White Paper: Efficacy of Psychiatric Drugs

Jacqueline A. Sparks, Ph.D.

Psychiatric drugs have been a mainstay of treatment for a range of behavioral and emotional problems over the past six decades in the US and around the world. While their use has skyrocketed, the problems they purport to alleviate have not diminished but, in fact, have increased (Whitaker, 2010). It is reasonable to ask whether the most widely prescribed drugs are effective in treating the conditions they claim to treat.

This paper summarizes current evidence for the efficacy of three major classes of psychiatric drugs—antidepressants, antipsychotics, and stimulants—with separate examinations for adults and children. It relies on meta-analyses and reviews, generally considered the best sources for assessing large amounts of data, as well as major studies, those considered seminal based on size, duration, or design. The paper concludes that psychiatric drugs lack adequate efficacy and calls into question their widespread use.

**Antidepressants**

**Adults and Antidepressants**

In a meta-analytic review of nineteen studies involving 2,318 people, Kirsch and Sapirstein (1998), showed that 75% of the response to antidepressants was duplicated by placebo. They speculated that the remaining 25 percent of the positive antidepressant effect might be attributable to the un-blinding power of side effects. Kirsch, Moore, Scoboria, and Nichols (2002) analyzed the efficacy data submitted to the US Food and Drug Administration (FDA) for the six most widely prescribed antidepressants approved between 1987 and 1999. Approximately 82% of the response to medication was duplicated by placebo control groups—57% of the studies failed to show a drug-placebo difference. When a difference was found, the drug/placebo difference was an average of 1.8 points on the clinician-rated Hamilton Depression Rating Scale (HDRS). FDA memoranda suggested the clinical significance of such a small difference was questionable (Laughren, 1998).

In a review of antidepressant trials involving 12,564 persons (Turner, Matthews, Eftihia, Linardatos, Tell, & Rosenthal, 2008), 94% of published trials had favorable results whereas the percentage of positive results for published and unpublished trials together drops to 51%. The authors warn that publication bias of this magnitude dramatically distorts reported effect sizes and has serious implications for researchers, health care professionals, and clients. Kirsch et al. (2008) meta-analytically examined all trials submitted to the FDA for the licensing of four popular SSRIs and found no clinically significant differences between placebo and the drugs, with the exception of the most distressed in the severely depressed group. The negligible difference in this group was found to be due not to the drug, but to a decreased response to placebo.
STAR*D (Sequenced Treatment Alternatives to Relieve Depression) (Rush et al., 2004), a 6-year, $35 million National Institute of Mental Health (NIMH)-funded study with nearly 2,900 participants at Level 1, examined the impact of sequenced augmentation or drug switching strategies on depression when a traditional regimen of a single SSRI failed. STAR*D was an un-blinded, non-placebo-controlled trial designed to simulate conditions faced in daily practice. The sample, however, did not represent a general clinical population since it excluded those with a history of intolerance or non-response to any SSRI and included only those who preferred a medication intervention. Due to the lack of a placebo and double blind, the authors acknowledge “Nonspecific treatment effects [e. g., the expectation of improvement] undoubtedly accounted for some unknown proportion of the acute response or remission rates” (Trivedi et al., 2006a, p. 37).

In the STAR*D, the average remission rate based on the primary outcome measure was 28% and 25% on the first two levels, and 14% and 13% on the last two—unimpressive considering the typical 30% placebo response in antidepressant trials (Thase & Jindal, 2004). At Level 1, 28% experienced moderate to intolerable side effects (Trivedi et al., 2006a). At Level 2 (participants augmented or switched), 51% experienced side effects ranging from moderate to intolerable (Rush et al., 2006a; Trivedi et al., 2006b). Data from the 12-month follow-up of those who either remitted or responded indicated a relapse rate of 58% (Rush et al., 2006b).i

The STAR*D was confounded by methodological irregularities. Researchers included ineligible mildly-depressed participants in the analysis, changed the primary outcome mid-stream, and cherry-picked data post-trial (see, https://www.psychologytoday.com/blog/mad-in-america/201008/the-stard-scandal-new-paper-sums-it-all). The NIMH website statement that 70% of those who completed the trial achieved remission is based on a projected response had no one dropped out. Taking these factors into account, the remission figure for those who continued through follow-up is 3%.

Reviews have demonstrated no advantage for combining psychotherapy and antidepressant treatments (e.g., Antonuccio, Danton, & DeNelsky, 1995), but Thase et al. (1997) found that combining the two offered some added benefit for the minority suffering with severe, recurrent depressions. Support for a combined regimen for more chronic depressions is also found in the Keller et al. (2000) trial. The combined group improved more than the medication or psychotherapy groups at 12 weeks. Results were weakened by the lack of a placebo control group and the use of only a single clinician-rated outcome measure.ii

Children and Antidepressants

Several large trials are often cited as evidence justifying child psychotropic prescription. For example, two randomized, placebo-controlled trials of fluoxetine (Prozac) (Emslie et al., 1997; Emslie et al., 2002) gained FDA approval for Prozac for youth aged 8-17 diagnosed with depression (FDA, 2003, January 3). However, both Emslie studies failed to find a statistical difference between Prozac and placebo on primary outcome measures.iv
The NIMH funded Treatment of Adolescent Depression Study (TADS) (TADS Team, 2004), again evaluated Prozac for the youth age group. TADS compared the efficacy of four treatment conditions: Prozac alone, cognitive behavioral therapy (CBT) alone, CBT plus Prozac, and placebo. Despite media claims, (The New York Times front page headline, “Antidepressant Seen as Effective in Treatment of Adolescents,” Harris, 2004), the FDA did not count TADS as a positive study for SSRIs due to the negative findings on its primary outcome measure. Other end-point comparisons in TADS favored the combined medication/CBT arm. However, treatment was unblind, and only the combined group received all intervention components (drug, psychotherapy, psychoeducation and family therapy, and supportive pharmacotherapy monitoring), creating a significant disparity in favor of the combination arm.

In the long-term (36-week) TADS efficacy study, partial and non-responders to placebo, and responders and partial responders to Prozac, CBT, and combination treatments in the 12-week trial were openly treated (The TADS Team, 2007). As in phase 1, Prozac and combination groups received additional encouragement and contact (medication management). Despite this, all treatment conditions converged by 30 weeks and remained so by week 36.

Jureidini et al. (2004) questioned the clinical significance of results that show no gains on primary or client/parent-rated measures in youth antidepressant trials and highlight other design weaknesses, including reliance on the last observation carried forward, an emphasis on secondary endpoints, and transforming continuous into categorical outcomes thereby inflating small differences. Moreover, publication bias—studies finding in favor of the investigative drug are published whereas unfavorable studies are not—results in over-estimations of antidepressant efficacy for pediatric use. An independent analysis by the FDA concluded that only 3 out of 15 published and unpublished trials of SSRIs showed them to be more effective than placebo on primary outcome measures (Laughren, 2004). None of the 15 found differences on client or parent-rated measures.

**Antipsychotics**

Adults and Antipsychotics

In the largest study of antipsychotics to date, the NIMH funded Clinical Antipsychotic Trials of Intervention (CATIE) (Lieberman et al., 2005), the primary outcome measure was discontinuation of treatment for any reason (not clinical improvement or remission). CATIE enrolled 1,400 participants at 57 US sites and used a triple blind—clinicians, raters, and participants did not know which drug participants were taking. CATIE had no placebo group, allowed clinicians to make flexible dosing decisions, and permitted multiple additional drugs (excluding antipsychotics). The goal of CATIE was to evaluate how well second generation antipsychotics (SGAs) (olanzapine—Zyprexa, quetiapine—Seroquel, resperidone—Risperdal) compared with one another and a FGA (perphenazine—Etrafon) in real world conditions.
Seventy four percent (74%) of CATIE participants discontinued before 18 months, largely due to inefficacy and intolerable side effects (Lieberman et al., 2005). The authors note that these rates are consistent with those observed in previous antipsychotic drug trials. Psychosocial functioning improved only modestly for the one third of CATIE participants who reached the primary Quality of Life Scale endpoint at 12 months (Swartz et al., 2007). Rates of moderate to severe adverse events revealed through systematic inquiry ranged from 42 to 69% (Zyprexa the worst) (Stroup et al., 2007). Hospitalization rates ranged from 11 to 20% over the study period, while a weight gain of over 7% occurred in 14 to 36% of participants (Zyprexa worst). The lead author of the CATIE studies admitted: "... the claims of superiority for [SGAs] were greatly exaggerated. This may have been encouraged by an overly expectant community of clinicians and patients eager to believe in the power of new medications. At the same time, the aggressive marketing of these drugs may have contributed to this enhanced perception of their effectiveness in the absence of empirical information" (Lieberman, 2006, p. 1070).

The Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD), another major investigation funded by the NIMH, examined the effectiveness of SGAs and anticonvulsants for persons diagnosed with bipolar disorder (Sachs et al., 2003). In one of two outcome reports, 30% experienced no recurrences of symptoms (Perlis et al., 2006); the second (Nierenberg et al., 2006) found lower rates of recovery (just under 15%). Results of the Work and Social Adjustment Scale evaluated during a period of remission revealed “considerable functional impairment” (Fagiolini et al., 2005, p. 284). Similar to CATIE findings, remission from clinically defined symptoms, even for the few who achieved this, did not mean adequate social functioning.

Children and Antipsychotics

The American Psychological Association Working Group on Psychoactive Medications for Children and Adolescents (APA Working Group) (2006) conducted a comprehensive investigation of the scientific literature related to pediatric antipsychotic use and found that studies supporting the use of antipsychotics to treat children contained significant methodological limitations including small sample sizes, open trials, and lower tier evidence (e.g., retrospective chart reviews and case reports) (APA Working Group on Psychoactive Medications for Children and Adolescents, 2006).

The FDA approved Risperdal for the treatment of children diagnosed with autism based on two 8-week trials and one 6-mos. open-label safety trial. Risperdal was approved for adolescents aged 13 -17 diagnosed with schizophrenia based on one 6 week trial and one unblind, 2-dose 8-week trial. One 3-week trial was conducted that served as the basis for FDA approval of Risperdal for children and teens aged 10-17 diagnosed with bipolar I disorder. Of the adolescents in this study, 36% were enrolled due to manic episodes; the remaining 64% were described as experiencing a behavior disorder—fifty percent (50%) had a diagnosis of ADHD. All of the pediatric Risperdal trials were sponsored by Janssen (maker of Risperdal).
Aripiprazol (Abilify) was approved for use by adolescents diagnosed with schizophrenia aged 13-17 based on a 6-week trial. It was approved for youth aged 10-17 diagnosed with bipolar I as a result of a 4-week trial. The evidence for approval of Zyprexa for adolescents diagnosed with schizophrenia was a 6-week trial; bipolar I, ages 13-17, 3-week trial. Finally, Seroquel was approved for acute bipolar mania, ages 10-17, based on a 3-week trial.

The NIMH-funded Treatment of Early Onset Schizophrenia Spectrum Disorders (TEOSS) (Sikich et al., 2008), considered a landmark trial, compared the efficacy, tolerability, and safety of two second generation antipsychotics (risperidone or Risperdal and olanzapine or Zyprexa) to a first generation antipsychotic (molindone or Moban) for youths, ages 8-19, diagnosed with early-onset schizophrenia spectrum disorder. At the end of eight weeks, the liberally defined response rate was 50% for those treated with Moban, 46% for Risperdal, and 34% for Zyprexa. Of the 116 participants in the acute phase, 46 (41%) withdrew due to adverse effects or inadequate efficacy. Sikich, lead author, described these short-term outcomes as indicative of limited efficacy for these drugs for the targeted problem and age group.

The short duration of these trials along with the significant ties of authors to the pharmaceutical industry must serve as qualifiers in assessing the studies' findings. For example, the 6-week Risperdal trial was funded by Johnson and Johnson, maker of the investigated drug. Six of seven authors were either employees of Johnson and Johnson or had significant ties to the company. In the 8-week Risperdal trial, all eight authors were employees of Johnson and Johnson. Seven of 8 study authors in the 3-week Risperdal bipolar trial were Johnson and Johnson employees and owned stock in the company. The remaining author received research support from Johnson and Johnson, consulted for the company, and was on its speakers bureau and advisory board.

Concerning the 6-week Abilify schizophrenia trial, 7 of the 10 authors were employees of Otsuka, maker of Abilify. One author was an employee of Bristol-Myers Squibb, U.S collaborator for development and commercialization of Abilify, and first author (Findling) and second author (Robb) received research support from and/or was a consultant for Otsuka. Finally, in the 8-week trial of Abilify for schizophrenia, 8 of 11 authors were employed by Eli Lilly, maker of Zyprexa and 3 others served as consultants for Eli Lilly. Studies have found a direct correlation between who funds a study and its outcome. For example, Heres et al. (2006) looked at published comparisons of five antipsychotic medications. In 9 out of 10 studies, the drug made by the company that sponsored the study was found to be superior. The findings from studies leading to approval Risperdal, Abilify, and Zyprexa for various indications in the youth population must be evaluated in light of conflicts of interest.

Additionally, the pediatric antipsychotic trials mentioned in this review contain design parameters that favor the investigative drug. For example, in the 6-week Abilify trial, subjects who experienced unacceptable dose-related tolerability problems before study day 25 were removed from the study. Eleven percent of the Zyprexa group in the 6-week Zyprexa trial had
responded well to the drug in previous use; those who had not responded were excluded. In this same trial, patients unable to tolerate the minimum dose and those who did not respond to treatment after 3 weeks were discontinued from the study. Use of such strategies weaken the validity of study findings, specifically claims of efficacy for the investigated drug.

Stimulants

Children and Stimulants

A review of forty years of trials supporting stimulant prescription (primarily methylphenidate [Ritalin]), found overall effect sizes in the moderate range (Conners, 2002). Effect sizes for academic productivity were low to moderate and in the zero range for academic achievement. The APA Working Group (2006) cited stimulant research limitations, including lack of data supporting long term efficacy or safety. Similarly, they noted that stimulant drugs, while often succeeding in lessening symptoms associated with ADHD diagnosis, show minimal efficacy in general life domains of the child, including social and academic success.

The Multimodal Treatment Study of Children with ADHD (MTA) (MTA Cooperative Group, 1999), the largest, most complexly designed trial of interventions for ADHD, is frequently cited as evidence of the superiority of stimulants over behavioral approaches. In this study, 3 of 19 measures (un-blinded) found differences favoring Ritalin. Blinded classroom observers, participant children, and the child peers failed to rate medication better than the behavioral interventions. At 14-months, assessments compared those actively medicated and those who had ended therapy; un-blinded measures found an advantage for Ritalin. Importantly, the behavioral group had ended therapy, with the last face-to-face therapeutic contact 4 to 6 months prior to assessment (Pelham, 1999). At 24-month follow-up, medication and combined groups lost much of their effect (up to 50%) while behavioral treatment and community groups retained theirs (MTA Cooperative Group, 2004). At 36 months, treatment groups did not differ significantly on any measure (Jensen et al., 2007).

The Preschool ADHD Treatment Study (PATS) investigated the efficacy and safety of Ritalin for preschoolers aged 3 to 5.5 (Greenhill et al., 2006). Twenty one percent (21%) of the children achieved MTA-defined criterion for remission.

Conclusions

Based on this review, the three classes of psychiatric drugs examined are weakly or not at all supported by meta-analyses or major trials. Findings in many studies are compromised by design flaws and conflicts of interest. Publication bias plays a role in what results are made known to academic, professional, and public domains. This review calls into question media reports and standard practices that presume the desirability or necessity of antidepressants, antipsychotics, and stimulants for a variety of diagnosed conditions and that promote their first-line use in many instances for adults and children.
References


Treatment for Adolescents with Depression Study (TADS) Team (2004). Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression. *JAMA, 292*(7), 807-820.


The TADS Team (2007). The Treatment for Adolescents with Depression Study (TADS): Long-term effectiveness and safety outcomes. *Archives of General Psychiatry, 64*(10), 1132-1144.


Portions of material in this document were previously published in Sparks, Duncan, Cohen, and Antonuccio (2010) and Sparks and Duncan (2012).

Various other psychotropic medications aimed to reduce SSRI-induced agitation or sexual dysfunctions were concomitantly prescribed to an unknown proportion of the participants.

The authors of this study, published in the *New England Journal of Medicine*, were heavily tied to the pharmaceutical industry. The editors stated “it would have used too much space to disclose them [financial ties to industry] fully in the Journal” (p. 1462). Additionally, the study’s investigative drug (nefazadone) has since been recalled, due to unacceptable liver toxicities.

Jureidini et al. (2004) reported that the first Emslie trial changed its primary outcome measure between the trial’s beginning and publication, using secondary measures to show superiority.

Researchers had significant ties to the pharmaceutical companies whose drugs were being investigated, including being stockholders.

These percentages must be understood in light of the study’s definition of response (Clinical Global Impression, CGI, score of at least 2, much improved, plus a ≥ 20% reduction in baseline on the Positive and Negative Syndrome Scale, PANNS). According to an analysis of cutoff and response scores for the PANSS, reduction of PANNS of ≥ 28% correlates with CGI “minimally improved.” The low cutoff on the PANSS in this trial calls into question the clinical meaningfulness of the response rates reported.