

A Comparison of the Efficacy of Medications for Adult Attention-Deficit/Hyperactivity Disorder Using Meta-Analysis of Effect Sizes

Stephen V. Faraone, PhD, and Stephen J. Glatt, PhD

Objectives: Medications used to treat attention-deficit/hyperactivity disorder (ADHD) in adults have been well researched, but comparisons among drugs are hindered by the absence of direct comparative trials. Our objectives were to (1) estimate the effect size of the medications used to treat adult ADHD, (2) determine if differences in the designs of studies confound comparisons of medication efficacy, (3) quantify the evidence for differences in effect sizes among medications, and (4) see if features of study design influence estimates of efficacy.

Data Sources: The following search engines were used: PubMed, Ovid, ERIC, CINAHL, MEDLINE, PREMEDLINE, the Cochrane database, e-psyche, and Social Sciences Abstracts. Presentations from the American Psychiatric Association and American Academy of Child and Adolescent Psychiatry meetings were reviewed.

Study Selection: A literature search was conducted to identify double-blind, placebo-controlled studies of ADHD in adults published in English after 1979. Only trials that used *DSM-III*, *-III-R*, or *-IV* ADHD criteria and followed subjects for ≥ 2 weeks were selected.

Data Extraction: Meta-analysis regression assessed the influence of medication type and study design features on medication effects.

Results: Nineteen trials met criteria and were included in this meta-analysis. These trials studied 13 drugs using 18 different outcome measures of hyperactive, inattentive, or impulsive behavior. After trials were stratified on the class of drug studied (short-acting stimulant vs long-acting stimulant vs nonstimulant), significant differences in effect size were observed between stimulant and nonstimulant medications ($P = .006$ and $P = .0001$, respectively, for short- and long-acting stimulants vs nonstimulants), but the effect for short-acting stimulants was not significant after correcting for study design features. The effect sizes for each drug class were similar in magnitude to what we previously reported for medication treatment studies of children with ADHD. We found significant heterogeneity of effect sizes for short-acting stimulants ($P < .001$) but not for other medication groups.

Conclusions: Although both stimulant and nonstimulant medications are effective for treating ADHD in adults, stimulant medications show greater efficacy for the short durations of treatment characteristic of placebo-controlled studies. We found no significant differences between short- and long-acting stimulant medications. Study design features vary widely among studies and can confound indirect comparisons unless addressed statistically as we have done in this study.

J Clin Psychiatry 2010;71(6):754–763

© Copyright 2009 Physicians Postgraduate Press, Inc.

Submitted: November 25, 2008; *accepted* February 19, 2009.

Online ahead of print: December 29, 2009

(doi:10.4088/JCP.08m04902pur).

Corresponding author: Stephen V. Faraone, PhD, Department of Psychiatry and Behavioral Sciences, SUNY Upstate Medical University, 750 East Adams St, Syracuse, NY 13210 (sfaraone@childpsychresearch.org).

Attention-deficit/hyperactivity disorder (ADHD) is a neurocognitive disorder with a high worldwide prevalence.¹ For decades, the stimulant medications methylphenidate, dextroamphetamine, and mixed amphetamine salts have been the most common drugs used in the treatment of ADHD. Stimulant medications reduce the overactivity, impulsivity, and inattention characteristic of patients with ADHD and improve associated behaviors, including on-task behavior, academic performance, and social functioning.²

Although stimulants have been the mainstay of ADHD pharmacotherapy for many decades, several nonstimulant medications have also shown evidence of efficacy. One of these, atomoxetine, is approved by the US Food and Drug Administration for the treatment of ADHD in adults. The others include tricyclic antidepressants,^{3–5} bupropion,^{6–8} modafinil,^{9,10} monoamine oxidase inhibitors,^{11,12} guanfacine,¹³ and clonidine.¹⁴

While the medications that treat ADHD have been well researched, comparisons among drugs are hindered by the absence of head-to-head trials. In the absence of such trials, physicians must rely on qualitative comparisons among published trials, along with their own clinical experience, to draw conclusions about the efficacy of different medication types on ADHD outcomes. Qualitative reviews of the literature are useful for summarizing results and drawing conclusions about general trends, but they cannot easily evaluate and control the many factors associated with study design that influence the apparent medication effect from a single study.

Meta-analysis provides a systematic quantitative framework for assessing the effects of medications reported by different studies; however, one problem faced by meta-analysis is that different studies use different outcome measures. Comparing such studies is difficult because the meaning of a 1-point difference between drug and placebo groups on a particular outcome measure is typically not the same as the meaning of a 1-point difference on another outcome measure. Meta-analysis partially solves this problem by computing an effect size for each measure. The effect size standardizes the unit of measurement across studies so that a change in 1 point on the effect size scale has the same meaning in each study. For example, in the case of the effect size known as the standardized mean difference (SMD), a 1-point difference means that

the drug and placebo groups differ by 1 standard deviation on the outcome measure.

Comparing effect sizes between studies is questionable if the studies differ substantially on design features that might plausibly influence drug-placebo differences. For example, if one study used a crossover design and the other a parallel design, we could not be sure whether the difference in effect size were due to differences in drug efficacy or differences in methodology. Or, if a study of one drug used physician ratings and the other used self ratings, the difference between studies could be due to the different raters used, not the different drugs. Meta-analysis can address this issue by using regression methods to determine if design features are associated with effect size and if differences in design features can account for differences among drugs.

The present study applies meta-analysis to published literature on the pharmacotherapy of ADHD in adults. For youth with ADHD, there have been prior meta-analyses of medication treatment¹⁵⁻¹⁷ and reviews of effect size for long-acting formulations.¹⁸ There are 2 meta-analyses of treatment for adult ADHD. One is limited to methylphenidate.¹⁹ The other did not address study design features and included some samples that had been selected on the basis of comorbid substance use disorders, which could have confounded medication comparisons.²⁰ Given this limitation of the prior literature, the present work sought to (1) estimate the effect size of the different medications used to treat adult ADHD, (2) determine if differences in the designs of studies might have confounded comparisons of medication efficacy, (3) quantify the evidence for significant differences in effect sizes among medications, and (4) see if features of study design influence the estimate of medication efficacy.

METHOD

A literature search was conducted to identify double-blind, placebo-controlled studies of ADHD in adults published in English after 1979. We searched for articles using the following search engines: PubMed, Ovid, ERIC, CINAHL, MEDLINE, PREMEDLINE, the Cochrane database, e-psyche, and Social Sciences Abstracts. In addition, presentations from the American Psychiatric Association and American Academy of Child and Adolescent Psychiatry meetings were reviewed. We included only studies using randomized, double-blind methodology with placebo controls that defined ADHD using diagnostic criteria from the *Diagnostic and Statistical Manual of Mental Disorders*, Third Edition (*DSM-III*); Third Edition, Revised (*DSM-III-R*); or Fourth Edition (*DSM-IV*), and that followed subjects for 2 weeks or more. To be included, studies must have presented the means and standard deviations of either change or endpoint scores for the drug and placebo groups. Some studies titrated patients to an optimal dose, whereas others titrated subjects to one of several fixed doses. Because it would not be fair to compare low fixed-dose treatment with optimally titrated treatment, for studies presenting data on more than 1 fixed dose, we used the highest dose. We excluded studies

that rated behavior in laboratory environments, studied fewer than 20 subjects in either the drug or placebo groups, were designed to explore appropriate doses for future work, or selected ADHD samples for the presence of a comorbid condition (eg, studies of ADHD among substance abusers).

We extracted the following data from each article: name of dependent outcome measure, name of drug, distribution of *DSM-IV* subtypes in study sample (for studies using *DSM-IV* criteria), design of study (parallel vs crossover), type of outcome score used (change score vs posttreatment score), type of rater (clinician, self), mean age of study sample, percentage of male subjects in study sample, dosing method (fixed dose vs titration to best dose), exclusion of nonresponders (yes/no), use of placebo lead-in (yes/no), and year of publication.

We analyzed 3 types of scores: total ADHD scores (either symptom counts or global ADHD symptom improvement ratings), hyperactive-impulsive symptom scores, and inattentive symptom scores. Effect sizes for dependent measures in each study were expressed as SMDs. The SMD is computed by taking the mean of the active drug group minus the mean of the placebo group and dividing the result by the pooled standard deviation of the groups. Studies reporting change scores provided endpoint minus baseline scores for drug and placebo groups. In this case, the SMD is computed as the difference between change scores. For studies reporting endpoint scores, the SMD is computed as the difference between endpoint scores. Our meta-analysis used the random-effects model of DerSimonian and Laird,²¹ which weights the effect of each study by its sample size. We used meta-analytic regression to assess the degree to which the effect sizes varied with the methodological features of each study described above. We estimated a separate model for each feature. The meta-analyses and meta-analytic regressions were weighted by the reciprocal of the variance of the effect size. We used Egger's²² method to assess for publication biases. We also assessed the number needed to treat (NNT), which is the number of patients who need to be treated to prevent 1 bad outcome. The NNT is typically computed as the inverse of the difference between the response rate on treatment and the response rate on placebo. This version of the NNT cannot be computed if the outcome is a continuous response such as change in ADHD symptoms. Kraemer and Kupfer²³ showed that NNT can be defined for continuous variables if we assume that a treated patient has a successful outcome if the patient's outcome is better than it would have been had they been given placebo rather than medication. They then compute $NNT = 1 / [(2 \times \text{normal}[SMD / \sqrt{2}]) - 1]$ where "normal(Z)" is the probability that a randomly selected standard normal variable is less than Z.

For each study, all dependent outcome measures reported were treated as a separate data point for entry into the analysis, with several studies providing data on more than 1 measure to permit comparison of measures as well as among drugs in this population. Because measures reported from the same study are not statistically independent of one another, standard statistical procedures will produce

Table 1. Studies Used in the Analysis^a

Study	Year	Drug	Mean Dose (mg/d)	Mean Dose (mg/kg/d)	Method of Dosing	N in Drug Group	N in Placebo Group	Mean Age (y)	Male (%)	DSM Version
Wender et al ²⁶	1985	MPH	43.2	0.6	Best	37	37	31	54	III
Spencer et al ²⁷	1995	MPH	65	0.9	Best	23	23	40	43	III-R
Spencer et al ²⁸	1998	Atomoxetine	76	DNP	Best	22	22	34	48	III-R
Wilens et al ²⁹	1999	ABT-418	75	DNP	Fixed	32	30	40	88	IV
Taylor and Russo ⁹	2000	Dextroamphetamine	21.8	DNP	Best	21	21	41	59	IV
Taylor and Russo ⁹	2000	Modafinil	206.8	DNP	Best	21	21	41	59	IV
Spencer et al ³⁰	2001	MAS	54	0.9	Best	27	27	39	56	IV
Wilens et al ³¹	2001	Bupropion SR	362	DNP	Best	21	19	38	55	IV
Michelson et al ³²	2003	Atomoxetine	90	DNP	Best	133	134	40	64	IV
Michelson et al ³²	2003	Atomoxetine	90	DNP	Best	124	124	42	66	IV
Kooij et al ³³	2004	MPH	DNP	0.945	Best	45	45	39	53	IV
Spencer et al ³⁴	2005	MPH	1.1	DNP	Best	78	32	37	58	IV
Wilens et al ³⁵	2005	Bupropion XL	393	DNP	Best	81	81	40	49	IV
Reimherr et al ³⁶	2005	Bupropion SR	298	DNP	Best	35	24	34	73	IV
Weisler et al ³⁷	2006	MAS-XR	60	DNP	Fixed	60	60	40	48	IV
Biederman et al ³⁸	2006	OROS MPH	80.9	DNP	Best	67	74	35	52	IV
Weiss and Hechtman ³⁹	2006	Dextroamphetamine	DNP	DNP	Best	23	20	38	64	IV
Weiss and Hechtman ³⁹	2006	Paroxetine	DNP	DNP	Best	24	20	38	64	IV
Reimherr et al ⁴⁰	2007	OROS MPH	62.7	DNP	Best	43	43	31	66	IV
Spencer et al ⁴¹	2007	<i>d</i> -MPH-ER	40	DNP	Fixed	55	53	39	60	IV
Adler et al ⁴²	2008	LDX	70	DNP	Fixed	120	62	36	52	IV

^aStudies are listed more than once if they studied more than 1 drug or presented more than 1 independent trial of the same drug.

Abbreviations: *d*-MPH-ER = dexamethylphenidate extended release, DNP = data not provided, LDX = lisdexamfetamine dimesylate, MAS = mixed amphetamine salts, MAS-XR = mixed amphetamine salts extended release, MPH = methylphenidate, OROS MPH = osmotic-release oral system methylphenidate, SR = sustained release, XL = extended release.

inaccurate *P* values. To address this intrastudy clustering, variance estimates were adjusted using Huber's²⁴ formula as implemented in STATA.²⁵ This formula is a "theoretical bootstrap" that produces robust statistical tests. The method works by entering the cluster scores (ie, sum of scores within families) into the formula for the estimate of variance. The Huber estimate is also called the "sandwich" estimate because it is calculated as the product of 3 matrices: the matrix formed by taking the outer product of the observation-level likelihood score vectors is in the middle, and this matrix is premultiplied and postmultiplied by the usual model-based variance matrix. The resulting *P* values are valid even when observations are not statistically independent. This approach was applied consistently for all studies.

RESULTS

Table 1 describes the 18 articles meeting criteria for inclusion in the meta-analysis. Studies are listed more than once if they studied more than 1 drug or if they reported independent studies of the same drug. The studies in Table 1 evaluated 13 drugs using 18 different measures of ADHD symptoms to assess efficacy. Each drug-placebo comparison provided information on more than one outcome score. These allowed us to compute 60 effect sizes. Table 2 shows the number of times each medication was studied and the numbers of subjects in the drug and placebo groups from those studies.

Figure 1 shows the results for long-acting stimulants. The mean effect size of 0.73 for long-acting stimulants is statistically significant ($z = 13.0, P < .001$). There was no significant heterogeneity among effect sizes ($\chi^2_{10} = 6.2, P = .8$). Figure

Table 2. Drug-Placebo Comparisons Tested in the Meta-Analysis Studies^a

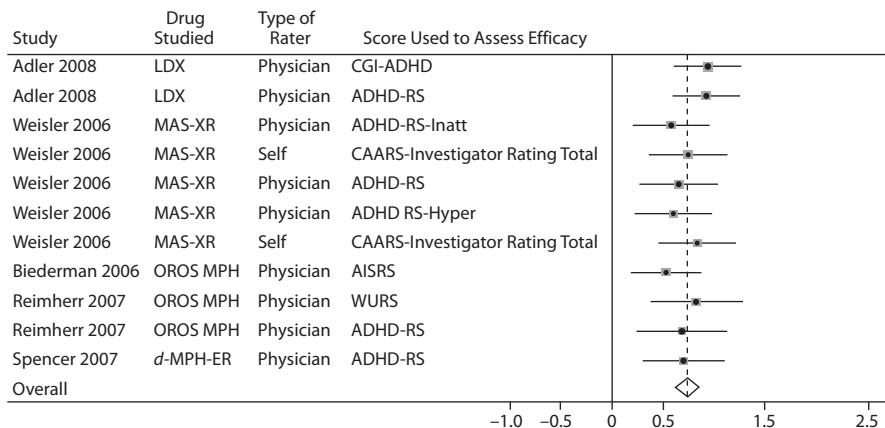
Medication	No. of Studies Testing Each Medication	No. of Subjects in Drug Group	No. of Subjects in Placebo Group
Nonstimulants/others			
ABT-418	1	32	30
Atomoxetine	3	279	280
Bupropion SR	2	56	43
Modafinil	1	21	21
Bupropion XL	1	81	81
Paroxetine	1	24	20
Short-acting stimulants			
MAS	1	27	27
Dextroamphetamine	2	44	41
MPH	4	183	137
Long-acting stimulants			
MAS-XR	1	60	60
OROS MPH	2	110	117
<i>d</i> -MPH-ER	1	55	53
LDX	1	120	62

^aIncludes crossover studies for which the medication and placebo subjects were the same; some studies examined more than 1 medication.

Abbreviations: *d*-MPH-ER = dexamethylphenidate extended release, LDX = lisdexamfetamine dimesylate, MAS = mixed amphetamine salts, MAS-XR = mixed amphetamine salts extended release, MPH = methylphenidate, OROS MPH = osmotic-release oral system methylphenidate, SR = sustained release, XL = extended release.

2 shows the results for short-acting stimulants. The mean effect size of 0.96 is statistically significant ($z = 7.0, P < .001$), but there was significant heterogeneity among effect sizes ($\chi^2_{17} = 58, P < .001$). Figure 3 shows results for nonstimulants. The mean effect size of 0.39 is statistically significant ($z = 13.9, P < .001$). There was no significant heterogeneity among effect sizes ($\chi^2_{30} = 25, P = .8$). Because Figure 3 suggests that our failure to find heterogeneity for nonstimulants

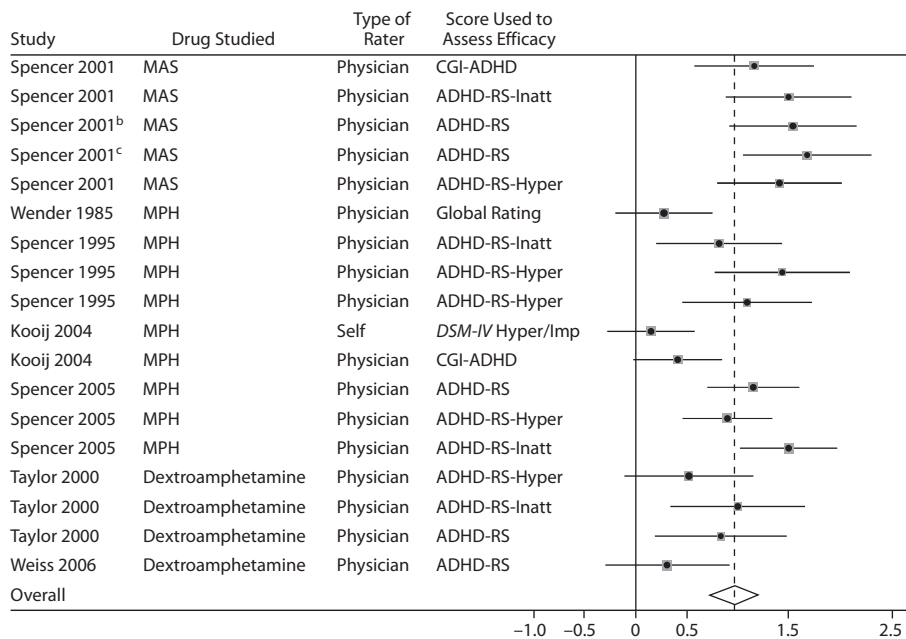
Figure 1. Standardized Mean Differences (SMDs) and 95% Confidence Intervals (CIs) for Long-Acting Stimulants^a



^aThe dark box indicates the effect size with the size of the box proportional to the power of the study. The horizontal line through each box gives the 95% confidence interval. At the bottom of the plot, the center of the diamond gives the mean SMD, and the width gives its 95% confidence interval. For Weisler 2006, there are 2 entries for the ADHD-RS; the first gives results based on morning ratings, and the second gives results based on afternoon ratings.

Abbreviations: ADHD-RS = Attention-Deficit/Hyperactivity Disorder Rating Scale, AISRS = Adult Attention-Deficit/Hyperactivity Disorder Investigator Symptom Rating Scale, CAARS = Conners' Adult ADHD Rating Scale, CGI = Clinical Global Impressions scale, d-MPH-ER = dexamfetamine extended release, Hyper = hyperactive symptoms, Inatt = inattentive symptoms, LDX = lisdexamfetamine dimesylate, MAS-XR = mixed amphetamine salts extended release, OROS MPH = osmotic-release oral system methylphenidate, WURS = Wender-Reimherr Utah Rating Scale.

Figure 2. Standardized Mean Differences (SMDs) and 95% CIs for Short-Acting Stimulants^a



^aThe dark box indicates the effect size with the size of the box proportional to the power of the study. The horizontal line through each box gives the 95% confidence interval. At the bottom of the plot, the center of the diamond gives the mean SMD, and the width gives its 95% confidence interval. ^bResults based on outcome scores. ^cResults based on change scores.

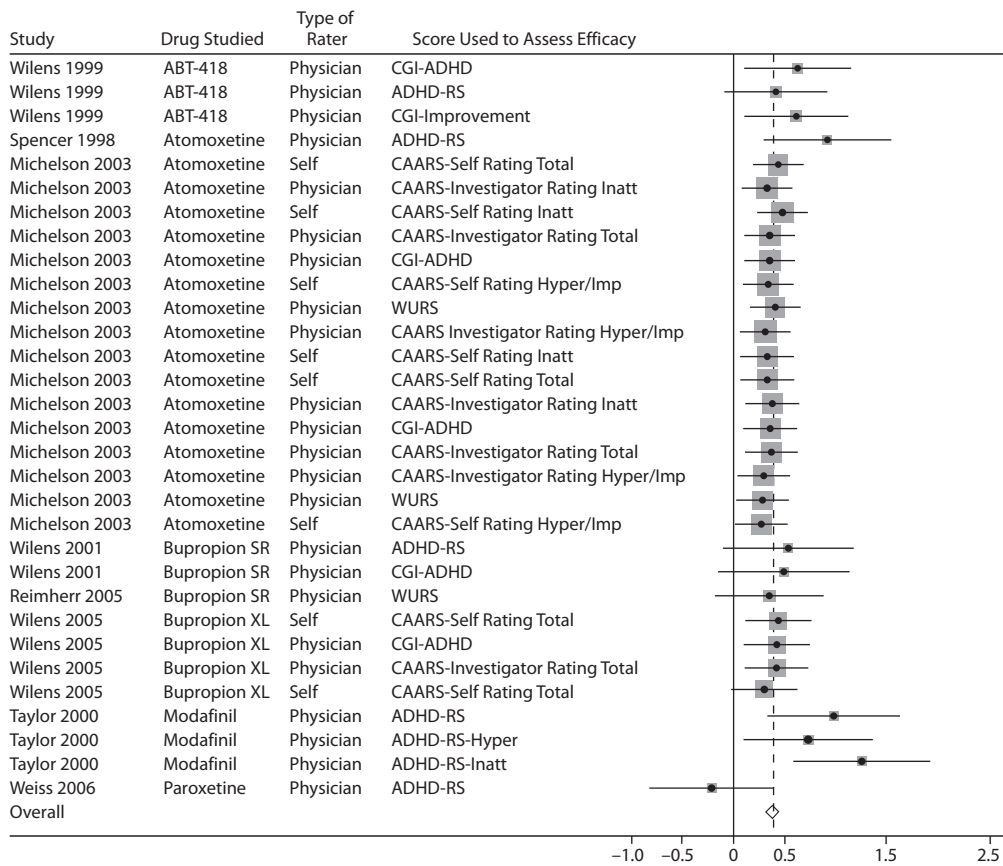
Abbreviations: ADHD-RS = Attention-Deficit/Hyperactivity Disorder Rating Scale, CGI = Clinical Global Impressions scale, Hyper = hyperactive symptoms, Inatt = inattentive symptoms, MAS = mixed amphetamine salts, MPH = methylphenidate.

was accounted for by the atomoxetine studies, we reran the heterogeneity analyses separately for atomoxetine and other nonstimulants. There was no significant heterogeneity among effect sizes for atomoxetine ($\chi^2_{16} = 6.1, P = 1.0$) or for the other nonstimulants ($\chi^2_{13} = 16.2, P = .24$).

We conducted separate random-effects meta-analyses for each class of medication to assess for publication bias. We

found no evidence of publication bias for the nonstimulants ($t_{30} = 1.7, P = .09$) or the long-acting stimulants ($t_{10} = 0.8, P = .5$). There was evidence of publication bias for short-acting stimulants ($t_{17} = 4.7, P < .001$). The trim and fill method to correct for publication bias⁴³ suggests that the pooled estimate of effect size for the short-acting stimulants (0.96) is an overestimate of a true effect size of 0.86.

Figure 3. Standardized Mean Differences (SMDs) and 95% CIs for Nonstimulants^a



^aThe dark box indicates the effect size with the size of the box proportional to the power of the study. The horizontal line through each box gives the 95% confidence interval. At the bottom of the plot, the center of the diamond gives the mean SMD, and the width gives its 95% confidence interval.

Abbreviations: ADHD-RS = Attention-Deficit/Hyperactivity Disorder Rating Scale, CAARS = Conners' Adult ADHD Rating Scale, CGI = Clinical Global Impressions scale, Hyper = hyperactive symptoms, Hyper/Imp = hyperactive/impulsive symptoms, Inatt = inattentive symptoms, SR = sustained release, WURS = Wender-Reimherr Utah Rating Scale, XL = extended release.

Table 3. Clinical and Demographic Features of Study Samples by Type of Medication

Variable	Nonstimulant/Other (n = 7)	Short-Acting Stimulant (n = 8)	Long-Acting Stimulant (n = 5)	Statistics		
				F	df	P
ADHD subtype, %						
Inattentive	29	36	15	0.7	2,9	.5
Hyperactive-impulsive	4	4	1	0.5	2,9	.6
Combined	67	60	85	0.8	2,9	.5
Gender, male, %	63	55	55	1.3	2,17	.3
Age, mean, y	38	38	36	0.8	2,17	.5

Abbreviation: ADHD = attention-deficit/hyperactivity disorder.

Meta-analysis regression found a significant effect of drug type ($F_{2,18} = 10.9, P = .0008$). The effect sizes for non-stimulant medications were significantly less than those for short-acting stimulants ($t_{18} = 3.1, P = .006$) and long-acting stimulants ($t_{18} = 4.2, P < .0001$). The 2 classes of stimulant medication did not differ significantly from one another ($t_{18} = 1.5, P = .3$). We also found no significant difference between methylphenidate-based and amphetamine-based medications ($t_{11} = 0.1, P = .9$).

Table 3 presents clinical and demographic variables that might potentially confound our meta-analytic comparisons

of medication types. We found no significant differences among studies for the distribution of *DSM-IV* subtypes (for studies using *DSM-IV* criteria), gender, or age. Table 4 describes the study design features for the 3 medication classes. We found significant differences for 1 study design feature: short-acting stimulant studies were less likely to use last-observation-carried-forward (LOCF) methodology compared with the other types of studies. The medication groups did not differ in the types of raters used, the score categories used to assess efficacy, whether or not subjects with a history of nonresponse were excluded at baseline, exclusion

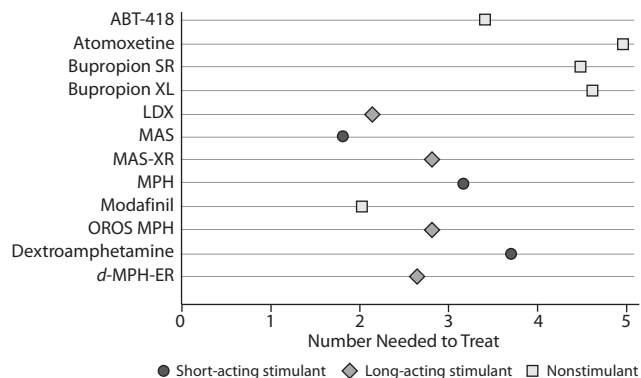
Table 4. Design Features of Studies by Type of Medication (% of each medication class)

	Nonstimulant/Other	Short-Acting Stimulant	Long-Acting Stimulant	P Value ^a
Type of score used				
Change	43	0	20	.2
Outcome	57	100	80	
Type of rater				
Self	43	14	0	.4
Physician	57	86	100	
Score category				
ADHD total	86	57	80	.7
Hyperactive-impulsive	0	29	0	
Inattentive	14	14	20	
Study design				
Crossover	14	71	20	.11
Parallel	86	29	80	
Placebo lead-in?				
No	57	100	100	.08
Yes	43	0	0	
Nonresponders excluded?				
No	71	100	100	.3
Yes	29	0	0	
Best vs fixed dose				
Best	86	100	40	.06
Fixed	14	0	60	
No. of sites				
Multisite	43	14	60	.4
Single site	57	86	40	
LOCF methodology?				
Yes	71	15	100	.01
No	29	86	0	

^aFisher exact test.

Abbreviations: ADHD = attention-deficit/hyperactivity disorder, LOCF = last observation carried forward.

Figure 4. Number Needed to Treat for Each Drug



Abbreviations: d-MPH-ER = dexamethylphenidate extended release, LDX = lisdexamfetamine dimesylate, MAS = mixed amphetamine salts, MAS-XR = mixed amphetamine salts extended release, MPH = methylphenidate, OROS MPH = osmotic-release oral system methylphenidate, SR = sustained release, XL = extended release.

of nonresponders, type of score (change score vs outcome score), or study design (crossover vs parallel). There were no significant differences in design features between studies using amphetamine-based medications and those using methylphenidate-based medications.

We used meta-analysis regression to determine if any of the design features in Tables 3 and 4 predicted effect size. Effect sizes were greater for crossover compared with parallel designs (0.88 vs 0.51; $F_{1,19} = 10.0, P = .005$), for physician versus self ratings (0.68 vs 0.43; $F_{1,18} = 4.8, P = .04$), for outcome compared with change scores (0.75 vs 0.45; $F_{1,18} = 11.4,$

$P = .003$), for studies that included prior nonresponders compared with those that did not (0.74 vs 0.36; $F_{1,18} = 7.2, P = .0006$), for studies that did not use a placebo lead-in compared with those that did (0.75 vs 0.36; $F_{1,18} = 7.5, P = .0006$), for those using completer compared with LOCF methodology (0.91 vs 0.47; $F_{1,18} = 11.3, P = .004$), and for those using 1 site compared with multiple sites to enroll subjects (0.83 vs 0.44; $F_{1,18} = 9.8, P = .006$). All other effects were nonsignificant (all P values $> .10$). Because use of LOCF methodology was significantly associated with both effect size and drug class, this could explain the observed differences among drug classes. After correcting for this confounding factor, the difference between long-acting stimulants and nonstimulants remained statistically significant ($t_{18} = 6.6, P < .001$), but the difference between short-acting stimulants and nonstimulants lost statistical significance ($t_{18} = 0.5, P = .6$).

Figure 4 presents the mean NNT for each drug computed across all data points for each drug. Paroxetine is not included in the figure because it had a negative effect size. As described in the Method, the NNT can be computed as a transformation of the standardized mean difference. In this context, a successful outcome means that the outcome of a treated patient is better than it would have been if the patient had not been treated. Thus, higher NNTs correspond to fewer successful treatments. As expected from Figures 1 through 3, Figure 4 shows higher NNTs for all nonstimulants except modafinil and lower NNTs for all stimulants.

Table 5 presents mean effect sizes stratified by the specific outcome scores used by the studies analyzed. It also gives the number of studies that used the outcome score. This table

Table 5. Mean Effect Sizes and Number of Studies, Stratified by Outcome Scores and Type of Medication

Score Used to Assess Efficacy	Nonstimulant		Stimulant	
	Mean Effect Size	No. of Studies	Mean Effect Size	No. of Studies
ADHD-RS				
Total	0.51	5	0.89	9
Inattentive symptoms	1.26	1	0.95	5
Hyperactive symptoms	0.74	1	0.94	6
AISRS	0.53	1
DSM-IV hyperactive/impulsive symptoms	0.23	1
CAARS-Investigator Rating				
Total	0.39	3	0.76	2
Inattentive symptoms	0.36	2
Hyperactive/impulsive symptoms	0.31	2
CAARS-Self Rating				
Total	0.38	4
Inattentive symptoms	0.41	2
Hyperactive/impulsive symptoms	0.31	2
WURS	0.35	3	0.83	1
CGI-Improvement	0.62	1
CGI-ADHD	0.46	5	0.79	3
Global rating of ADHD	0.28	1

Abbreviations: ADHD = attention-deficit/hyperactivity disorder, ADHD-RS = Attention-Deficit/Hyperactivity Disorder Rating Scale, AISRS = Adult Attention-Deficit/Hyperactivity Disorder Investigator Symptom Rating Scale, CAARS = Conners' Adult ADHD Rating Scale, CGI = Clinical Global Impressions scale, WURS = Wender-Reimherr Utah Rating Scale.

should be interpreted cautiously given that each scale was used by only a handful of studies. Of note, for both stimulants and nonstimulants, effect sizes tend to be higher for the ADHD rating scale.

DISCUSSION

Our meta-analysis of efficacy outcomes for medication studies of adults with ADHD found significant differences in effect size between stimulant and nonstimulant medications, even after correcting for study design features that might have confounded the results. This finding and the magnitude of the effect sizes for each drug class were similar to what we previously reported for medication treatment studies of children with ADHD.⁴⁴

We found evidence of publication bias for short-acting stimulants, which reduced the mean effect size from 0.96 to 0.86, which is more in line with the effect size for long-acting stimulants (0.73). Although our analyses indicate that effect sizes for stimulants are significantly greater than those for other medications, the presence of confounds suggests that, in the absence of confirmatory head-to-head studies, caution is warranted when comparing the effects of different medications across studies. Although head-to-head trials are needed to make definitive statements about efficacy differences, our results comparing stimulants and nonstimulants are compatible with the efficacy differences from studies of children between atomoxetine and mixed amphetamine salts reported by Wigal et al^{45,46} and the conclusions of a prior review limited to a smaller subset of studies that excluded short-acting stimulants.¹⁸

Table 1 shows little uniformity in the study design parameters used to assess medication efficacy. Although this does not affect the interpretation of individual studies, it makes difficult the comparison of the efficacy of different medications in the absence of direct comparisons within

the same study. This problem is further compounded by the fact that effect sizes, which compare treatment efficacy, differ according to some study design variables. Comparing medication effect sizes in different studies, without accounting for these influences, will lead to spurious conclusions. We found that only 1 design variable differed significantly among the medication groups: short-acting stimulant studies were less likely to use LOCF methodology compared with the other types of studies. When we adjusted for this difference using meta-analysis regression, the difference between short-acting stimulants and nonstimulants lost statistical significance.

The robust effects of most of the medications studied can be seen in Figures 1 through 3, which show that most measures of effect from all studies were statistically significant (ie, the horizontal line indicating the 95% confidence interval does not overlap with zero). The only drug that was worse than placebo was paroxetine, which is clearly not effective for ADHD (Figure 3). For the long-acting stimulants (Figure 1) and nonstimulants (Figure 3), heterogeneity within each medication class was not statistically significant. In contrast, for short-acting stimulants, we found evidence of significant heterogeneity among studies. Visual inspection of Figure 2 suggests that the overall pattern of greater dispersion, as compared with Figures 1 and 3, cannot be attributed to 1 or a few extreme study results.

The NNT results (Figure 4) are useful because they give some sense of the clinical implications of the variability of effect sizes among different drugs. For example, some nonstimulants have NNTs approaching 5, and some short- and long-acting stimulants have NNTs near 2. For these more effective stimulants, only 2 patients need to be treated to have 1 positive outcome, compared with 5 patients for the nonstimulants. These NNTs also tell us that for 50% of patients, stimulant trials are wasted (because 1 of the 2 treated patients would have done better or as well on placebo) and for 80% of

patients, nonstimulant treatments are wasted. Note that we are using the term “wasted” to mean that the patient would have done as well or better if they had been treated with placebo. The 50% response rate for stimulants and the 20% response rate for nonstimulants are less than response rates reported in the literature because those response rates are not placebo adjusted. Regardless of the absolute magnitude of effect, though, the NNT has implications for both the outcome of individual patients and the costs of wasted treatments. The NNT results also suggest that most patients will need more than 1 drug trial to achieve a positive outcome that cannot be attributed to a placebo effect. Ideally, our NNT result would have been based on a common, binary definition of response across all studies, but that was not possible due to how data were presented in each article.

Our work has methodological implications for the design of ADHD treatment studies because several methodological features of the studies were associated with the magnitude of their effect sizes. Physician ratings led to higher effect sizes compared with self ratings. Although self-ratings may be sufficiently reliable and sensitive in some contexts, they do not appear to be as useful as physician ratings for detecting response to ADHD medications. Studies using outcome scores had higher effect sizes than studies using change scores. The use of change scores may be more conservative because they adjust for baseline group differences. Crossover designs had greater effect sizes than parallel designs; this finding makes sense given the inherent matching in the crossover design that removes between-patient variability from the analysis. Studies using only 1 site had greater effect sizes than studies using multiple sites. This may be due to difficulties managing large multisite trials, especially as regards standardization of methods. As expected, studies using the conservative LOCF methodology had smaller effect sizes than studies that used only completers in the analysis. Of note, all of the multisite trials used LOCF methodology.

It is difficult to compare the short- and long-acting stimulants because the studies we reviewed were not intended to assess the duration of the effect of a medication throughout the day. Moreover, the use of short-acting medications in a clinical trial environment may not reflect the real-world effects of missed doses if compliance is enhanced by the clinical trial. In contrast to our conclusions, Peterson and colleagues²⁰ meta-analysis of medication studies of adult ADHD found that short-acting stimulants were more effective than long-acting stimulants. Although our initial analysis showed short-acting stimulants to have greater effect sizes than long-acting stimulants, that difference could be accounted for by study confounds and publication bias. As is evident from Table 4, many of the methodological features predictive of lower effect sizes were more common among the long-acting compared with short-acting stimulant studies. For example, all of the long-acting studies used conservative LOCF methodology, compared with only 15% of the short-acting studies. There are 3 other differences between our study and the study by Peterson et al,²⁰ which may explain differences in results: (1) we excluded studies

of substance abusers, whereas they did not; (2) we used continuous measures of ADHD outcomes, whereas they used the proportion of responders; and (3) we included Biederman and colleagues⁴⁷ study of LDX, which was previously unavailable to Peterson et al but which also had the highest effect size among the long-acting stimulants.

Our work must be interpreted in light of several limitations. Compared with studies of child and adolescent ADHD, there are relatively few treatment studies of adult ADHD. And because meta-analysis relies on data collected by others, the validity and strength of our conclusions rely on the quality of the individual studies, which was beyond our control. Although our inclusion criteria required each study to meet criteria for study quality (use of double-blind placebo-controlled methods, selection of subjects with structured diagnostic criteria), other features of these studies may account for some of the interstudy variability in results. Because we relied on data presented by authors, we could not assess the effects of all potential confounds; ie, we were limited by what other investigators chose to present. For example, we could not compute the effect sizes at specific time points, because such data were rarely provided. Similarly, we did not assess differential duration of action between medication classes, because this effect is rarely presented. Our conclusions about short-acting stimulants are limited by the significant heterogeneity we found for their effect sizes. All meta-analyses are limited by the quality of the studies analyzed. For that reason, we limited our review to double-blind placebo-controlled studies that diagnosed ADHD using DSM criteria. Nevertheless, although our analyses controlled for several study design features, it is possible that systematic methodological differences between drugs or classes of drugs might have led to spurious results.

Our meta-analysis excluded studies not meeting our inclusion criteria. We excluded 2 open trials. One reported a 70% response rate using doses of 40 mg/d.⁴⁸ We excluded a double-blind parallel study by Wood et al⁴⁹ because they did not use structured diagnostic criteria. Instead, they used the diagnosis of minimal brain dysfunction to select patients for study. Using a low mean dose of 27 mg/d, they found a 53% improvement rate in methylphenidate-treated adults with minimal brain dysfunction. We excluded Iaboni and colleagues⁵⁰ double-blind crossover study because it did not provide the data needed to compute an effect size. We also excluded studies that did not provide data to compute the SMD⁵¹ or that only provided data on ADHD patients with substance use disorders.^{52–58}

Although indirectly comparing treatments using meta-analysis cannot replace direct comparison of treatments within the same study, there is evidence from other medical fields that meta-analysis is a valid approach. Song et al⁵⁹ examined 44 published meta-analyses that used measures of effect magnitude to indirectly compare treatments. In most cases, the results using indirect meta-analysis comparisons did not differ from the results of direct comparisons within the same study. However, for 3 of the 44 comparisons, they found significant differences between the direct and the

indirect estimates. Chou et al⁶⁰ compared initial highly active antiretroviral therapy (HAART) with a protease inhibitor (PI) versus a nonnucleoside reverse transcriptase inhibitor (NNRTI). They did a direct meta-analysis of 12 head-to-head comparisons and an indirect meta-analysis of 6 trials of NNRTI-based HAART and 8 trials of PI-based HAART. In the direct meta-analysis, NNRTI-based regimens were better than PI-based regimens for virologic suppression. In contrast, the indirect meta-analyses showed that NNRTI-based HAART was worse than PI-based HAART for virologic suppression. From these studies, it seems reasonable to conclude that indirect comparisons usually agree with the results of head-to-head direct comparisons. Nevertheless, when direct comparisons are lacking, the results of indirect comparisons using measures of effect magnitude should be viewed cautiously.

Despite these limitations, our findings highlight the remarkable variability in methods among adult ADHD treatment studies. Although this does not argue against the validity of individual studies, it highlights the difficulty one faces when interpreting differential medication efficacy when a direct comparative trial is not available. Yet, despite this variability, we found substantial and significant differences in efficacy between stimulant and nonstimulant medications. Although efficacy effect sizes should not be the sole guide for clinicians to use when choosing an ADHD medication, they do provide useful information for clinicians to consider when planning treatment regimens for patients with ADHD.

Drug names: atomoxetine (Strattera), bupropion (Wellbutrin, Aplenzin, and others), clonidine (Catapres, Jenloga, and others), dextroamphetamine extended release (Focalin XR), dextroamphetamine (Dexedrine, Dextroamp, and others), guanfacine (Intuniv, Tenex, and others), lisdexamfetamine dimesylate (Vyvanse), methylphenidate (Ritalin, Daytrana, and others), mixed amphetamine salts (Adderall), modafinil (Provigil), osmotic-release oral system methylphenidate (Concerta), paroxetine (Paxil, Pexeva, and others).

Author affiliations: From the Departments of Psychiatry (both authors) and Neuroscience and Physiology (Dr Faraone), SUNY Upstate Medical University, Syracuse, New York.

Potential conflicts of interest: Dr Faraone has received consulting fees or research support from or has been on advisory boards or a speaker for Shire, Eli Lilly, Pfizer, McNeil, and the National Institutes of Health. Dr Glatt reports no additional financial or other relationship relevant to the subject of this article.

Funding/support: This work was supported by a grant from Shire US to Dr Faraone.

REFERENCES

1. Faraone SV, Sergeant J, Gillberg C, et al. The worldwide prevalence of ADHD: is it an American condition? *World Psychiatry*. 2003;2(2):104–113.
2. Greenhill LL, Pliszka S, Dulcan MK, et al. Summary of the practice parameter for the use of stimulant medications in the treatment of children, adolescents, and adults. *J Am Acad Child Adolesc Psychiatry*. 2001;40(11):1352–1355.
3. Spencer T, Biederman J, Coffey B, et al. A double-blind comparison of desipramine and placebo in children and adolescents with chronic tic disorder and comorbid attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry*. 2002;59(7):649–656.
4. Wilens TE, Biederman J, Baldessarini RJ, et al. Cardiovascular effects of therapeutic doses of tricyclic antidepressants in children and adolescents. *J Am Acad Child Adolesc Psychiatry*. 1996;35(11):1491–1501.
5. Biederman J, Baldessarini RJ, Wright V, et al. A double-blind placebo controlled study of desipramine in the treatment of ADD, I: efficacy. *J Am Acad Child Adolesc Psychiatry*. 1989;28(5):777–784.
6. Conners CK, Casat CD, Gualtieri CT, et al. Bupropion hydrochloride in attention deficit disorder with hyperactivity. *J Am Acad Child Adolesc Psychiatry*. 1996;35(10):1314–1321.
7. Casat CD, Pleasants DZ, Schroeder DH, et al. Bupropion in children with attention deficit disorder. *Psychopharmacol Bull*. 1989;25(2):198–201.
8. Casat CD, Pleasants DZ, Van Wyck Fleet J. A double-blind trial of bupropion in children with attention deficit disorder. *Psychopharmacol Bull*. 1987;23(1):120–122.
9. Taylor FB, Russo J. Efficacy of modafinil compared to dextroamphetamine for the treatment of attention deficit hyperactivity disorder in adults. *J Child Adolesc Psychopharmacol*. 2000;10(4):311–320.
10. Rugino TA, Copley TC. Effects of modafinil in children with attention-deficit/hyperactivity disorder: an open-label study. *J Am Acad Child Adolesc Psychiatry*. 2001;40(2):230–235.
11. Ernst M. MAOI treatment of adult ADHD. Presented at: NIMH Conference on Alternative Pharmacology of ADHD; 1996; Washington, DC.
12. Shekim WO, Davis LG, Bylund DB, et al. Platelet MAO in children with attention deficit disorder and hyperactivity: a pilot study. *Am J Psychiatry*. 1982;139(7):936–938.
13. Biederman J, Melmed RD, Patel A, et al. SPD503 Study Group. A randomized, double-blind, placebo-controlled study of guanfacine extended release in children and adolescents with attention-deficit/hyperactivity disorder. *Pediatrics*. 2008;121(1):e73–e84.
14. Connor DF, Fletcher KE, Swanson JM. A meta-analysis of clonidine for symptoms of attention-deficit hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry*. 1999;38(12):1551–1559.
15. Faraone SV, Biederman J, Roe CM. Comparative efficacy of Adderall and methylphenidate in attention-deficit/hyperactivity disorder: a meta-analysis. *J Clin Psychopharmacol*. 2002;22(5):468–473.
16. Faraone SV, Biederman J. Efficacy of Adderall for attention-deficit/hyperactivity disorder: a meta-analysis. *J Atten Disord*. 2002;6(2):69–75.
17. Schachter HM, Pham B, King J, et al. How efficacious and safe is short-acting methylphenidate for the treatment of attention-deficit disorder in children and adolescents? a meta-analysis. *CMAJ*. 2001;165(11):1475–1488.
18. Banaschewski T, Coghill D, Santosh P, et al. Long-acting medications for the hyperkinetic disorders: a systematic review and European treatment guideline. *Eur Child Adolesc Psychiatry*. 2006;5(8):476–495.
19. Faraone SV, Spencer T, Aleardi M, et al. Meta-analysis of the efficacy of methylphenidate for treating adult attention-deficit/hyperactivity disorder. *J Clin Psychopharmacol*. 2004;24(1):24–29.
20. Peterson K, McDonagh MS, Fu R. Comparative benefits and harms of competing medications for adults with attention-deficit hyperactivity disorder: a systematic review and indirect comparison meta-analysis. *Psychopharmacology (Berl)* 2008;197(1):1–11.
21. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177–188.
22. Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629–634.
23. Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. *Biol Psychiatry*. 2006;59(11):990–996.
24. Huber PJ. The behavior of maximum likelihood estimates under non-standard conditions. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Vol 1. Berkeley, CA: University of California Press; 1967:221–233.
25. Stata Corporation. *Stata User's Guide: Release 9*. College Station, TX: Stata Corp LP; 2005.
26. Wender PH, Reimherr FW, Wood DR. Stimulant therapy of "adult hyperactivity." *Arch Gen Psychiatry*. 1985;42(8):840.
27. Spencer T, Wilens T, Biederman J, et al. A double-blind, crossover comparison of methylphenidate and placebo in adults with childhood-onset attention-deficit hyperactivity disorder. *Arch Gen Psychiatry*. 1995;52(6):434–443.
28. Spencer T, Biederman J, Wilens T, et al. Effectiveness and tolerability of tomoxetine in adults with attention deficit hyperactivity disorder. *Am J Psychiatry*. 1998;155(5):693–695.
29. Wilens TE, Biederman J, Spencer TJ, et al. A pilot controlled clinical trial of ABT-418, a cholinergic agonist, in the treatment of adults with attention deficit hyperactivity disorder. *Am J Psychiatry*. 1999;156(12):1931–1937.
30. Spencer T, Biederman J, Wilens T, et al. Efficacy of a mixed amphetamine

- salts compound in adults with attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry*. 2001;58(8):775–782.
31. Wilens TE, Spencer TJ, Biederman J, et al. A controlled clinical trial of bupropion for attention deficit hyperactivity disorder in adults. *Am J Psychiatry*. 2001;158(2):282–288.
 32. Michelson D, Adler L, Spencer T, et al. Atomoxetine in adults with ADHD: two randomized, placebo-controlled studies. *Biol Psychiatry*. 2003;53(2):112–120.
 33. Kooij JJ, Burger H, Boonstra AM, et al. Efficacy and safety of methylphenidate in 45 adults with attention-deficit/hyperactivity disorder: a randomized placebo-controlled double-blind cross-over trial. *Psychol Med*. 2004;34(6):973–982.
 34. Spencer T, Biederman J, Wilens T, et al. A large, double-blind, randomized clinical trial of methylphenidate in the treatment of adults with attention-deficit/hyperactivity disorder. *Biol Psychiatry*. 2005;57(5):456–463.
 35. Wilens TE, Haight BR, Horrigan JP, et al. Bupropion XL in adults with ADHD: a randomized, placebo-controlled study. *Biol Psychiatry*. 2005;57(7):793–801.
 36. Reimherr FW, Marchant BK, Strong RE, et al. Emotional dysregulation in adult ADHD and response to atomoxetine. *Biol Psychiatry*. 2005;58(2):125–131.
 37. Weisler RH, Biederman J, Spencer TJ, et al. Mixed amphetamine salts extended-release in the treatment of adult ADHD: a randomized, controlled trial. *CNS Spectr*. 2006;11(8):625–639.
 38. Biederman J, Mick E, Surman C, et al. A randomized, placebo-controlled trial of OROS methylphenidate in adults with attention-deficit/hyperactivity disorder. *Biol Psychiatry*. 2006;59(9):829–835.
 39. Weiss M, Hechtman L; Adult ADHD Research Group. A randomized double-blind trial of paroxetine and/or dextroamphetamine and problem-focused therapy for attention-deficit/hyperactivity disorder in adults. *J Clin Psychiatry*. 2006;67(4):611–619.
 40. Reimherr FW, Williams ED, Strong RE, et al. A double-blind, placebo-controlled, crossover study of osmotic release oral system methylphenidate in adults with ADHD with assessment of oppositional and emotional dimensions of the disorder. *J Clin Psychiatry*. 2007;68(1):93–101.
 41. Spencer TJ, Adler LA, McGough JJ, et al. Efficacy and safety of dexamethylphenidate extended-release capsules in adults with attention-deficit/hyperactivity disorder. *Biol Psychiatry*. 2007;61(12):1380–1387.
 42. Adler LA, Goodman DW, Kollins SH, et al. Double-blind, placebo-controlled study of the efficacy and safety of lisdexamfetamine dimesylate in adults with attention-deficit/hyperactivity disorder. *J Clin Psychiatry*. 2008;69(9):1364–1373.
 43. Duval S, Tweedie R. A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis. *J Am Stat Assoc*. 2000;95(449):89–98.
 44. Faraone SV, Biederman J, Spencer TJ, et al. Comparing the efficacy of medications for ADHD using meta-analysis. *MedGenMed*. 2006;8(4):4.
 45. Wigal SB, McGough JJ, McCracken JT, et al. A laboratory school comparison of mixed amphetamine salts extended release (Adderall XR) and atomoxetine (Strattera) in school-aged children with attention deficit/hyperactivity disorder. *J Atten Disord*. 2005;9(1):275–289.
 46. Faraone SV, Wigal SB, Hodgkins P. Forecasting three-month outcomes in a laboratory school comparison of mixed amphetamine salts extended release (Adderall XR) and atomoxetine (Strattera) in school-aged children with ADHD. *J Atten Disord*. 2007;11(1):74–82.
 47. Biederman J, Krishnan S, Zhang Y, et al. Efficacy and safety of lisdexamfetamine (NRP-104) in children with attention-deficit/hyperactivity disorder: a phase 3, multicenter, randomized, double-blind, forced dose, parallel-group study. *Clin Ther*. 2007;29(3):450–463.
 48. Shekim WO, Asarnow RF, Hess E, et al. A clinical and demographic profile of a sample of adults with attention deficit hyperactivity disorder, residual state. *Compr Psychiatry*. 1990;31(5):416–425.
 49. Wood DR, Reimherr FW, Wender PH, et al. Diagnosis and treatment of minimal brain dysfunction in adults: a preliminary report. *Arch Gen Psychiatry*. 1976;33(12):1453–1460.
 50. Iaboni F, Bouffard R, Minde K, et al. *The Efficacy of Methylphenidate in Treating Adults With Attention-Deficit/Hyperactivity Disorder*. Philadelphia, PA: American Academy of Child and Adolescent Psychiatry; 1996.
 51. Kooij JS, Boonstra AM, Vermeulen SH, et al. Response to methylphenidate in adults with ADHD is associated with a polymorphism in SLC6A3 (DAT1). *Am J Med Genet B Neuropsychiatr Genet*. 2008;147B(2):201–208.
 52. Levin FR, Evans SM, McDowell DM, et al. Methylphenidate treatment for cocaine abusers with adult attention-deficit/hyperactivity disorder: a pilot study. *J Clin Psychiatry*. 1998;59(6):300–305.
 53. Paterson R, Douglas C, Hallmayer J, et al. A randomised, double-blind, placebo-controlled trial of dexamphetamine in adults with attention deficit hyperactivity disorder. *Aust N Z J Psychiatry*. 1999;33(4):494–502.
 54. Levin ED, Conners CK, Silva D, et al. Effects of chronic nicotine and methylphenidate in adults with attention deficit/hyperactivity disorder. *Exp Clin Psychopharmacol*. 2001;9(1):83–90.
 55. Levin FR, Evans SM, Brooks DJ, et al. Treatment of cocaine dependent treatment seekers with adult ADHD: double-blind comparison of methylphenidate and placebo. *Drug Alcohol Depend*. 2007;87(1):20–29.
 56. Levin FR, Evans SM, Brooks DJ, et al. Treatment of methadone-maintained patients with adult ADHD: double-blind comparison of methylphenidate, bupropion and placebo. *Drug Alcohol Depend*. 2006;81(2):137–148.
 57. Schubiner H, Saules KK, Arfken CL, et al. Double-blind placebo-controlled trial of methylphenidate in the treatment of adult ADHD patients with comorbid cocaine dependence. *Exp Clin Psychopharmacol*. 2002;10(3):286–294.
 58. Carpentier PJ, de Jong CA, Dijkstra BA, et al. A controlled trial of methylphenidate in adults with attention deficit/hyperactivity disorder and substance use disorders. *Addiction*. 2005;100(12):1868–1874.
 59. Song F, Altman DG, Glenny AM, et al. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ*. 2003;326(7387):472.
 60. Chou R, Fu R, Huffman LH, et al. Initial highly-active antiretroviral therapy with a protease inhibitor versus a non-nucleoside reverse transcriptase inhibitor: discrepancies between direct and indirect meta-analyses. *Lancet*. 2006;368(9546):1503–1515.