# Evidence-Based Somatic Treatment of Depression in Adults

Daniel Carlat, MD*

**KEYWORDS**
- Antidepressants • Selective serotonin reuptake inhibitors
- Serotonin–norepinephrine reuptake inhibitors
- Efficacy studies • STAR-D trial

This article reviews recent studies and controversies about the effectiveness of antidepressant medications for depression. These medications are used for treatment of a wide variety of nondepressive conditions. Most notably, selective serotonin reuptake inhibitors (SSRIs) and serotonin–norepinephrine reuptake inhibitors (SNRIs) are often prescribed for most anxiety disorders, and they are clearly effective in this sphere, with robust differences between active drugs and placebos in controlled clinical trials. As one studies the contentious literature on antidepressants for depression, it is important to not lose sight of the fact that the very term "antidepressant" has become anachronistic. These drugs are actually broad-spectrum emotional and physical pain-relieving agents with efficacy in depression, anxiety, insomnia, as well several nonpsychiatric pain syndromes such as migraines, irritable bowel syndrome, neuropathies, and fibromyalgia.

This review is divided into three broad topics: (1) the heated controversy regarding how one should interpret "efficacy studies" of antidepressants, that is, the standard placebo-controlled double-blind studies used by drug companies to gain Food and Drug Administration (FDA) approval; (2) "effectiveness studies," open-label studies enrolling types of patients who are excluded from efficacy studies, focusing on one enormous study in particular—the STAR-D trial; and (3) the evidence base that can guide clinicians in the choice of antidepressants for particular patients.

---

Disclosure: Dr Carlat has nothing to disclose regarding the topic of this article.
Department of Psychiatry, Tufts University School of Medicine, 800 Washington Street, Boston, MA 02111, USA
* Carlat Publishing, PO Box 626, Newburyport, MA 01950.
*E-mail address:* drcarlat@comcast.net

Psychiatr Clin N Am 35 (2012) 131–142
doi:10.1016/j.psc.2011.11.002 **psych.theclinics.com**
0193-953X/12/$ – see front matter 

## EFFICACY STUDIES OF ANTIDEPRESSANTS: ARE ANTIDEPRESSANTS SIMPLY PLACEBOS WITH SIDE EFFECTS?

Psychologist Irving Kirsch has published papers and a recent book (*The Emperor's New Drugs: Exploding the Antidepressant Myth*)[1] arguing that antidepressants work no better than placebos. Kirsch is neither an "antipsychiatrist" nor a scientologist. Instead, he is a well-respected professor of psychology with a keen understanding of research methodology, which is why his articles have led to serious debate among both professionals and patients who have a stake in depression treatment.

In a series of papers over a long career, Kirsch has taken a magnifying glass to the gold standard research trials in antidepressant research—the placebo-controlled double-blind studies. In a nutshell, he has argued that, when averaged together, such studies actually show only a very small benefit of active treatment, and even this small benefit can likely be explained as an artifact of the side effects of antidepressants.

Many practicing psychiatrists are tempted to discount this claim automatically. After all, patients regularly appear to improve dramatically on antidepressants. But the history of medicine is filled with examples of treatments that were widely assumed to be effective based on individual clinical experience, which turned out to have little effect whatsoever. Benjamin Rush, for example, who is commonly considered the "Father of American Psychiatry," believed strongly in the effectiveness of bloodletting and purging with ipecac to "cure" a range of psychiatric illnesses.[2] Such dramatic treatments likely did lead to improvements, but subsequent experience and study revealed that these treatments were nonspecific—and that any effectiveness was due to placebo factors such as expectancy and hopefulness.

It was not until the mid-1930s that medical researchers began to compare treatments with placebo controls—inactive treatments that were packaged to look identical to active treatments. Such trials were able to subtract out the placebo effect and therefore empirically prove that a supposedly effective treatment was, indeed, effective. The first fully randomized placebo-controlled trials in medicine were first performed by psychiatrists in the 1950s in studies of thorazine and lithium.[2] Eventually, the FDA formalized the importance of these methods, when in 1980 it adopted a requirement that any drug submitted for approval must include at least two "pivotal" trials, generally meaning placebo-controlled trials.

### *The Kirsch Findings*

More than a dozen antidepressants have been FDA approved within these strict guidelines, leading most people, both patients and physicians, to assume that these agents are genuinely effective. But in 1998, Kirsch and his colleagues published the first of many papers questioning these studies. His work culminated in a recent meta-analysis on all clinical trials submitted to the FDA in support of four antidepressants: fluoxetine, venlafaxine, nefazodone, and paroxetine.[3] In summarizing the results of all these trials, they found that the average improvement on the Hamilton Depression Scale (HamD) was 9.6 points for patients randomly assigned to medications versus a mean improvement of 7.8 points for patients assigned to placebo, yielding a mean drug–placebo difference of 1.8. This corresponded to an average effect size of 0.31, considered small.

Although acknowledging that there is a statistically significant benefit of antidepressants over placebos, Kirsch argued that this separation was not clinically significant. To put this in perspective, consider that the Hamilton Depression Scale used in most studies is a 17-item scale with a maximum score of 51 points. The higher the score is, the worse the depression. Mild depression is generally defined as a score

of 8 to 13, moderate depression, 14 to 18; severe depression, 19 to 23; and very severe depression greater than 23. The 2-point difference reported by Kirsch (rounded up from 1.8) may not seem like much. As Kirsch notes in his book, a 2-point difference on the Hamilton scale "can be obtained by no longer waking during the night, or by no longer waking early in the morning, or by being less fidgety during the interview, or by eating better." These appear, indeed, to be clinically rather trivial improvements. Kirsch also cites guidelines from the United Kingdom's National Institute for Health and Clinical Excellence (NICE), which considers a minimum of a 3-point difference on the HamD between drug and placebo as being clinically significant.

Nor is Kirsch the only researcher to report these small differences. Erick Turner (a psychiatrist and former FDA analyst) and colleagues conducted a larger analysis, looking at data on all 12 antidepressants approved by the FDA between 1987 and 2004 (this included 8 SSRIs, 2 SNRIs, Wellbutrin, and Serzone). Although Turner and colleagues did not calculate the mean differences in HamD scores (because these data were not available for all of the studies they analyzed), they did calculate the average effect size. They reported an effect size of 0.32, nearly identical with Kirsch's 0.31.[4] This number, too, falls well short of NICE guidelines, which set an effect size of 0.5 as being the minimum to be considered clinically significant.

### *Are Antidepressants Placebos with Side Effects?*

In his book *The Emperor's New Drugs* Kirsch goes even further in his skepticism of antidepressant efficacy. He believes that even the small, clinically questionable benefit of antidepressants that he calculated may be a meaningless artifact of the side effects of antidepressants. Placebo-controlled studies are based on how successfully patients and physicians are fooled into believing a placebo might be the active drug. But active drugs such as antidepressants often cause side effects that can clue patients to their treatment assignments. This risks "unblinding" the study, and might give the active treatment an unfair advantage, in the sense that patients might be more hopeful of improvement if they believe they are receiving a real drug. This would neutralize the scientific integrity of a placebo-controlled trial.

Indeed, there is some evidence that in clinical trials, patients respond better to antidepressants when they notice more side effects,[5] implying that studies may become unblinded and therefore less valid. Is it possible that antidepressants are seemingly more effective than placebos purely because they cause side effects? Are patients responding to side effects, and not to the antidepressant?

Another study analyzed the effect of so-called "active placebos," which are sometimes used in antidepressant trials. In such studies, patients assigned to placebo are given a drug such as atropine, which has no inherent antidepressant effect but causes anticholinergic effects such as dry mouth and constipation. If antidepressants work via their side effect alone, one would expect no difference between antidepressants with side effects and placebos with side effects. However, a review of nine such clinical trials did not find this. In this meta-analysis, the average antidepressant versus active placebo effect size was 0.39, larger than the antidepressant versus regular placebo effect size of 0.31 reported by Kirsch.[6]

### *Defining "Clinical Significance"*

It seems likely, therefore, that antidepressants do, in fact, exert a specific antidepressant effect—but what of the charge that their effects are so small as to be clinically negligible? In an editorial, Turner and Rosenthal reviewed Kirsch's data and discussed the appropriateness of his dependence on the United Kingdom's NICE's

criteria as a benchmark for clinical significant. They argued that NICE's cutoff of an effect size of 0.5 is arbitrary because "effect size" is a continuous measure. Even the father of effect sizes, Cohen (the effect size is sometimes called "Cohen's *d*"), acknowledged when he proposed effect size benchmarks that "the values chosen had no more reliable a basis than my own intuition."[7]

For the average clinician, the concepts of Hamilton depression scores and effect sizes are somewhat abstract. The outcomes that seem more clinically relevant are the response rate (the percentage of patients who show a 50% or more decrease in HamD score) and the remission rate (percentage of patients with final scores below 7).

In a recent comprehensive review of the literature, Levkovitz and colleagues analyzed all published studies from 1980 to 2009 of antidepressants for either major depressive disorder or dysthymia.[8] These studies included 28,807 patients randomized to medication versus 16,887 patients randomized to placebo. For major depressive disorder, the pooled response rate (RR) for antidepressants was 54.3% versus 37.9% for placebo, resulting in a number needed to treat (NNT) of 6.1. The NNT is a practical measure of the effectiveness of a treatment, and refers to the number of patients one would have to treat with a medication to yield one response beyond what would be produced by placebo. For dysthymia (for which there was far fewer data), the pooled RR for antidepressants was 52.4% versus 29.9% for placebo, for an NNT of 4.4.

These significant differences in response rates appear more impressive than the very small differences in HamD scores. One caveat of this study, however, is that the researchers reviewed only data that had been published. Drug companies must submit at least two positive studies to receive FDA approval, but they typically conduct many more to account for the possibility of negative studies—that is, studies showing no significant difference between drug and placebo. They are not required to publish negative studies, and the aforementioned research by Turner and colleagues[4] found that studies with positive outcomes were 12 times more likely to be published than studies with negative outcomes. Further, they found that 94% of published antidepressant trials were positive (drug superior to placebo), but when both published and unpublished trials were pooled, only 51% of all antidepressant trials were positive. Overall, they computed that selective publication inflated the apparent effect size of antidepressants by 32%. Nonetheless, even including the negative data, they reported that all 12 antidepressants were statistically superior to placebo.

### *Generalizability of Clinical Trials*

A serious critique of any meta-analysis that focuses on placebo-controlled trials is that such trials are not representative of the kinds of patients practitioners see in clinical settings. Controlled trials exclude a wide range of patients, including those with comorbid anxiety, substance abuse, bipolar disorder, psychosis, personality disorders, and suicidality. In addition, patients whose depression is too mild or too severe, or who have active medical problems, will also often be excluded. One study found that only 14% of patients presenting to a clinic for depression treatment would be enrolled in a typical randomized controlled trial.[9]

What is the implication of this lack of generalizability? Many of the patients barred from these studies are the patients clinicians see in their offices every day. These are patients with "messy" symptom pictures, some with multiple disorders and others with atypical symptoms that may not qualify for an official *Diagnostic and Statistical Manual* (DSM) category. There are few studies of antidepressant efficacy in such patients, but even in the absence of clear evidence, one must treat patients who seek help. Rightly or wrongly, many clinicians tend to use the results of randomized

controlled trials (RCTs) as rough indicators of a "signal" of antidepressant efficacy, and then assume that if drugs work in these rarified populations, they will presumably also work for their patients.

Researchers have called for companies to conduct studies more representative of actual patient populations, but it is unclear that companies will heed the call. The more heterogeneous the patient mix, the harder it is to demonstrate a clear difference between drug and placebo—and therefore the harder it is to recoup the millions of dollars spent on clinical trials.

### *Effect of Depression Severity*

A number of studies, including one conducted by Kirsch and colleagues, have indicated that antidepressants work most robustly for patients with the most severe depression. Kirsch found that the higher the baseline HamD score, the more antidepressants separated from placebo. However, he also reported the interesting finding that this greater benefit for more severely depressed patients was due more to a lower placebo response than a greater response to medication.[3]

A more recent study reported similar results, in which patients with a baseline HamD score of 25 or more benefited more from antidepressants than less severely depressed patients.[10]

## EFFECTIVENESS STUDIES OF ANTIDEPRESSANTS: THE STAR*D TRIALS

Thus far, this article has reviewed only studies that meet the "gold standard" of clinical trials—that is, randomized, double-blind, placebo-controlled. In 2006, the first results of the STAR-D (Sequenced Treatment Alternatives to Relieve Depression) trial were published. This was the largest clinical trial for depression ever conducted, as well as one of the few large trials performed without industry funding and thus with less potential for strategic biases in research design that might favor one drug over another.[11] STAR-D was designed as an "effectiveness" trial as opposed to an "efficacy" trial. The researchers recently provided an excellent explanation of the unique features of this type of trial:

> "STAR*D has key features that define it as an effectiveness trial. Design elements such as broadly inclusive selection criteria and enrollment of patients from primary and specialty settings and with multiple concurrent medical and psychiatric illnesses give STAR*D results high external validity. Comparison of STAR*D participants with the U.S. population highlights the generalizability. The racial-ethnic composition of the enrolled participants approximates that of the U.S. population on the basis of data from the 2000 Census, and the distribution of depressive severity seen in STAR*D participants is consistent with the spectrum reported by Kessler and colleagues in a nationally representative sample (10% mild, 38% moderate, 39% severe, and 13% very severe). Both facts suggest that the sample was representative of depressed patients in the United States."[12]

There were other elements of the STAR-D methodology that make the results unusually generalizable to real-world practice. For example, there was no placebo condition. As is true in real practice settings, both physicians and patients knew which drugs were being taken. Although patients were randomly assigned to different drugs in STAR-D, they were allowed some input into the kind of treatment they received.

On the other hand, these methodologic decisions may well have backfired on the study designers, as there is much debate regarding whether the STAR-D results have provided any information of true value to clinicians, as noted later.

**Table 1**
**Results of augmentation**

| Celexa plus . . . | Remission (HRSD-17, %) | Response (QIDS-SR-16, %) |
|---|---|---|
| BuSpar | 30.1 | 26.9 |
| Wellbutrin SR | 29.7 | 31.8 |

### *STAR-D Level 1: Monotherapy with Celexa*

In Level 1 of the trial, 4041 patients were enrolled and started to receive Celexa up to 60 mg/d for up to 12 weeks. After an average length of treatment of 10 weeks, and on an average dose of 41.7 mg/d, patients receiving Celexa had a response rate of 47% (as measured by a nonstandard outcome measure, the 16-item Quick Inventory of Depressive Symptomatology-Self-Report [QIDS-SR]) and a remission rate of 27% (as measured by the standard HAM-D). How does this compare to other antidepressant trials? As noted in the previous section, the average response rates in placebo-controlled trials is 54.3%, so in this population Celexa appeared to be somewhat less effective than might be expected. On the other hand, this was a more treatment-resistant population than patients enrolled in typical trials, so it is difficult to compare these numbers with other studies.

### *STAR-D Level 2: Augmentation Versus Switching*

While 30% of STAR*D patients remitted on Celexa alone, 70% did not. Of these remaining patients, many discontinued participation in the study for various reasons, but 1474 remained and were assigned to one of three different groups, based on patient preference: 565 to augmentation treatment, 727 to a switch strategy, and 182 to cognitive therapy (either alone or added to Celexa).

#### *Augmentation track*

Patients who chose augmentation were assigned to treatment with either Wellbutrin SR (279 patients, mean dose 267.5 mg/d) or BuSpar (286 patients, mean dose 40.9 mg/d), and remained on their Celexa. Remission took an average of 6.3 weeks for the Wellbutrin SR patients and 5.4 weeks for those assigned to BuSpar (not a statistically significant difference). The remission and response numbers are in **Table 1**.

Although both BuSpar and Wellbutrin SR augmentation produced about 30% remission, various aspects of effectiveness and tolerability, reproduced in **Table 2**, place Wellbutrin SR slightly ahead of BuSpar.

#### *Switch track*

Patients who chose the switching track were assigned to switch from Celexa to one of three agents: Effexor XR, Wellbutrin SR, or Zoloft. **Table 3** provides the results of the switching arm; there were no statistically significant differences between the treatments.

**Table 2**
**Wellbutrin SR Versus BuSpar for augmentation**

| | Reduction in QIDS-SR Depression Score (%) | Time Adhered to Treatment | Rate of Discontinuation due to Side Effects |
|---|---|---|---|
| Wellbutrin SR | 25.3 | 10.2 weeks | 12.5% |
| BuSpar | 17.1 | 9.2 weeks | 20.6% |

All differences are statistically significant.

**Table 3**
**Results of switching antidepressants in level 2 of STAR-D**

| Switch from Celexa to . . . | Remission (HRSD-17, %) | Response (QIDS-SR-16, %) |
|---|---|---|
| Effexor XR | 24.8 | 28.2 |
| Wellbutrin SR | 21.3 | 26.1 |
| Zoloft | 17.6 | 26.7 |

Finally, a relatively small number of patients (182) chose the cognitive behavior therapy arm of the study, with a remission rate of 31%.

### *STAR-D Levels 3 and 4: More Augmentation and Switching*

At Level 3, patients were offered the opportunity of continuing in the study through two more levels of treatment, though relatively few did so. There were no significant differences between switching patients to Remeron (mirtazapine) or nortriptyline, though patients augmented with triiodothyronine (T3) had marginally better results than those augmented with lithium and experienced fewer side effects. Finally, at heroic Level 4, a very small number of patients were randomized to either "California Rocket Fuel" (Effexor XR plus Remeron) or the monoamine oxidase inhibitor Parnate (tranylcypromine), with no significant outcome differences between these two agents.

### *A 67% Cumulative Remission?*

In a press release by the American Psychiatric Association announcing the first results of the STAR*D trial, the lead was "In STAR*D, the nation's largest depression treatment study, results indicate that 67 percent of patients who complete from one to four treatment steps can reach remission." But recent papers have closely analyzed these results and found that they were significantly inflated as a result of post hoc changes in the statistical plan. The two main changes were: (1) a decision to drop the original primary outcome variable, the HAMD remission rate, and to replace it with the QIDS, a patient-rated scale that yielded higher remissions; and (2) another post hoc decision to include the most mildly depressed patients in the final analysis (those with initial HamD scores of less than 14). Thus the actual cumulative remission rate is unclear, but is likely less than 40% according to the original analytic plan. Further, looking at the 12-month continuation care results, only 108 of the original 4041 (2.6%) survived the trial without dropping out or relapsing.[13]

### *Lessons of STAR-D*

The goal of STAR-D was to provide us with evidence-based guidelines for how to proceed with treatment when a patient does not respond to an initial SSRI trial at a robust dose continued for 2 to 3 months. What is the next step in treating such patients?

Before STAR-D, most experts on treatment resistant depression would answer this question with, "If the patient has responded at least partially to the first medication, add a second. But if your patient's symptoms have not budged at all, switch to a different treatment." This reasonable approach was based on cobbling together various small studies, mostly without control groups. The great hope was that NIMH-funded study would finally provide the answer.

But as the authors of STAR-D recently acknowledged:

> "The data collected did not allow direct comparison of the benefits of switching versus augmenting. Patient preferences were a part of the equipoise randomization strategy, and most patients preferred either augmentation or switching at level 2. Consequently, patient groups were not equivalent at the point of randomization at the beginning of level 2; the augmentation group at level 2 was somewhat less depressed than the group that switched."[12]

In other words, STAR-D's design was excellent for recruiting many real-world patients, but was poor for creating clinical guidelines. It does not indicate which patients should be switched to a different antidepressant or which patients augmented. But assuming that a decision is made to switch or to augment—does STAR-D help us to choose the best medication? In a one sense, it does because it provides the valuable information that *it does not really matter which medication one chooses*—they are all equally effective. One can reasonably switch patients to an SNRI, Wellbutrin, or another SSRI; and one can reasonably augment with either Wellbutrin or BuSpar (though Wellbutrin might be a better choice in terms of side effects).

For more treatment-resistant patients, we learn from STAR-D that switching to Remeron or nortriptyline is identical, and that augmenting with lithium or T3 are similar (though one might lean toward thyroid in terms of tolerability). Finally, "California Rocket Fuel" (Effexor XR plus Remeron) or Parnate will provide similar benefits for the most treatment resistant. In a sense, STAR-D enhances our confidence in a clinical truism that has lasted for decades—that all antidepressants are created equal.

### *Choosing an Antidepressant*

Thus far, we have found that gold standard placebo-controlled studies show a small advantage of antidepressants over placebo for patients with mild to moderate depression, and the advantage is larger and more convincing as the severity of depression deepens. In real patient populations with comorbidities and with depressive symptoms that may not qualify for a specific diagnosis, we have no clear guidance for how well antidepressants will work, though most clinicians assume that the "signal" of efficacy from RCTs implies that they will be efficacious for other patients.

In addition, the largest trial of antidepressants in history provided relatively little clinical guidance, but did indicate a lack of difference between different antidepressant choices, whether for augmentation or switching.

The following paragraphs consider the information obtained from head-to-head studies of antidepressants.

### *SNRIs Versus SSRIs*

For a time, the SNRI Effexor (venlafaxine) was thought to be somewhat more effective than SSRIs, based on a famous 2001 meta-analysis.[14] In this study, based primarily on studies comparing Effexor with Prozac (fluoxetine), Effexor registered a 45% remission rate versus a 35% rate for SSRIs, an NNT of 10. However, over time, more Effexor has been compared with more agents, and the magnitude of its advantage has narrowed. In 2008, a company-sponsored meta-analysis found that the NNT of venlafaxine over SSRIs is 17, meaning that a clinician would have to treat 17 patients with venlafaxine to find one additional patient who would not have responded to an SSRI.[15] Generally, any NNT above 10

is considered to be clinically insignificant, although, as for the NICE criteria for clinical significance, this is a judgment call.

The most recent meta-analysis was not funded by a manufacturer (it was funded as a health technology assessment by the German government), and was likely the most comprehensive review yet compiled, including both published and unpublished data of comparisons among duloxetine, venlafaxine, SSRIs, and tricyclic antidepressants (TCAs).[16] The analysis included 10 randomized controlled trials comparing Cymbalta to SSRIs, 31 trials comparing Effexor to SSRIs, and 11 trials comparing Effexor to TCAs.

The researchers found that response rates were statistically significantly higher for Effexor than for SSRIs—but by only a small margin (66.5% vs 61.3%). There was no advantage in terms of remission rates. Effexor did not differ from either Cymbalta or TCAs on response or remission, nor did Cymbalta differ from SSRIs or TCAs. Both Effexor (11.8% vs 9.2%) and Cymbalta (8.4% vs 5.8%) had higher rates of discontinuation due to adverse events than did SSRIs. Effexor's adverse event–related discontinuation rates were not different from rates on TCAs.

The authors of this German study reported that data from several additional unpublished studies had been included in a different meta-analysis sponsored by the manufacturer of Effexor.[15] When they asked for these data, Wyeth declined to release them, casting doubt on the validity of the earlier findings that Effexor had a higher remission rate than SSRIs.

Overall, the most recent findings imply that there is little to no efficacy advantage of SNRIs over SSRIs, at least for the selected patients enrolled in randomized clinical trials. Because of the worse tolerability of SNRIs, they should continue to be considered second-line agents.

### *Choosing Among the SSRIs*

Given that SSRIs continue to be our first-line antidepressants, are there any reasons to choose one over the other? Many head-to-head studies of antidepressants have been conducted over the years, and recently two groups of researchers have assembled these trials and have performed meta-analyses.

The two studies used different methods of choosing studies and doing their analyses, and they therefore came up with somewhat different conclusions. The first study was arguably the more careful of the two. It was commissioned by the Agency for Healthcare Research and Quality, and surveyed a wide range of studies using different potential methods for comparing drugs. In addition to analyzing 105 head-to-head studies, they also analyzed 66 placebo-controlled studies. They found that no single antidepressant was more efficacious than the others. If this was the extent of the findings, the study would have been of limited use.

But this is only where the interesting results began. Among the clinically useful results were the following:

- Mirtazapine has a significantly faster onset of action than citalopram, fluoxetine, paroxetine, or sertraline.
- Meta-analysis of 15 fair-quality studies indicates that venlafaxine is associated with a higher rate of nausea and vomiting than SSRIs as a class (33% vs 22%).
- Evidence from 15 fair-quality studies indicates that sertraline is associated with a higher incidence of diarrhea than bupropion, citalopram, fluoxetine, fluvoxamine, mirtazapine, nefazodone, paroxetine, or venlafaxine (11% vs 8%).

- Seven fair-quality trials indicate that mirtazapine leads to higher weight gain than citalopram, fluoxetine, paroxetine, or sertraline (0.8–3.0 kg after 6–8 weeks).
- A good-quality systematic review provides evidence that paroxetine and venlafaxine have the highest rates of the discontinuation syndrome; fluoxetine has the lowest (data not reported).
- Evidence from five fair-quality trials provide shows that bupropion causes significantly less sexual dysfunction than fluoxetine, paroxetine, or sertraline. Among SSRIs, paroxetine has the highest rates of sexual dysfunction. Overall, more than 50% report sexual dysfunction.[17]

Cipriani and coworkers took a different approach to their meta-analysis, limiting the articles reviewed to 117 head-to-head comparisons (and omitting trials comparing single drugs vs placebo, observational trials, and pooled trials, all of which were included in the paper by Gartlehner and colleagues). Although less comprehensive, this analysis allowed the researchers to produce rankings of different antidepressants

**Table 4**
**Second-generation antidepressants**

| Medication | Indications (Includes Long-Acting Forms) | Evidence-Based Features |
|---|---|---|
| Bupropion (Wellbutrin, Zyban, Generic) | MDD, seasonal affective disorder, smoking cessation | Less sexual dysfunction, no weight gain |
| Citalopram (Celexa, Generic) | MDD | |
| Desvenlafaxine (Pristiq) | MDD | |
| Duloxetine (Cymbalta) | MDD, diabetic peripheral neuropathy, fibromyalgia, GAD | Unique efficacy for pain in dispute |
| Escitalopram (Lexapro) | MDD, GAD | Particularly high efficacy/tolerability |
| Fluoxetine (Prozac, Sarafem, Generic) | MDD, OCD, bulimia nervosa, PMDD, panic disorder | Lower risk of discontinuation syndrome |
| Fluvoxamine (Luvox, Generic) | OCD | |
| Mirtazapine (Remeron, Generic) | MDD | High weight gain, somnolence; fast onset of action |
| Paroxetine (Paxil, Generic,) | MDD, OCD, panic, social phobia, GAD, PTSD | High sexual dysfunction, weight gain, discontinuation symptoms |
| Sertraline (Zoloft, Generic) | MDD, OCD, panic, PTSD, PMDD | Particularly high efficacy/ tolerability/low cost; high diarrhea |
| Venlafaxine (Effexor, Generic) | MDD, GAD | Efficacy advantage over SSRIs now disputed; high nausea/ vomiting, discontinuation symptoms |
| Vilazodone (Viibryd) | MDD | |

*Abbreviations:* GAD, generalized anxiety disorder; MDD, major depressive disorder; OCD, obsessive–compulsive disorder; PMDD, premenstrual dysphoric disorder; PTSD, posttraumatic stress disorder.

in terms of efficacy, tolerability, and economics. They concluded that venlafaxine, mirtazapine, sertraline, and escitalopram were slightly more effective than the eight other new-generation antidepressant medications examined. Among these, escitalopram and sertraline had the best tolerability, while sertraline was the most economical; sertraline therefore took the "grand prize" by scoring high in efficacy, tolerability, and price.[18] However, the methodology of this article has been critiqued because all of the trials were phase 4 trials sponsored by manufacturers, and it is possible that some drug companies were more skillful than others in designing trials to best their competitors. Nonetheless, the results seem intuitively consonant with the experiences of many clinicians.

## SUMMARY

The efficacy of antidepressants has become a contentious topic over the last decade, and yet a review of the literature shows that they are consistently more effective than placebo. Although the average magnitude of this effect is unclear, many individual patients respond well to a course of antidepressants, and relapse when the medication is discontinued.

Choosing the right antidepressant for a given patient remains more art than science, but the studies reviewed here provide some helpful guidance. **Table 4** lists the second-generation antidepressants along with potential reasons for choosing one over the other (based on side effects, costs, or possible therapeutic advantages.)

Based on these data, the following conclusions can be drawn:

- For an all-around first-line antidepressant, sertraline is hard to beat, given its combination of efficacy, tolerability, and low expense. Once escitalopram becomes generic, it will join sertraline in this category.
- Bupropion is often a first-line alternative to sertraline, because of its lack of sexual side effects; although it has less efficacy for anxiety disorders, it is helpful for other comorbidities, such as tobacco dependence and attention-deficit/hyperactivity disorder.
- Both paroxetine and mirtazepine are often maligned because of side effects of sedation and weight gain; however, these side effects may be advantageous for those whose depressive symptoms include insomnia and excessive weight loss.
- Although not specifically reviewed in this article, certain antidepressants are liable to cause more drug–drug interactions than others; the most prominent of these are fluoxetine, paroxetine, and fluvoxamine.

## REFERENCES

1. Kirsch I. The emperor's new drugs: exploding the antidepressant myth. Philadelphia: Basic Books; 2010.
2. Healy D. The antidepressant era. Cambridge (MA): Harvard University Press; 2000. p. 91.
3. Kirsch I, Deacon BJ, Huedo-Medina TB, et al. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. PLoS Med 2008;5:e45.
4. Turner EH, Matthews AM, Linardatos E, et al. Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med 2008;358:252–60.
5. Greenberg RP, Bornstein RF, Zborowski MJ, et al. A meta-analysis of fluoxetine outcome in the treatment of depression. J Nerv Ment Dis 1994;82:547–51.
6. Moncrieff J, Wessely S, Hardy R. Active placebos versus antidepressants for depression. Cochrane Database Syst Rev 2004;1:CD003012.

7. Turner E, Rosenthal R. Efficacy of antidepressants. BMJ 2008;336:516–7.
8. Levkovitz Y, Tedeschini E, Papakostas G. Efficacy of antidepressants for dysthymia: a meta-analysis of placebo-controlled randomized trials. J Clin Psychiatry 2011;72(4): 509–14.
9. Zimmerman M, Mattia JI, Posternak MA. Are subjects in pharmacological treatment trials of depression representative of patients in routine clinical practice? Am J Psychiatry 2002;159:469–73.
10. Fournier JC, DeRubeis RJ, Hollon SD et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. JAMA 2010;303(1):47–53.
11. Trivedi MH, Rush AJ, Wisniewski SR, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. Am J Psychiatry 2006;163:28–40.
12. Gaynes BN, Warden D, Trivedi MH, et al. What did STAR*D teach us? Results from a large-scale, practical, clinical trial for patients with depression. Psychiatr Serv 2009; 60:1439–45.
13. Pigott HE, Leventhal AM, Alter GS, et al. Efficacy and effectiveness of antidepressants: current status of research. Psychother Psychosom 2010;79:267–79.
14. Thase ME, Entsuah AR, Rudolph RL. Remission rates during treatment with venlafaxine or selective serotonin reuptake inhibitors. Br J Psychiatry 2001;178:234–41.
15. Nemeroff CB, Entsuah R, Benattia I, et al. Comprehensive Analysis of Remission (COMPARE) with Venlafaxine versus SSRIs. Biol Psychiatry 2008;63(4):424–34.
16. Schueler YB, Koesters M, Wieseler B, et al. A systematic review of duloxetine and venlafaxine in major depression, including unpublished data. Acta Psychiatrica Scand 2011; 23:247–65.
17. Gartlehner G, Gaynes BN, Hansen RA. Comparative benefits and harms of second-generation antidepressants: background paper for the American College of Physicians. Ann Intern Med 2008;149:734–50.
18. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. Lancet 2009; 373:746–58.