Eur Child Adolesc Psychiatry (2010) 19:353–364
DOI 10.1007/s00787-009-0054-3

ORIGINAL CONTRIBUTION

Comparing the efficacy of stimulants for ADHD in children and adolescents using meta-analysis

Stephen V. Faraone · Jan Buitelaar

Received: 14 January 2009 / Accepted: 20 August 2009 / Published online: 10 September 2009

Abstract Stimulants used to treat attention-deficit/hyperactivity disorder (ADHD) have been well researched, but comparisons among stimulants are hindered by the absence of direct comparative trials. The goal of this work was to compare the efficacy of methylphenidate and amfetamine formulations through a meta-analysis of double-blind placebo-controlled trials. We analyzed recent published literature on the stimulant therapy of ADHD to describe the variability of drug–placebo effect sizes. A literature search was conducted to identify double-blind, placebo-controlled studies of ADHD in children and adolescents published after 1979. Meta-analysis regression assessed the influence of medication type and study design features on medication effects. Twenty-three trials met criteria and were included in this meta-analysis. These trials studied 11 drugs using 19 different outcome measures of hyperactive, inattentive, or impulsive behavior. We found significant differences between amfetamine and methylphenidate products, even after correcting for study design features that might have confounded the results. Our analyses indicate that effect sizes for amfetamine products are significantly, albeit moderately, greater than those for methylphenidate. We found that most measures of effect from all studies were statistically significant. Our findings suggest that amfetamine products may be moderately more efficacious than methylphenidate products, even after controlling for potentially confounding study design features. This difference in effect size may be due to differences between amfetamine and methylphenidate in the molecular mechanisms involved in facilitating the dopaminergic neurotransmission.

Keywords ADHD · Medications · Efficacy · Effect size · Meta-analysis

S. V. Faraone
Departments of Psychiatry and Neuroscience & Physiology, SUNY Upstate Medical University, Syracuse, NY, USA

J. Buitelaar
Department of Psychiatry, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands

S. V. Faraone (✉)
Departments of Psychiatry and Behavioral Sciences, SUNY Upstate Medical University (SVF), 750 East Adams St, Syracuse, NY 13210, USA
e-mail: faraones@upstate.edu

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is a neurocognitive disorder with a high worldwide prevalence [22]. For decades, the stimulant medications methylphenidate, dexamfetamine, and mixed amfetamine salts have been the most common drugs used in the treatment of ADHD. The stimulants as a class increase the availability of synaptic dopamine [45, 46], but the mechanism involved differs between methylphenidate and amfetamines. Methylphenidate can be viewed as a dopamine reuptake inhibitor, which facilitates dopaminergic neurotransmission at the dopamine transporter, and elicits little presynaptic dopamine release [35]. In contrast, amfetamines are thought to block the reuptake of both norepinephrine and dopamine into the presynaptic neuron and to facilitate neurotransmitter release through reverse transport [9, 39, 43].

Therapeutic effects of stimulants include a reduction of the hyperactivity, impulsivity, and inattention characteristic of patients with ADHD, and improvement of associated behaviors, including on-task behavior, academic performance, and social functioning [26]. Studies demonstrate

robust effects in both children and adults [41], and long-acting formulations extend the action of these medications over 8–13 h to allow once-daily dosing [4, 25, 49].

While the stimulants 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.

When the results of clinical trials are statistically significant, comparative treatment choices should not be made based on comparisons of statistical significance. The reason is that the magnitude of statistical significance is heavily influenced by the number of patients studied. Therefore, it is possible for a small trial of a highly effective treatment to have a less statistically significant result than a large trial of a modestly effective treatment. Thus, while the results of statistical analyses provide crucial information, the magnitude of statistical significance does not necessarily indicate the magnitude of the treatment effect. As such, it is impossible to determine from the degree of statistical significance how, for example, a novel treatment evaluated in one study compares in relative terms to the efficacy of other established or emerging treatments for the same condition.

This interpretative problem with statistical significance can be addressed using the concept of "effect size," which was developed to allow clinically meaningful comparisons of efficacy between treatment trials. The effect size can help clinicians decide whether the often modest increases in efficacy of newer treatments are important enough to influence clinical decisions. This is done by referencing acceptable effects of widely recognized treatments for specific disorders. Without using this concept, comparing two treatment trials can be difficult. As the name suggests, an effect size estimate places an interpretable value on the direction and magnitude of an effect of a treatment. This measure of effect can then be used to compare the efficacy of the treatment in question with similarly computed measures of effect of treatment efficacy in other studies that may use seemingly non-comparable measures.

Meta-analysis provides a systematic quantitative framework for comparing the effect sizes reported by different studies. 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, a value of zero means that there is no drug–placebo difference and a value of one means that the drug and placebo groups differ by 1 standard deviation (SD) on the outcome measure. Cohen [5] offered the following guidelines for interpreting the magnitude of the standardized mean differences (SMD) in the social sciences: small, SMD = 0.2; medium, SMD = 0.5 and large, SMD = 0.8.

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 a study of one drug using double-blind methodology found a smaller effect size than a study of a second drug that was not blinded, we could not be sure whether the difference in effect size were due to differences in drug efficacy or differences in methodology. 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 stimulant therapy of ADHD. A prior meta-analysis of ADHD medications found a significant increased efficacy of stimulant medications when compared with non-stimulants such as atomoxetine [19]. Although there have been prior meta-analyses of stimulants and reviews of effect size for long-acting formulations [3], these have been limited by a focus on one stimulant or failing to take into account methodological differences among studies [15, 17, 19, 23, 38]. Moreover, prior studies have not compared methylphenidate and amfetamine while also addressing confounding study design variables. Differences in effect sizes between stimulants can be anticipated, given the mechanism of action involved in increasing synaptic dopamine differs between methylphenidate and amfetamines. Thus, we sought to extend the available literature by determining whether (a) available studies provide evidence for significant differences in effect sizes between methylphenidate and amfetamine products and (b) if features of study design influence estimates of medication efficacy.

Materials and methods

A literature search was conducted to identify double-blind, placebo-controlled studies of ADHD in children and

adolescents published in English after 1979. We searched for articles using the following search engines: PubMed, Ovid, ERIC, CINAHL, Medline, PREMEDLINE, EMBASE, the Cochrane database, e-psyche, and social sciences abstracts. We included studies that (1) evaluated a stimulant medication for the treatment of ADHD in children and adolescents; (2) were published in English after 1979. (3) Used randomized, double-blind methodology with placebo controls; (4) defined ADHD using diagnostic criteria from the *Diagnostic and Statistical Manual of Mental Disorders* second edition (DSM-II), third edition (DSM-III), revised third edition (DSM-III-R) or fourth edition (DSM-IV); (5) followed subjects for 2 weeks or more, and (6) presented the mean and SD of either change or end point scores for the drug and placebo groups. For studies presenting data on more than one fixed dose, we used the highest dose. We excluded studies that rated behavior in laboratory environments, were designed to explore appropriate doses for future work, or selected ADHD samples for the presence of a comorbid condition (e.g., studies of ADHD among mentally retarded children).

All articles were completely read by one of the authors (SVF). The following data were extracted: 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 (parent, teacher, 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); year of publication; number of sites (single vs. multisite), and use of last observation carried forward (LOCF) methodology (yes/no). Data were extracted by reading the articles, identifying the needed information and entering that information into an Excel spreadsheet.

Effect sizes for dependent measures in each study were expressed as SMD. 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 SD of the groups. Studies reporting change scores provided end point minus baseline scores for drug and placebo groups. In this case, the SMD is computed as the difference between change scores. For studies reporting end point scores, the SMD is computed as the difference between end point scores. Studies were weighted according to the number of participants included. Our meta-analysis used the random effects model of DerSimonian and Laird [8]. We use the I^2 index to assess the heterogeneity of effect sizes [29]. Its value lies between 0 and 100 and estimates the percentage of variation among effect sizes that can be attributed to heterogeneity. A significant I^2 suggests that the effect sizes analyzed are not estimating the same population effect size. We used meta-analytic regression to assess the degree to which the effect sizes varied with the methodologic features of each study [28, 31]. We used Egger's [12] method to assess for publication biases and adjusted SMDs for publication bias using the "trim and fill" method of Duval and Tweedie [10].

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 one 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 inaccurate P values. To address this intra-family clustering, variance estimates were adjusted using Huber's [30] formula as implemented in STATA [42]. This formula is a "theoretical bootstrap" that produces robust statistical tests. The method works by entering the cluster scores (i.e., sum of scores within families) into the formula for the estimate of variance. The resulting P values are valid even when observations are not statistically independent.

We also computed the number needed to treat (NNT) effect size. For binary outcomes, the NNT is the number of patients who need to be treated to prevent one failure to respond. Because binary outcomes were rarely presented in the reviewed studies, we computed the NNT from quantitative outcomes. For quantitative outcomes, the NNT is the number of patients one needs to treat to have a successful outcome. In this context, a "successful outcome" means that a medication-treated patient responded better than a randomly selected placebo-treated patient [32]. To compute the NNT, one must first compute the probability of benefit (POB), which is defined as the probability that a randomly selected treated patient will show a level of improvement exceeding that of a randomly selected placebo patient [34]. The probability of benefit is equivalent to the area under the receiver operating characteristic curve and is also proportional to the Mann–Whitney statistic comparing drug and placebo outcome scores[6, 27]. For details, see Faraone et al. [20], Kraemer and Kupfer [32], and prior applications of the method [16, 18, 20, 21, 40, 44]. To compute the POB, one-first computes $Z = \text{SMD}/\sqrt{2}$. This Z statistic is distributed as a standard normal distribution and the probability of benefit is computed as the probability that a randomly selected standard normal variable is less than Z. We then compute $\text{NNT} = 1/(2 \times \text{POB} - 1)$ [32].

Results

Table 1 describes the 23 articles meeting the criteria for inclusion in the meta-analysis. Studies are listed more than

Table 1 Study features

First author	Publication date	Drug	Dosing method	*N* in drug group	*N* in placebo group	Mean age	% Male	DSM version
Klorman	1987	MPH	Fixed	19	19	15	84	3
Taylor	1987	MPH	Best	37	37	9	100	3
Douglas	1988	MPH	Fixed	19	19	9	89	3
Arnold	1989	D-Amph	Fixed	18	18	–	100	3
Klorman	1994	MPH	Best	44	44	9	84	3R
Schachar	1997	MPH	Best	37	29	8	77	3R
Manos	1999	MAS	Best	42	42	10	79	4
Manos	1999	MPH	Best	42	42	10	79	4
Zeiner	1999	MPH	Best	36	36	9	100	3R
Pliszka	2000	MAS	Best	20	18	8	–	4
Pliszka	2000	MPH	Best	20	18	8	–	4
James	2001	MAS	–	35	35	9	60	4
James	2001	D-Amph	–	35	35	9	60	4
James	2001	D-Amph ER	–	35	35	9	60	4
Wolraich	2001	MPH	Best	97	90	9	87	4
Wolraich	2001	OROS MPH	Best	95	90	9	78	4
Biederman	2002	MAS-XR	Fixed	120	203	9	80	4
Greenhill	2002	MPH-MR	Best	155	159	9	83	4
Biederman	2003	MPH-LA	Best	63	71	9	77	4
Wigal	2004	MPH	Best	41	41	10	88	4
Wigal	2004	D-MPH	Best	42	41	10	88	4
Findling	2005	OROS MPH	Best	89	85	9	70	4
Wilens	2006	OROS MPH	Best	87	90	15	80	4
Spencer	2006	MAS-XR	Fixed	27	24	14	64	4
Spencer	2006	MAS-XR	Fixed	26	28	14	64	4
Findling	2006	MPH	Best	120	39	10	79	4
Findling	2006	MPH-MR	Best	120	39	10	81	4
Biederman	2007	LDX	Fixed	73	72	9	71	4
Palumbo	2008	MPH	Best	29	30	9	83	4
Findling	2008	OROS MPH	Best	78	77	9	66	4
Findling	2008	MTS	Best	82	77	9	60	4
Newcorn	2008	OROS MPH	Best	220	74	10	71	4

Studies are listed multiple times if they studied more than one drug

MPH Methylphenidate, *MAS* mixed amfetamine salts, *NA* not available, *D-Amph* dextroafhetamine, *ER* extended release, *OROS* osmotic release oral system, *XR* extended release, *MR* modified release, *LA* long acting, *D-MPH* dexmethylphenidate, *MTS* methylphenidate transdermal system, *LDX* lisdexamfetamine dimesylate

once if they studied more than one drug or if they reported independent studies of the same drug. The studies in Table 1 evaluated 11 drugs using 19 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 99 effect sizes. We categorized these scores into three subgroupings: total ADHD symptom scores ($N = 73$), inattention subscale scores ($N = 9$), and hyperactivity-impulsivity subscale scores ($N = 17$). As these varying N's indicated, most studies did not provide information on all of these scores. 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.

We used regression analyses to determine if any of the potentially confounding variables were associated with SMDs. The following variables were not significantly associated with the SMDs: date of study publication, duration of action (short vs. long-acting stimulant); gender, DSM-IV ADHD subtype (for DSM-IV studies only), type of dosing (the use of fixed dose vs. titration to best dose designs), use of a placebo lead-in, type of design (parallel vs.

Table 2 Drug–placebo comparisons tested in the meta-analysis studies

Medication	No. of studies of each medication	Number of subjects in drug group	Number of subjects in placebo group
Amfetamine			
MAS	3	97	95
MAS XR	2	147	227
D-Amph	2	53	53
D-Amph ER	1	35	35
LDX	1	73	72
Methylphenidate			
MPH	12	546	444
MPH MR	2	275	198
OROS® MPH	5	581	414
D-MPH	1	42	41
MTS	1	96	85
MPH-LA	1	63	71

The table includes crossover studies for which the medication and placebo subjects were the same; some studies examined more than one medication

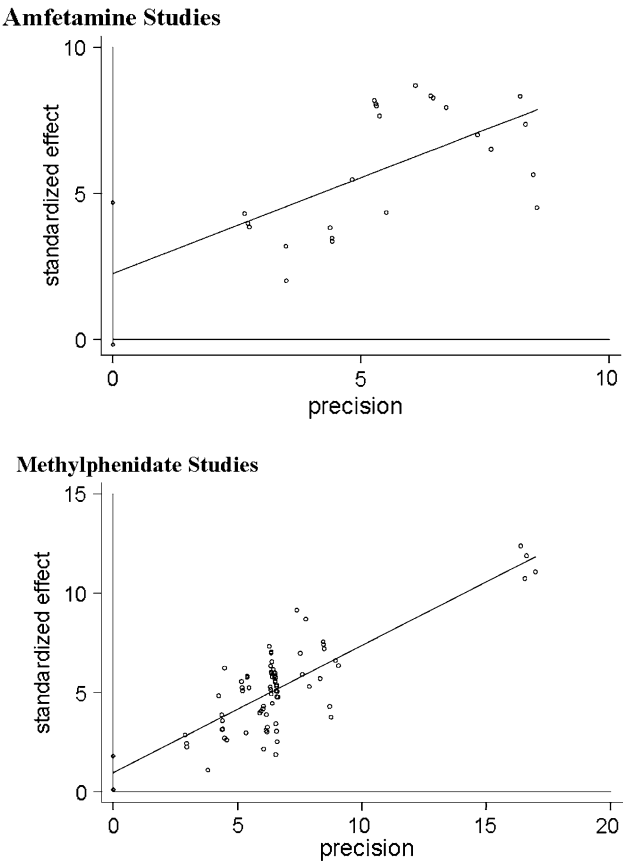

Fig. 1 Publication bias plots

crossover), and use of LOCF methodology or use of multiple sites (all P's > 0.12). We did find a significant association of the SMDs with three variables. SMDs were greater for children compared with adolescents [SMD = 0.89 vs. 0.64; $t(21) = 4.3$, $P < 0.001$]. Teacher (0.92) and physician raters (0.96) had higher SMDs than parent (0.73) or self (0.47) raters [$F(3,22) = 26.5$, $P < 0.001$]. We also found that studies presenting outcome scores had higher SMDs (0.93) than studies presenting change scores [0.75; $t(22) = 2.4$, $P = 0.03$]. Additionally, effect sizes were negatively correlated with the length of the study protocol ($r = -0.29$, $P = 0.034$). This latter finding may be due to the fact that the longer-duration studies tended to report change scores, which tend to have lower SMDs. These potentially confounding variables did not significantly differ between the amfetamine and methylphenidate studies (all P's > 0.20), and after controlling for these potentially confounding variables, we found that the SMDs for studies of amfetamine were significantly greater than the SMDs for studies of methylphenidate [$t(21) = 1.4$, $P = 0.008$].

Using Egger's test, we found no significant publication bias for studies of amfetamine [$t(23) = 3.3$, $P = 0.07$] but did find significant publication bias for studies of methylphenidate [$t(74) = 2.2$, $P = 0.03$]. The lack of significance for amfetamine studies could be due to the small number of such studies. In this regard, it is notable that the publication bias statistics (and their 95% confidence intervals are 2.3 {−0.17, 4.7} for amfetamine and 0.9 {0.09, 1.8} for methylphenidate. Thus, the estimated bias is greater, albeit not significant, for amfetamine studies and, because the confidence intervals for the two drug classes overlap, one cannot conclude that one type shows more or less bias than the other. The nature of the publication bias can be seen in Fig. 1. Each of the two graphs in Fig. 1 plots the standardized SMD against the precision of the study. In the absence of publication biases, the more precise studies should yield larger standardized effects and the regression of standardized effect on precision should intersect the horizontal axis at zero. The publication bias statistics given above is simply the point on the horizontal axis intersected by the regression line. The graphs show the point of intersection and the 95% confidence intervals describe above. Because the lines do not intersect zero, we can conclude that some studies are missing due to publication bias. These missing studies, if published, would have fallen below the regression lines in the lower left quadrant of each graph, which would have the effect of moving the intercept of the regression toward zero. Thus, the missing studies are studies with relatively low precision that estimated smaller effect sizes than the published studies of similar precision. We estimated the results of these missing studies using the "trim and fill" method [10]. Doing so reduced the amfetamine effect size from 1.10 to 0.99. For methylphenidate,

the SMD was reduced from 0.79 to 0.72. Although the average reduction for amfetamine studies (0.11) was greater than the reduction for methylphenidate studies (0.07), the difference between the two medications remained significant after subtracting these reductions from each SMD [$t(21) = 2.8$, $P = 0.01$].

The results for the meta-analyses are presented graphically in Figs. 2, 3, 4, for different types of ADHD outcome scores (total scores, global ratings, hyperactive-impulsive scores and inattentive scores). The left column of each figure gives the first author and date of the relevant article and in the second and third columns, the drug studied and the rater making the rating, respectively; entries in the left column are duplicated when the study provided more than one estimate of effect size. The right side of each graph gives the effect size (standardized mean difference) as a dark dot. The size of the shaded box around the box is proportional to the sample size. The horizontal line through the box gives the 95% confidence interval. The results obtained when pooling within drug types and across the entire sample are given by diamonds. The center of the diamond marks the estimate of the pooled effect size. The left and right ends of the diamond mark the 95% confidence interval.

The difference between medication groups was significant for measures that assessed all ADHD symptoms [Fig. 1: SMD = 1.03 vs. 0.77, $t(19) = 2.5$, $P = 0.02$], and hyperactive-impulsive symptoms [Fig. 2; SMD = 1.20 vs. 0.91; $t(7) = 3.5$, $P = 0.01$]. Because there was only one amfetamine study [lisdexamfetamine dimesylate (LDX)] assessing inattentive symptoms, we could not do a statistical comparison but note that the effect size for the only amfetamine study of inattentive symptoms had a greater effect size than all the methylphenidate studies (SMD = 1.52 vs. 0.84) with little overlap of the 95% confidence intervals (See Fig. 3). The small increased efficacy of amfetamine over methylphenidate is also seen when using the NNT. For amfetamine studies, the NNT was 2.0 with a 95% confidence interval of {1.7, 2.2}. For methylphenidate, the NNT was 2.6 with a 95% CI of {2.4, 2.8}. The NNT is defined as the number of patients that one needs to treat to have a successful outcome. In this context, a "successful outcome" means that a medication-treated patient responded better than a randomly selected placebo-treated patient [32].

We found substantial heterogeneity of the SMD among studies for ADHD total scores. The I^2 heterogeneity statistics were 74.5% ($P < 0.001$) for amfetamine and 45.4% ($P < 0.001$) for methylphenidate. For hyperactive-impulsive scores the I^2 statistic indicated no significant heterogeneity for amfetamine ($I^2 = 0.0\%$; $P = 0.52$) but did indicate significant and substantial heterogeneity for methylphenidate ($I^2 = 68.5\%$; $P = 0.001$). For inattentive ratings, we could not assess heterogeneity for amfetamine as only one data point was available. For methylphenidate, the heterogeneity statistic was not significant ($I^2 = 0.0\%$, $P = 0.83$).

We found no statistically significant evidence of publication bias for amfetamine studies of total ADHD symptoms [$t(15) = 1.9$, $P = 0.08$] or methylphenidate studies of total ADHD symptoms [$t(56) = 0.9$, $P = 0.35$]. We found significant evidence of publication bias for amfetamine studies of hyperactive-impulsive symptoms [$t(6) = 5.4$, $P = 0.003$]. After adjusting the SMD for this bias, it was reduced to 1.15. We found significant evidence of publication bias for methylphenidate studies of hyperactive-impulsive symptoms [$t(9) = 4.8$, $P = 0.001$]. After adjusting the SMD for this bias, it was reduced to 0.73. We also found significant evidence of publication bias for methylphenidate studies of inattentive symptoms [$t(6) = 4.0$, $P = 0.007$]. After adjusting the SMD for this bias, it was reduced to 0.73. We could not assess publication bias for amfetamine studies of inattentive symptoms as there was only one study.

Discussion

Our meta-analysis of ADHD efficacy outcomes found significant differences between amfetamine and methylphenidate products, even after correcting for study design features that might have confounded the results. Our analyses indicate that effect sizes for amfetamine products are statistically, albeit moderately, greater than those for methylphenidate. The robust effects of all stimulant medications can be seen in Figs. 1, 2, 3, which show that most measures of effect from all studies were statistically significant. These figures also show a good deal of variability among studies with overlapping confidence intervals between many methylphenidate and amfetamine studies, which is to be expected given the small differences in the mean SMDs between medication groups.

The results from the NNT statistic are instructive. According to the NNT results, when using amfetamine, clinicians need to treat two patients for each positive outcome for total ADHD symptoms, but for methylphenidate, an average of 2.6 need to be treated. The NNT data also allow us to address the costs of treatment options. One approach to this issue is to consider the costs of failed treatments. A simple way to compute the probability of a failed treatment would be to use the observed failure rate in the drug treatment group. But this measure is not appropriate because it does not take into account the placebo response rate. We can compute the probability of a failed treatment adjusted for the placebo response as (NNT-1)/NNT, which is equivalent to adding the failure rate in the drug group to the response rate in the placebo group. In this

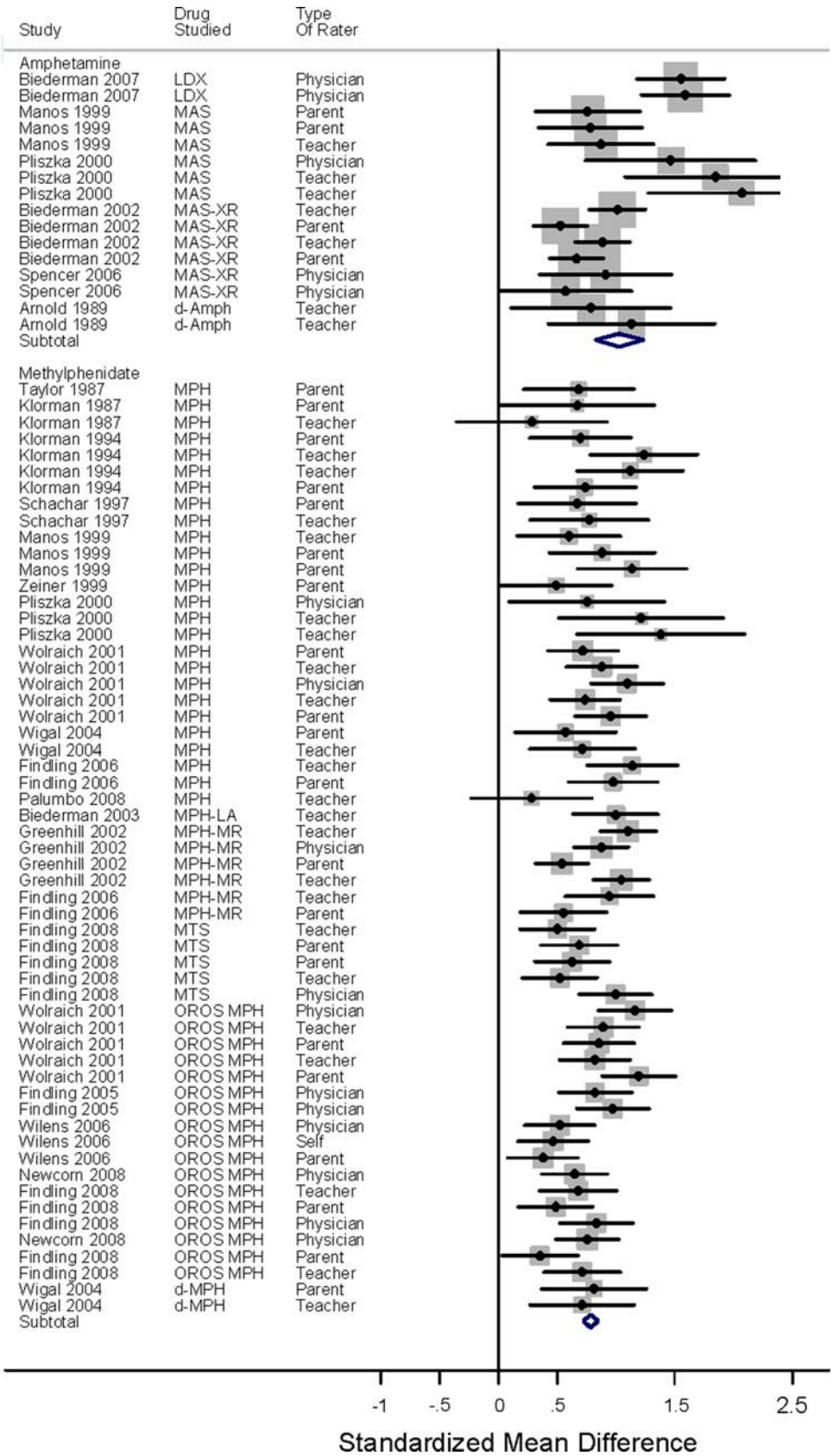

Fig. 2 Effect sizes and 95% confidence intervals (CIs) for total ADHD symptoms. Note: see text for description of graph

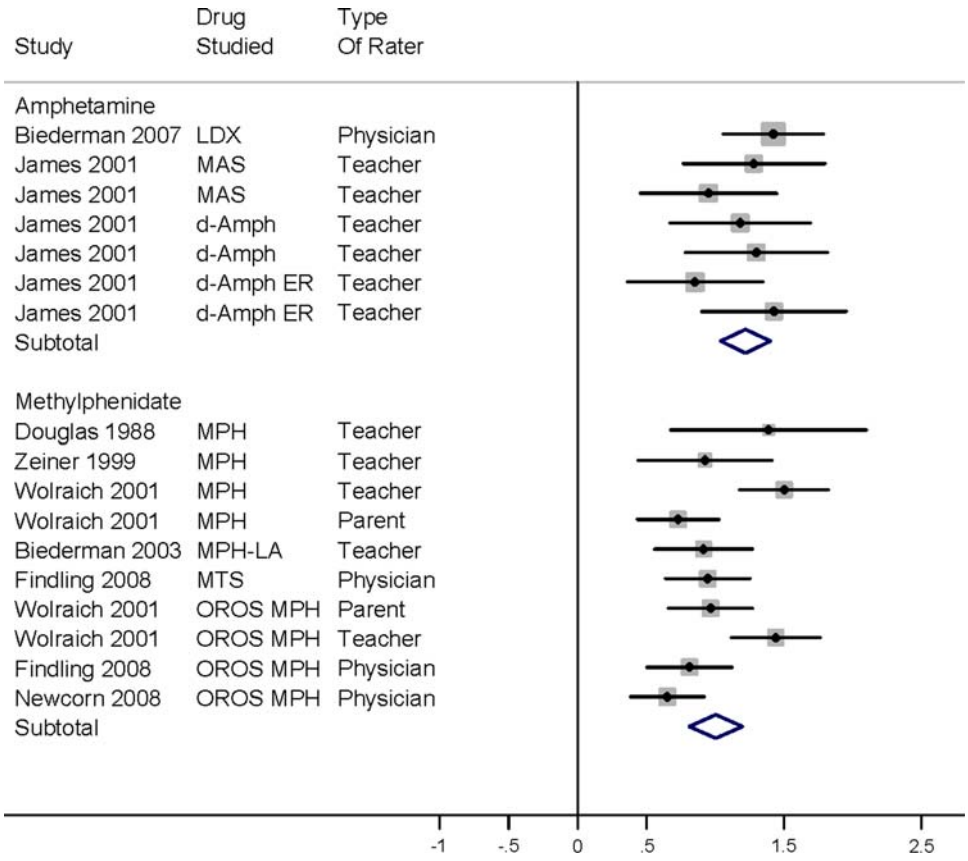

Fig. 3 Effect sizes and 95% confidence intervals (CIs) for hyperactivity-impulsivity. Note: see text for description of graph

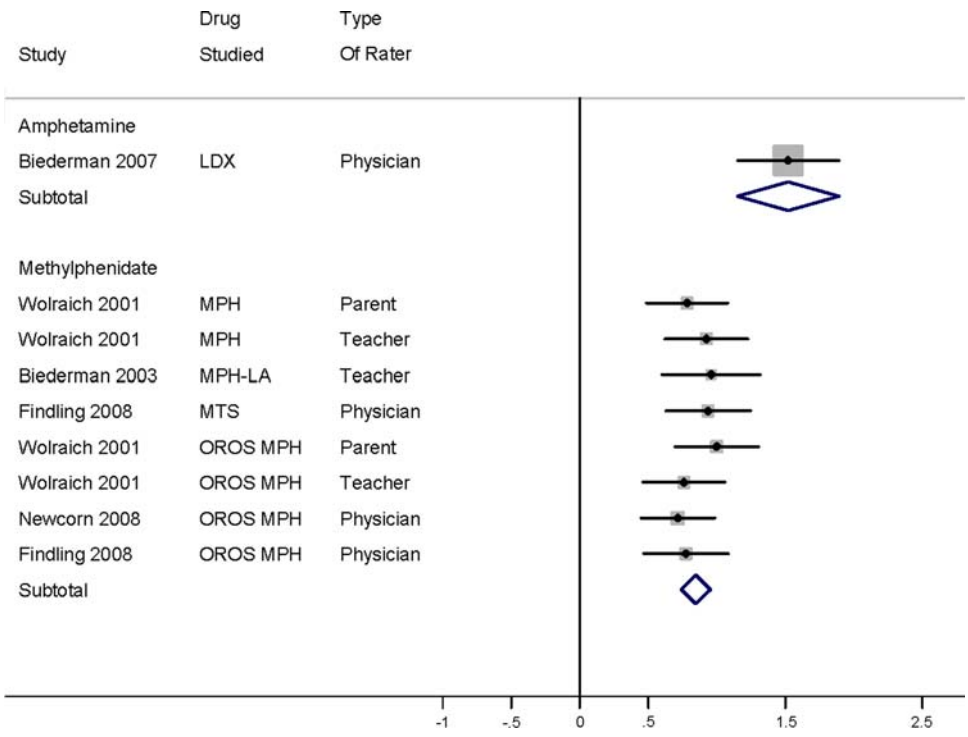

Fig. 4 Effect sizes and 95% confidence intervals (CIs) for inattention. Note: see text for description of graph

context, a treatment failure occurs when a drug-treated patient has a worse outcome than a randomly selected placebo treated patient [32].

Now, consider a health care system that treats 100,000 patients annually. Because we can compute the probability of a treatment failure, we can easily compute the number of

wasted treatments as a function of the NNT. The amfetamine NNT of two translates into a treatment failure probability of 50%, which means that for every 100,000 treatments, 50,000 will be wasted. For methylphenidate, the NNT is 2.6 and the probability of treatment failure is 62%, which means that for every 100,000 treatments, 62,000 will be wasted. The failure rates for each class of medication are higher than the observed failure rates in the drug groups because, as discussed above, they are adjusted for the placebo response. These results illustrate the fact that small differences in the NNT can have marked differences in the overall costs to health care systems.

There are few direct comparisons of amfetamine and methylphenidate in the literature, but Arnold et al. [2] reported a double-blind placebo-controlled crossover comparison of methylphenidate and dexamfetamine in 29 children with minimal brain dysfunction (MBD). Of 26 drug responders, 12 responded best to dexamfetamine, ten to methylphenidate and four to neither. Using subjects from a day hospital school, Elia et al. [14] assessed the effects of methylphenidate and dexamfetamine on mathematics testing in 33 hyperactive boys. Both drugs increased the number of attempts made to solve math problems but only dexamfetamine improved the percent correct. In a summer camp program, Pelham et al. [36] compared immediate release methylphenidate, a sustained-release form of methylphenidate (SR-20 Ritalin) and a sustained-release form of dexamfetamine (Dexedrine Spansule). Dependent measures include evaluations of social behavior during group recreational activities, classroom performance, and performance on a continuous performance task. The authors found equivalent and beneficial effects of all four medications, but sustained-release dexamfetamine produced a more consistent effect than sustained release methylphenidate. Elia et al. [13] studied 48 ADHD boys using a double-blind crossover of methylphenidate, dexamfetamine, and placebo. There were no significant group differences on any of the outcome measures. Based on efficacy and adverse events, there was no significant preference for either drug at the end of the trial. The response rate to dexamfetamine was non-significantly greater than the response to methylphenidate (88 vs. 79%) and the rates of adverse events were similar. Arnold [1] reviewed comparative studies of methylphenidate and amfetamine completed before 1997. He found a small advantage of amfetamine over methylphenidate. Of 174 patients in six crossover studies, he reported that 48 responded best to amfetamine, 27 responded best to methylphenidate, and 72 responded equally to both. He concluded that 87% of ADHD children should respond well to one of the two stimulant classes.

Efron et al. [11] completed a double-blind, crossover trial of methylphenidate versus dexamfetamine in 125 ADHD children. There were significant group mean improvements from baseline score on all measures for both stimulants. Although the drugs did not differ in their efficacy for treating inattentive symptoms, response was better for methylphenidate for teacher ratings of conduct problems and hyperactivity. Parents rated 73% of subjects as globally improved on MPH and 69% improved on dexamfetamine, compared with baseline and 46% of parents chose methylphenidate as the preferred drug, compared with 37% who chose dexamfetamine. The two drugs did not differ on continuous performance test measures of correct responses or errors. Manos et al. [33] also reported a double-blind, placebo-controlled study of MAS-IR and methylphenidate. No differences between methylphenidate and MAS-IR were observed for either teacher or parent ratings of behavior or side effects.

Pliszka et al. [37] reported a 3-week, double-blind, placebo-controlled study of immediate release mixed amfetamine salts (MAS-IR) and methylphenidate. MAS-IR produced significantly more improvements on teacher ratings and the clinical global impressions scale than did methylphenidate. There was a trend for MAS-IR to be associated with more sadness and stomachaches. Other adverse events did not differ between groups. In a study of driving, adolescent drivers with ADHD were compared on a driving simulator after taking 72 mg of OROS methylphenidate, 30 mg of mixed amfetamine salts extended release, or placebo in a randomized, double-blind, placebo-controlled, crossover study design [7]. On an overall measure of driving impairment, OROS methylphenidate led to better driving performance compared with placebo and mixed amfetamine salts extended release; mixed amfetamine salts extended release demonstrated no statistical improvement over placebo.

Wilson et al. [48] studied the effects of two long-acting stimulant medications (MAS-XR and OROS methylphenidate) on neuropsychological functioning among ADHD adolescents. They studied two neuropsychological tasks, which measure visual memory, attention span, and response inhibition (the Delayed Matching-to-Sample and the Go/No-go tasks). There were no significant differences between the two drugs. Both medications showed some improvement in neuropsychological functioning compared with placebo, but only the comparisons with OROS reached statistical significance.

When taken together, our meta-analysis and the prior head-to-head studies of amfetamine and methylphenidate suggest the former may be modestly more efficacious. This conclusion, however, has some limitations. First, as Figs. 1, 2, 3 show, there is a good deal of variability among studies. Second, methylphenidate may have some advantages for specific outcomes as was seen in the driving simulator study by Cox et al. [7]. Also, when choosing

medications, efficacy is not the only issue that should be taken into account. For example, Green [24] reported an open label crossover trial of 100 ADHD children who had been treated with both dexamfetamine and methylphenidate. Eighty-two of the 100 parents had a clear preference for one drug over the other. Among these, 61 preferred methylphenidate and 21 preferred dexamfetamine. These preference ratings were a combined reaction to both the efficacy and side effects of the medications.

We found 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 study design variables. Comparing medication effect sizes in different studies will lead to spurious conclusions without accounting for these influences. We found that three study design variables, age of subject, type of score (change score or outcome score), and rater (physician vs. parent vs. teacher vs. self) differed significantly among the medication groups and was also predictive of the effect size. When we adjusted for these differences using meta-analysis regression, however, we continued to find significant differences between amfetamine and methylphenidate. Of note, effect sizes are lower for studies of adolescents compared with studies of children, lower for parents and self ratings compared with teachers and parents, and lower for change compared with outcome scores. These findings could be useful in the design and interpretation of studies.

This work must be interpreted in the context of several limitations. Because we relied on data presented by authors, and were limited by what these investigators chose to present, we could not assess the effects of all potential confounds. For example, we could not compute the effect sizes at specific time points, because such data were rarely provided. Future work should review studies of time course such as the analog school laboratory paradigm. Although that work does not assess outcome in the patient's environment, it would provide useful data about efficacy peak or trough effect. Similarly, we did not assess differential duration of action between medication classes, but this effect is rarely presented. 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. 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. For example, as Fig. 1 shows, the effect size in Wilens et al.'s [47] study of OROS-MPH was much lower that that seen for other OROS MPH studies and other long-acting stimulants. Wilens et al. note that because there were limitations on the highest dose used, their results might be less that expected with optimal dosing. Similarly, Zeiner et al.'s [50] study of methylphenidate reported a low effect size using doses of 0.5 mg/kg, which is relatively low by current standards. Effects of this sort that are idiosyncratic to one or a few studies cannot be adjusted for in a meta-analysis context.

Despite these limitations, our findings suggest that amfetamine products may be moderately more efficacious than methylphenidate products among children and adolescents. This difference in effect size may be due to differences between amfetamine and methylphenidate in the molecular mechanisms involved in facilitating the dopaminergic neurotransmission. 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.

Acknowledgment This work was supported by Shire Development.

Conflict of interest statement Dr. Stephen V. Faraone is/has been a consultant for, receives/d research support from, is/has been a speaker for, or is/has been on the advisory board for the following pharmaceutical companies: McNeil, Pfizer, Shire, Eli Lilly and Company, and the Novartis Corporation. Dr. Jan K. Buitelaar is/has been a consultant for, receives/d research support from, is/has been a speaker for, or is/has been on the advisory board for the following pharmaceutical companies: Janssen Cilag BV, Eli Lilly and Company, UCB, Shire, Servier, Bristol-Myer Squibb, Organon, and Bioproject.

References

1. Arnold L (2000) Methylphenidate vs. amphetamine: comparative review. J Atten Disord 3:200–211
2. Arnold LE, Christopher J, Huestis R, Smeltzer DJ (1978) Methylphenidate vs dextroamphetamine vs caffeine in minimal brain dysfunction. Arch Gen Psychiatry 35:463–473
3. Banaschewski T, Coghill D, Santosh P, Zuddas A, Asherson P, Buitelaar J, Danckaerts M, Dopfner M, Faraone SV, Rothenberger A, Sergeant J, Steinhausen HC, Sonuga-Barke EJ, Taylor E (2006) Long-acting medications for the hyperkinetic disorders: a systematic review and European treatment guideline. Eur Child Adolesc Psychiatry 15:476–495
4. Biederman J, Lopez FA, Boellner SW, Chandler MC (2002) A randomized, double-blind, placebo-controlled, parallel-group study of SLI381 in children with attention deficit hyperactivity disorder. Pediatrics 110:258–266
5. Cohen J (1988) Statistical power analysis for the behavioral sciences. Erlbaum, Hillsdale
6. Colditz GA, Miller JN, Mosteller F (1988) Measuring gain in the evaluation of medical technology. The probability of a better outcome. Int J Technol Assess Health Care 4:637–642
7. Cox DJ, Merkel RL, Moore M, Thorndike F, Muller C, Kovatchev B (2006) Relative benefits of stimulant therapy with OROS methylphenidate versus mixed amphetamine salts extended release in improving the driving performance of adolescent

drivers with attention-deficit/hyperactivity disorder. Pediatrics 118:e704–e710
8. Dersimonian R, Laird N (1986) Meta-analysis in clinical trials. Controlled Clin Trials 7:177–188
9. Dlugos A, Freitag C, Hohoff C, Mcdonald J, Cook EH, Deckert J, De Wit H (2007) Norepinephrine transporter gene variation modulates acute response to D-amphetamine. Biol Psychiatry 61:1296–1305
10. Duval S, Tweedie R (2000) A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. J Am Stat Assoc 95:89–98
11. Efron D, Jarman F, Barker M (1997) Methylphenidate versus dexamphetamine in children with attention deficit hyperactivity disorder: a double-blind, crossover trial. Pediatrics 100:E6
12. Egger M, Davey Smith G, Schneider M, Minder C (1997) Bias in meta-analysis detected by a simple, graphical test. Br Med J 315:629–634
13. Elia J, Borcherding BG, Rapoport JL, Keysor CS (1991) Methylphenidate and dextroamphetamine treatments of hyperactivity: are there true nonresponders? Psychiatry Res 36:141–155
14. Elia J, Welsh P, Gullota C, Rapoport J (1993) Classroom academic performance: Improvement with both methylphenidate and dextroamphetamine in ADHD boys. J Child Psychol Psychiatry Allied Disciplines 34:785–804
15. Faraone S, Biederman J (2002) Efficacy of adderall, for attention-deficit/hyperactivity disorder: a meta-analysis. J Atten Disord 6:69–75
16. Faraone S, Short E, Biederman J, Findling RL, Roe CM, Manos M (2002) Efficacy of Adderall and Methylphenidate in attention deficit hyperactivity disorder: a drug–placebo and drug–drug response curve analysis of a naturalistic study. Int J Neuropsychopharmacol 5:121–129
17. Faraone SV, Biederman J, Roe CM (2002) Comparative efficacy of adderall and methylphenidate in attention-deficit/hyperactivity disorder: a meta-analysis. J Clin Psychopharmacol 22:468–473
18. Faraone SV, Biederman J, Spencer T, Michelson D, Adler L, Reimherr F, Glatt S (2005) Efficacy of atomoxetine in adult attention-deficit/hyperactivity disorder: a drug–placebo response curve analysis. Behav Brain Funct 1:16
19. Faraone SV, Biederman J, Spencer TJ, Aleardi M (2006) Comparing the efficacy of medications for ADHD using meta-analysis. Medscape Gen Med E J 8:4
20. Faraone SV, Biederman J, Spencer TJ, Wilens TE (2000) The drug–placebo response curve: a new method for assessing drug effects in clinical trials. J Clin Psychopharmacol 20:673–679
21. Faraone SV, Pliszka SR, Olvera RL, Skolnik R, Biederman J (2001) Efficacy of adderall and methylphenidate in attention-deficit/hyperactivity disorder: a drug–placebo and drug–drug response curve analysis. J Child Adolesc Psychopharmacol 11:171–180
22. Faraone SV, Sergeant J, Gillberg C, Biederman J (2003) The worldwide prevalence of ADHD: is it an American condition? World Psychiatry 2:104–113
23. Faraone SV, Spencer T, Aleardi M, Pagano C, Biederman J (2004) Meta-analysis of the efficacy of methylphenidate for treating adult attention deficit hyperactivity disorder. J Clin Psychopharmacol 54:24–29
24. Green C (1996) Methylphenidate vs dexamphetamine: a clinical audit. J Paediatr Child Health 32:470
25. Greenhill LL, Findling RL, Swanson JM (2002) A double-blind, placebo-controlled study of modified-release methylphenidate in children with attention-deficit/hyperactivity disorder. Pediatrics 109:E39
26. Greenhill LL, Pliszka S, Dulcan MK, Bernet W, Arnold V, Beitchman J, Benson RS, Bukstein O, Kinlan J, Mcclellan J, Rue D, Shaw JA, Stock S, Kroeger K (2001) 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 40:1352–1355
27. Hanley JA, Mcneil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29–36
28. Hedges LV, Olkin I (1985) Statistical methods for meta-analysis. Academic Press, Orlando
29. Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. BMJ 327:557–560
30. Huber PJ (1967) The behavior of maximum likelihood estimates under non-standard conditions. Proc Fifth Berkeley Symp Math Stat Probab 1:221–233
31. Hunter JE, Schmidt FL (1990) Methods of meta-analysis: correcting error and bias in research findings. Sage, Newbury Park
32. Kraemer HC, Kupfer DJ (2006) Size of treatment effects and their importance to clinical research and practice. Biol Psychiatry 59:990–996
33. Manos MJ, Short EJ, Findling RL (1999) Differential effectiveness of methylphenidate and Adderall in school-age youths with attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry 38:813–819
34. Mcgraw KO, Wong SP (1992) A common language effect size statistic. Psychol Bull 111:361–365
35. Patrick KS, Caldwell RW, Ferris RM, Breese GR (1987) Pharmacology of the enantiomers of threo-methylphenidate. J Pharmacol Exp Ther 241:152–158
36. Pelham W, Greenslade K, Vodde-Hamilton M, Murphy D, Greenstein J, Gnagy E, Guthrie K, Hoover M, Dahl R (1990) Relative efficacy of long-acting stimulants on children with attention deficit-hyperactivity disorder: a comparison of standard methylphenidate, sustained-release methylphenidate, sustained-release dextroamphetamine, and pemoline. Pediatrics 86:226–237
37. Pliszka SR, Browne RG, Olvera RL, Wynne SK (2000) A double-blind, placebo-controlled study of Adderall and methylphenidate in the treatment of attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry 39:619–626
38. Schachter HM, Pham B, King J, Langford S, Moher D (2001) How efficacious and safe is short-acting methylphenidate for the treatment of attention-deficit disorder in children and adolescents? A meta-analysis. CMAJ 165:1475–1488
39. Seiden LS, Sabol KE, Ricaurte GA (1993) Amphetamine: effects on catecholamine systems and behavior. Annu Rev Pharmacol Toxicol 33:639–677
40. Short EJ, Manos MJ, Findling RL, Schubel EA (2004) A prospective study of stimulant response in preschool children: insights from ROC analyses. J Am Acad Child Adolesc Psychiatry 43:251–259
41. Spencer T, Biederman J, Wilens T (2000) Pharmacotherapy of attention deficit hyperactivity disorder. Child Adolesc Psychiatr Clin N Am 9:77–97
42. Stata Corporation (2001) Stata user's guide: release 7.0. Stata Corporation, College Station
43. Thornburg JE, Moore KE (1973) The relative importance of dopaminergic and noradrenergic neuronal systems for the stimulation of locomotor activity induced by amphetamine and other drugs. Neuropharmacology 12:853–866
44. Tiihonen J, Hallikainen T, Ryynanen OP, Repo-Tiihonen E, Kotilainen I, Eronen M, Toivonen P, Wahlbeck K, Putkonen A (2003) Lamotrigine in treatment-resistant schizophrenia: a randomized placebo-controlled crossover trial. Biol Psychiatry 54:1241–1248
45. Volkow ND, Fowler JS, Wang G, Ding Y, Gatley SJ (2002) Mechanism of action of methylphenidate: insights from PET imaging studies. J Atten Disord 6:S31–S43

46. Volkow ND, Wang G, Fowler JS, Logan J, Gerasimov M, Maynard L, Ding Y, Gatley SJ, Gifford A, Franceschi D (2001) Therapeutic doses of oral methylphenidate significantly increase extracellular dopamine in the human brain. J Neurosci 21:RC121
47. Wilens T, Mcburnett K, Bukstein O, Mcgough J, Greenhill L, Lerner M, Stein MA, Conners CK, Duby J, Newcorn J, Bailey CE, Kratochvil CJ, Coury D, Casat C, Denisco MJ, Halstead P, Bloom L, Zimmerman BA, Gu J, Cooper K, Lynch JM (2006) Multisite, controlled trial of OROS® methylphenidate (CONCERTA®) in the treatment of adolescents with attention-deficit/hyperactivity disorder. Arch Pediatr Adolesc Med 160:82–90
48. Wilson HK, Cox DJ, Merkel RL, Moore M, Coghill D (2006) Effect of extended release stimulant-based medications on neuropsychological functioning among adolescents with attention-deficit/hyperactivity disorder. Arch Clin Neuropsychol 21:797–807
49. Wolraich M, Greenhill LL, Pelham W, Swanson J, Wilens T, Palumbo D, Atkins M, Mcburnett K, Bukstein O, August G (2001) Randomized controlled trial of OROS methylphenidate qd in children with attention deficit/hyperactivity disorder. Pediatrics 108:883–892
50. Zeiner P, Bryhn G, Bjercke C, Truyen K, Strand G (1999) Response to methylphenidate in boys with attention-deficit hyperactivity disorder. Acta Paediatr 88:298–303