conclusions without access to all relevant information about the study, and, without having an understanding of the factors that can lead to misleading conclusions.

Escitalopram versus Venlafaxine in the Treatment of Depression

Two other examples of misleading representation of clinical trial findings involve studies comparing venlafaxine and escitalopram. The first study, descriptively titled A Flexible Dose Comparison of Escitalopram and Venlafaxine XR (6) was presented in 2002 and widely disseminated before the final paper was published in 2004. In this comparative trial conducted in primary care centers in Europe, outpatients aged 18–85 years with a major depressive episode (MADRS total score of 18 or greater), were randomized to receive 8 weeks of double-blind treatment with flexible doses of escitalopram (10 to 20 mg/day) or venlafaxine XR (75 to 150 mg/day). After a 1-week washout period, patients started treatment with 10 mg escitalopram or 75 mg venlafaxine XR. Based on clinical response, medication doses could be doubled (to 20 mg/day escitalopram or 150 mg/day venlafaxine XR) after 2 or 4 weeks of treatment. The primary efficacy measure was the change from baseline on the MADRS. In the observed case analysis, escitalopram was significantly more effective at week 2 but no statistically significant differences were observed at end point in the observed cases or LOCF analysis. The analysis shown in was included in the lecture kit for escitalopram and used in the poster, creating the impression that it worked better and sooner than venlafaxine.

Yet there was a basic aspect of the study that favored escitalopram. There are a number of ways that dosing can be compared in an RCT. An assessment of the dose range allowable in the study compared with the therapeutic dose range recommended by the manufacturer and a comparison of the mean doses as a percentage of the maximal dose range for each agent are useful ways to determine whether dosing was comparable.

This 8-week, randomized, double-blind study compared the efficacy and tolerability of escitalopram to that of venlafaxine XR in primary care patients with major depressive disorder. The efficacy of escitalopram (10–20 mg; n=148) was similar to venlafaxine XR (75–150 mg; n=145), based on mean change from baseline to week 8 in Montgomery and Åsberg Depression Rating Scale total score. In ad hoc analyses, escitalopram-treated patients achieved sustained remission significantly faster than did venlafaxine-treated patients. More venlafaxine-treated patients had nausea, constipation, and increased sweating ($p < 0.05$). When treatment was completed after 8 weeks, significantly more venlafaxine-treated patients had discontinuation symptoms ($p < 0.01$). Thus escitalopram treatment was similar to venlafaxine treatment with respect to efficacy and was better tolerated by patients in primary care.

There was a small number of patients–to-sites ratio. The failure to disclose the large number of cites is a major omission. Given the ratio of number of subjects in the study (44 sites in 8 countries with about 250 patients—to be confirmed and calculated), it further diminishes the ability to see treatment differences that might exist. The point again, is that it misleads the reader into thinking that the two treatments are equivalent, when in fact it obscures differences that might exist. It is also noteworthy that the 8-week remission rates of 69.9% (Escitalopram) and 69.7% (Venlafaxine XR) are higher rates than usually seen in clinical trials. The probable reasons, as in the previous example, are the absence of placebo arm and the use of a more "liberal" definition of remission (MADRS of 12).

The major limitation of the study, however, was the fact that all patients in the escitalopram arm were taking at least the therapeutic dose of that drug. On the other hand, the average dose of venlafaxine was lower than many clinicians consider adequate.

In order to study the two drugs in which it would be certain that both agents were used at their maximum doses, a second trial was performed. This was titled A Fixed Dose Comparison of Escitalopram and Venlafaxine XR (7). In this randomized, double-blind trial to assess the comparative efficacy, safety, and tolerability of escitalopram and venlafaxine XR at their highest recommended doses, patients with DSM-IV-defined major depression were titrated (in accordance with labeling information) to receive 20 mg/day escitalopram or 225 mg/day of venlafaxine XR. Following a 1 week, single-blind, placebo lead-in, patients received 8 weeks of double-blind treatment.

This study compared escitalopram and venlafaxine extended release (XR) in depressed outpatients at the highest doses recommended in the United States. In this randomized trial, patients (diagnosis of DSM-IV-defined major depressive disorder; baseline Hamilton Rating Scale for Depression score of >/= 20) received 1 week of single-blind placebo treatment, followed by 8 weeks of double-blind, fixed-dose treatment with either escitalopram or venlafaxine XR (rapidly titrated to 20 mg/day and 225 mg/day, respectively, in accordance with prescribing information). The venlafaxine XR group had a higher incidence of discontinuation due to adverse events (16.0% vs. 4.1%; $p < .01$).

Analysis of remission rates showed that patients in both groups showed no statistical difference. However, given the disproportionate dropout rate in the venlafaxine group in the first week of the study—mainly due to the rapid, forced increase in venlafaxine dosage—an LOCF analysis clearly stacked the results in favor of escitalopram.

The most important decision in the design of the study was the forced titration. The paper explains the decision away by stating that both drugs were "titrated in accordance with labeling information." In fact, the venlafaxine label says you may increase the dose after 4 days, but in practice this is rarely

done, and in most instances dosing begins at 37.5 mg, not the 75 mg used in the study. The paper thus accurately concludes that "results of this study indicate that, when titrated rapidly to their maximum recommended doses, escitalopram is at least as effective as venlafaxine XR and significantly better tolerated." The unanswered question in this study is how well these drugs would have compared had the titration not maximized the early intolerability of venlafaxine.

CONCLUSIONS

Until physicians become more sophisticated about the interpretation of research findings they will continue to be fooled by smoke-and –mirror campaigns. Yet, most practicing psychiatrists lack the training in research methodology and often are too busy to dig beneath the surface of study results furnished by or with the support of a manufacturer.

The fact that all drugs work equally well doesn't mean that some patients respond better to one drug and not another. Even among agents within the same class, such as the SSRIs, for example, for reasons that are not understood, these idiosyncratic patient preferential responses may be pronounced. Other considerations that may increase the likelihood of a drug working are its effects on comorbid disorders and its side effect profile. Findings from studies are, however, used to create an impression that their drug has demonstrated some superiority in terms of efficacy, speed of action, tolerability, safety, ease of use, potential drug interactions, or cost.

What can the interested reader or listener do to reduce the risk of being misled? Ask questions about the following:

- Source of Information: Is it a recognized journal?
- Sponsorship: Since the conduct of the study, the choice of which data are highlighted, and the meeting itself are invariably industry-funded, healthy skepticism is a good thing
- Study duration: One reason not to compare results of separate studies, in addition to possible differences in patient populations and investigators, is the range of study duration. Most antidepressant clinical trials last eight weeks, but others can be 6, 9, or 12 weeks long.
- Tables and Graphs: Should contain, at the very least, the findings of statistical significance, error bars, number of subjects in each arm, and an indication of which method is used in accounting for dropouts (mainly LOCF, OC, or MMRM). Also, clinicians should be alert for variations on the Y-Axis. When the axis starts at 0 or at some higher cutoff point it makes any changes in rating scores seem more impressive.
- Posters and Publications: Beware of posters. Posters are often meant to communicate the most basic information about a study. Much crucial detail is not included, and in many instances the posters do not reflect the final statistical analysis. Most importantly, posters are not rigorously reviewed for content. In the appropriate context, where they are push-pinned to a corkboard with the author of the poster available to answer questions, they can be useful.
- Size: How many subjects were in each arm?
- Design: Was the study open-label or was it an RTC? Was placebo included?
- Dosing: Flexible, fixed, forced titration?
- Definition of outcome (response, remission).
- How drop outs are handled (intent-to-treat, observed cases)?

Ultimately, it needs to be kept in mind that no single study Is definitive. To date, all comparative antidepressant trials suffer from the same limitation: They have low statistical power and have low "assay sensitivity" due to the use of use of equivalence designs. Finally, unless you request information from the medical information department of a pharmaceutical company, you might not know if the positive study results presented reflect the entirety of completed trials examining the same agents for the same indication in a similar population.

REFERENCES

1. Sackett DL, Rosenberg WM, Gray JA, Hayned RB, Richardson WS. Evidence- based medicine: What it is and what it isn't. *BMJ* 1996;312:71–72
2. Thase, M. E., Entsuah, A. R. & Rudolph, R. L. Remission rates during treatment with venlafaxine or selective serotonin reuptake inhibitors. *Br J Psychiatry* 2001;178:234–241
3. Mallinckrodt CH, Clark WS, David SR. Accounting for dropout bias using mixed effects models. *J Biopharm Stat* 2001 Feb-May; 11:9–21
4. Patris M, Bouchard JM, Bougerol T, Charbonnier JF, Chevalier JF, Clerc G, Cyran C, Van Amerongen P, Lemming O, Hopfner Petersen HE. Citalopram versus fluoxetine: A double-blind, controlled, multicentre, phase III trial in patients with unipolar major depression treated in general practice. *Int Clin Psychopharmacol*. 1996;11:129–36
5. Golden RN, Nemeroff CB, McSorley P, Pitts CD, Dube EM. Efficacy and tolerability of controlled-release and immediate-release paroxetine in the treatment of depression. *J Clin Psychiatry* 2002 Jul;63:577–84
6. Montgomery SA, Huusom AKT, Bothmer J. A randomised study comparing escitalopram with venlafaxine XR in primary care patients with major depressive disorder. *Neuropsychobiology* 2004;50:57–64
7. Bielski RJ, Ventura D, Chang CC. J *Clin Psychiatry* 2004 Sep; 65:1190–6