

# The Impact of Failed Antidepressant Trials on Outcomes in Children and Adolescents with Anxiety and Depression: A Systematic Review and Meta-Analysis

Sarah A. Mossman, MACPR,<sup>1</sup> Jeffrey A. Mills, PhD,<sup>2</sup> John T. Walkup, MD,<sup>3</sup> and Jeffrey R. Strawn, MD<sup>1</sup>

## Abstract

**Objective:** To identify predictors of medication-placebo differences in double-blind placebo-controlled antidepressant trials in children and adolescents with anxiety and depression.

**Methods:** Clinical trials in patients <18 years of age with major depressive disorder or generalized, separation or social anxiety disorders were obtained from PubMed, the Cochrane Database and clinicaltrials.gov searches from inception through 2019. Forty-nine trials (43 published and 6 unpublished) of anxiety ( $\kappa = 13$ ) and depression ( $\kappa = 36$ ) evaluated 19 antidepressants in 8642 child and adolescent patients; placebo and medication response rates, trial characteristics, disorder, medication class, and funding source were extracted. Antidepressant-placebo differences were examined using Bayesian hierarchical models and estimates of response were determined for trial design, disorder, and medication class variables. Using meta-regression, correlates of antidepressant-placebo difference and placebo response were examined.

**Results:** Funding source differentiated medication-placebo differences regardless of disorder. Industry trials had larger placebo response rates (mean difference:  $0.189 \pm 0.066$ , credible interval [CrI]: 0.067 to 0.33,  $p = 0.0008$ ) and smaller medication-placebo differences ( $-0.235 \pm 0.078$ , CrI:  $-0.397$  to  $-0.086$ ,  $p = 0.005$ ) compared with federally funded trials. However, medication response was similar for industry- and federally-funded studies ( $-0.046 \pm 0.042$ , CrI:  $-0.130$  to 0.038,  $p = 0.252$ ).

**Conclusions:** The impact of study sponsorship on trial outcome supports the assertion that industry-funded trials with high placebo response rates and small drug-placebo differences are “failed trials” and should not be described as “negative trials” or used to determine public health estimates of antidepressant efficacy in children and adolescents with anxiety and depression. Identifying the proper role and value of industry-funded trials is critical to establishing the evidence base for antidepressants in youth.

**Keywords:** antidepressant, depression, pediatric, placebo response, funding

## Introduction

THE RANDOMIZED DOUBLE-BLIND placebo-controlled trial (DBPCT) has become the standard for evaluating medication efficacy (Bothwell and Podolsky 2016). However, trial design and study implementation directly influence the validity of trial data and our ability to detect medication-placebo differences. Reviews and meta-analyses of antidepressant trials in children and adolescents with anxiety and depression often fail to observe antide-

pressant efficacy, particularly in depression trials. Many trials included in these meta-analyses have active drug response rates in the 55%–60% range, as well as placebo response rates that sometimes approach 60% and small medication-placebo differences ( $\sim 10\%$ ), which account for the apparent lack of efficacy. Frequently, these trials are deemed “negative” trials leading to the conclusion that antidepressants are ineffective. However, the quality of study implementation in the majority of these trials has been questioned suggesting that studies with high placebo response rates

<sup>1</sup>Anxiety Disorders Research Program, Department of Psychiatry and Behavioral Neuroscience, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA.

<sup>2</sup>Department of Economics, Carl H. Lindner College of Business at the University of Cincinnati, Cincinnati, Ohio, USA.

<sup>3</sup>Department of Psychiatry and Behavioral Health, Robert and Anne Lurie Children’s Hospital of Chicago, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA.

**Presentations:** These data were presented as a poster at the American Academy of Child and Adolescent Psychiatry’s 66th Annual Meeting, Chicago, IL, October 14–19, 2019.

**Funding:** This work was supported by the Eunice Kennedy Shriver National Institute of Child Health and Development (NICHD) through Grants R01HD098757 and R01HD099775 and by the Yung Family Foundation.

should be considered “failed” trials rather than “negative” trials and thus inappropriate for inclusion in reviews and meta-analyses (Walkup 2017).

Declining drug-placebo differences are consistently observed in registration trials of antidepressants in adults, but the cause remains elusive (Khin et al. 2011). Meta-analyses and systematic reviews suggest that dosing strategies, more recently completed studies, population heterogeneity, study sponsorship, and the inclusion of active comparator medications contribute to antidepressants not demonstrating efficacy (Rutherford and Roose 2013). In adults with major depressive disorder (MDD), increased placebo response rates are associated with antidepressants not demonstrating benefit (Timothy et al. 2002; Papakostas et al. 2015; Furukawa et al. 2016) and these high placebo response rates are seen in trials with more treatment arms, fixed (rather than flexible) dosing strategies, higher completion rates, longer trials, more study sites, and fewer patients enrolled from academic sites (Sinyor et al. 2010; Enck et al. 2011; Dunlop et al. 2012; Furukawa et al. 2016). Placebo response rates also increase at the end of study recruitment suggesting the pressure to complete the trial may impact rigor in implementation (McGlothlin and Viele 2018). Corresponding with the rise in placebo response rates, the pharmaceutical industry has moved from academic sites with university-based Institutional Review Boards (IRBs) to commercial clinical trials sites and centralized nonacademic IRBs. Although more efficient in trial start up and implementation, reducing academic sites may boost placebo response rates and decrease the likelihood of detecting drug-placebo differences (Sinyor et al. 2010; Enck et al. 2011; Dunlop et al. 2012; Strawn and Croarkin 2018).

Taken together, studies with high placebo response rates and small drug-placebo differences force the question as to whether these trials are truly “negative” trials or, more appropriately considered, “failed” trials due to implementation problems. It is important to note the incentive structure in pursuing a clinical indication is clearly favorable to producing a high-quality trial; however, the costs and time involved in enhancing rigor in large multisite multinational trials may work against quality implementation and result in an increasing number of failed trials.

In children and adolescents, few studies have examined trial success variables beyond the impact of placebo response (Dobson and Strawn 2016). Studies with more study sites, fewer Caucasians, fewer males, fewer patient randomizations per site, and studies led by inexperienced investigators (those lacking significant experience in the specific disorder and/or population being studied) have higher placebo response rates (Bridge et al. 2009; Cohen et al. 2010; Walkup 2017).

The incentive structure for antidepressant trials in youth sets the goal as trial *completion*, rather than *outcome*. Enacted in 1997, the Food and Drug Modernization Act (FDAMA) incentivized medication trials in youth by providing 6 months of additional patent exclusivity if a pediatric trial was undertaken, regardless of the outcome. Extended patent exclusivity (and consequent revenue) resulted in a surge of antidepressant (and other drug) trials from 1997 to 2007. Then, the Pediatric Research Equity Act of 2003 (PREA) required testing of newly approved medications in children and adolescents if the “new drug is likely to be used in children” to report safety and efficacy findings on the product label (Field and Boat 2012). However, demonstrating efficacy was still not required. Although financial incentives were initially provided to the pharmaceutical industry (extended patent exclusivity), civil penalties later followed for those who failed to conduct post-marketing studies in children and adolescents. These factors and

the demand to rapidly complete these studies implicitly shifted study motivation from “outcome” (efficacy evaluation for pediatric indications) to “completion” (for regulatory and commercial purposes).

This situation is perhaps best understood in terms of economic theory. First, FDAMA/PREA failed to create an incentive mechanism that would produce high-quality adequately powered successful trials that could demonstrate both efficacy and safety. Rather, these policies created an *adverse incentive structure* (i.e., incentives that produce the opposite effect to that intended); FDAMA and PREA incentivized low-quality underpowered hastily implemented trials in youth.

Economically, this problem can also be viewed in terms of *externalities* (i.e., economic activity that affects a third party that is not directly involved in or benefiting from the activity). Successfully implemented trials produce positive externalities in that they increase clinical knowledge and lead to superior treatments for all of society. Poorly implemented trials produce negative externalities because they reduce the clarity of accumulated evidence or present misleading information and reduce knowledge acquisition. Since only *internal benefits* (i.e., activities that are financially advantageous to corporate sponsors) are considered by the industry sponsors of clinical trials, one of the main economic roles of government is to produce/incentivize *positive externalities* and limit/disincentivize *negative externalities*. In the current climate, many of the trials conducted under FDAMA/PREA have disincentivized the positive externality associated with carefully implemented meticulously conducted trials that were positioned to demonstrate efficacy. Future policy initiatives should, therefore, carefully consider the impact a proposed policy will have on incentives, and structure the policy to obtain the desired external effects on knowledge acquisition and overall benefit to society.

Despite recognition of multiple factors that influence drug-placebo differences, evaluating these differences across studies involves formidable challenges. These include (1) accounting for both observed and unobserved heterogeneity, (2) capturing the fundamentally hierarchical structure of the studies, (3) pooling data across trials to achieve an acceptable risk of type II error, and (4) leveraging prior data to accurately assess outcomes (Quintana et al. 2017; Lewis and Angus 2018). However, these limitations in conventional meta-analytic approaches can be addressed by Bayesian hierarchical models (BHMs). BHMs integrate multilevel information that allows multiple effects to be estimated in parallel and separates the observed variability into parts attributable to “random differences and true differences” (McGlothlin and Viele 2018). This allows better assessment of heterogeneity and provides more precise estimates of the treatment effect without relying on large sample assumptions. BHMs do not require assuming complete heterogeneity, as has been common in prior meta-analyses, instead modeling the degree of heterogeneity through estimation of an individual effects distribution.

The pressure to complete trials within budget, and on time, may come with the risk of trial failure. It is critical to identify predictors of “failed” clinical trials as “the inevitable outcome is delay and increasing inefficiency in the development of new antidepressants” (Mancini et al. 2014). Moreover, the inclusion of failed trials—particularly when the result of poor trial implementation—degrade the estimated effect determined in meta-analyses, which results in confusion among clinicians, patients, families, and policy makers (Walkup 2017). With these considerations in mind, we examined factors that potentially influence clinical trial outcome in children and adolescents with anxiety and depressive disorders.

## Methods

### Search strategy

The studies included were primarily obtained through an electronic database search of English language articles in PubMed, the Cochrane Database, and the U.S. government clinical trials registry located at [www.clinicaltrials.gov](http://www.clinicaltrials.gov). The following search strategies were utilized within databases: (adolescent\* OR children OR pediatric OR youth) AND (anxiety OR generalized anxiety disorder OR GAD OR depression OR MDD) AND (selective serotonin reuptake inhibitor OR SSRI OR serotonin norepinephrine reuptake inhibitor OR SNRI OR selective serotonin norepinephrine reuptake inhibitor OR fluoxetine OR paroxetine OR sertraline OR citalopram OR escitalopram OR venlafaxine OR duloxetine OR nefazodone OR imipramine OR clomipramine OR fluvoxamine OR mirtazapine OR nortriptyline OR desipramine OR amitriptyline OR desvenlafaxine OR vilazodone). Search results were filtered for clinical trials in the child and adolescent population. Cochrane reviews, reference lists, and Supplementary Materials were hand-searched for additional relevant trials. Results were then cross-checked and limited to randomized DBPCTs of antidepressants in anxious or depressed youth (Fig. 1).

### Study inclusion criteria

Randomized DBPCTs evaluating the safety, efficacy, and tolerability of antidepressants for the treatment of youth with anxiety disorders and MDD were included for review. Consistent with recent meta-analyses in this population with MDD and anxiety disorders, studies were excluded if they included adults (age

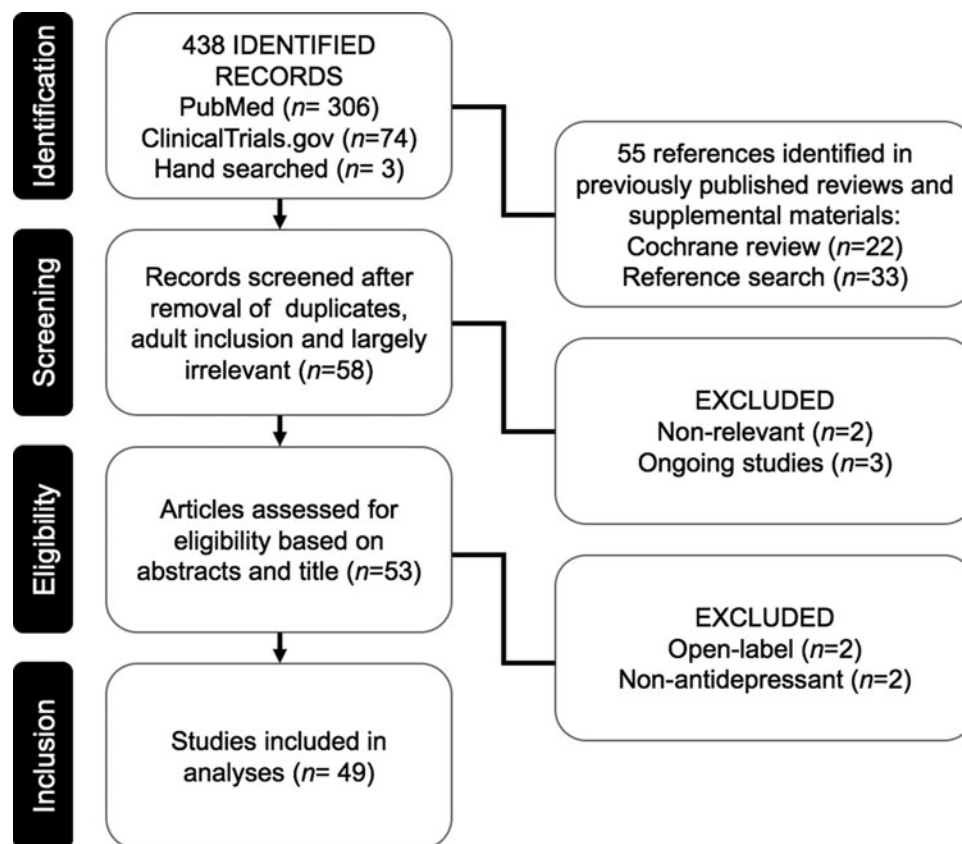
>19 years), did not study an antidepressant, were not randomized, were not placebo controlled, or provided adjunctive psychotherapy to the active or control groups being analyzed (Dobson and Strawn 2016; Varigonda et al. 2016).

### Data extraction

Extracted data were recorded into an Excel (Microsoft, Redmond, WA) spreadsheet. Based on prior research in adults, our hypotheses and considerations for the child and adolescent population, we extracted data regarding the following: (1) patient characteristics (i.e., age, gender, and patient enrollment per site), (2) study design (i.e., treatment arms, trial duration, fixed vs. flexible dosing strategies, study drug and class, and placebo lead-in periods), and (3) implementation (i.e., funding type, enrollment period, international sites, completion rates, and discontinuation rates). The efficacy of antidepressants was reported in terms of primary categorical outcome measures, placebo response rates (%), and medication response rates (%), when applicable.

### Statistical methods

Analyses were based on data from acute treatment phases; data from extension and placebo lead-in phases were excluded. As previously described, a BHM was applied to evaluate response. BHMs—which assume that the degree of heterogeneity across studies is unknown—facilitate the examination of the sensitivity of the results to varying degrees of heterogeneity, including complete heterogeneity (i.e., the studies are entirely unrelated) (Mills and Strawn 2020). The BHM was specified as a Beta-Binomial with



**FIG. 1.** Preferred reporting items for systematic reviews and meta-analyses flow diagram for selection of antidepressant trials in child and adolescent anxiety and depression.

hierarchical parameters,  $\theta$ , the mode across studies, and  $K$ , the degree of heterogeneity across studies. The hierarchical prior used herein is  $\theta \sim \text{Beta}(a_\theta, b_\theta)$ , with  $a_\theta = 3$ ,  $b_\theta = 2$  for a relatively uninformative prior,  $K \sim \text{Gamma}(a_K, b_K)$ , with  $a_K = 10$ ,  $b_K = 20$  to estimate the degree of heterogeneity. Models were generated from Hamiltonian Monte Carlo simulation, as previously described (Mills and Strawn 2020). From the BHM posterior simulation samples, we determined the probability for each trial outcome (i.e., medication-related improvement, placebo-related improvement, difference in placebo-medication improvement) and report this probability with the 95% credible interval (CrI).

Sensitivity analyses were conducted in which the posterior densities of differences in risk rates were examined for SSRIs, SNRIs, and tricyclic antidepressant (TCA) trials in patients with depressive or anxiety disorders. The extent to which the estimated unobserved heterogeneity across trials influenced our findings was examined by comparing the BHM with a model assuming complete homogeneity of studies so that patient outcomes are exchangeable across all studies and risk of incidence is assumed the same for all studies.

Statistical heterogeneity was quantified with standard measures: the  $Q$  statistic (i.e., weighted sum of squared differences between individual study effects and the pooled effect across trials),  $I^2$  (i.e., heterogeneity-related variance rather than variance attributed to sampling error), and  $\tau^2$  (i.e., variance among true effect sizes) (Higgins et al. 2003; Higgins 2008). Furthermore, the variance parameter  $K$  in the BHM can be considered a Bayesian measure of heterogeneity and reflects between study heterogeneity variance in probability of response associated with medication or placebo.

Analyses were conducted in *R* (version: 3.4.3) and *Julia* (version: 1.2.0) (Bezanson et al. 2014). The BHMs were each estimated using 5000 iterations of the “No U-Turn” Hamiltonian Monte Carlo sampler in Turing.jl (Ge et al. 2018). The posterior Markov

Chain Monte Carlo samples provide a simulation sample of the difference in response between medication and placebo. Kernel density estimates for this sample numerically approximate the exact posterior density for the difference, with a larger simulation sample leading to a higher degree of accuracy. Mean probability of response rate, standard deviations (SDs), and CrIs are then computed directly from the simulated posterior sample, along with the posterior density ratio (PDR) representing posterior odds against the null hypothesis (i.e., there is no difference) and a Bayesian posterior  $p$ -value (Kruschke 2014; McGlothlin and Viele 2018; Mills and Strawn 2019).

For all analyses, Bayesian PDRs and Bayesian equivalents of a frequentist  $p$ -value for evidence against the null hypothesis ( $H_0$ ) were determined and considered statistically significant at the  $p < 0.05$  level. Means are represented  $\pm$  their SDs and precision is expressed as 95% CrIs. Regarding multiple comparisons, a Bayesian inferential approach was selected as it eliminates the need to adjust for multiple comparisons given the sequential relationship between inference and hypothesis testing (Kruschke 2014).

## Results

### Selection of studies and study characteristics

The preferred reporting items for systematic reviews and meta-analyses diagram in Figure 1 provides a visual for clinical trial selection, which yielded 49 studies for inclusion. Overall, through the specific search strategy previously described, 438 citations were identified by the initial search and 58 potentially eligible articles were screened with 53 assessed for eligibility; 49 DBPCTs including >900 study sites and 8642 patients between the years 1971 and 2019 were included (Table 1). Of these included trials, 19 antidepressants from 5 medication classes were analyzed, with

TABLE 1. TRIAL DESIGN AND PATIENT CHARACTERISTICS

|                                 | Anxiety studies            | Depression studies          | Significance | Federally funded           | Industry funded             | Significance |
|---------------------------------|----------------------------|-----------------------------|--------------|----------------------------|-----------------------------|--------------|
| Trial category, $n$ (%)         | Federal<br>( $n=9$ ; 69%)  | Federal<br>( $n=7$ ; 21%)   | 0.001        | Anxiety<br>( $n=9$ ; 56%)  | Anxiety<br>( $n=3$ ; 11%)   | 0.001        |
|                                 | Industry<br>( $n=3$ ; 23%) | Industry<br>( $n=21$ ; 62%) | 0.010        | MDD<br>( $n=7$ ; 44%)      | MDD<br>( $n=21$ ; 81%)      | 0.008        |
|                                 | Unknown<br>( $n=1$ ; 8%)   | Unknown<br>( $n=6$ ; 18%)   | 0.260        | Unknown<br>( $n=0$ ; 0%)   | Unknown<br>( $n=2$ ; 8%)    | 0.219        |
| Dosing, $n$ (%)                 | Fixed ( $n=5$ ; 38%)       | Fixed ( $n=11$ ; 32%)       | 0.333        | Fixed ( $n=7$ ; 44%)       | Fixed ( $n=5$ ; 19%)        | 0.048        |
|                                 | Flexible<br>( $n=8$ ; 62%) | Flexible<br>( $n=20$ ; 59%) | 0.450        | Flexible<br>( $n=9$ ; 56%) | Flexible<br>( $n=17$ ; 65%) | 0.276        |
|                                 | Unknown<br>( $n=0$ ; 0%)   | Unknown<br>( $n=3$ ; 9%)    | 0.248        | Unknown<br>( $n=0$ ; 0%)   | Unknown<br>( $n=4$ ; 15%)   | 0.075        |
| Placebo lead-in period, $n$ (%) | Lead-in<br>( $n=2$ ; 15%)  | Lead-in<br>( $n=14$ ; 41%)  | 0.055        | Lead-in<br>( $n=5$ ; 31%)  | Lead-in<br>( $n=9$ ; 35%)   | 0.427        |
|                                 | (2/13)                     | (14/34)                     |              | (5/16)                     | (9/26)                      |              |
| Enrollment per site, $n$        | 31.9 $\pm$ 24.7            | 19.6 $\pm$ 19.7             | 0.355        | 39.7 $\pm$ 26.4            | 10.3 $\pm$ 5.0              | 0.147        |
| Screening size, $n$             | 215.4 $\pm$ 170.6          | 411.3 $\pm$ 239.0           | 0.259        | 251.1 $\pm$ 283.7          | 439.1 $\pm$ 166.2           | 0.292        |
| Randomized participants         | 136 $\pm$ 121              | 197 $\pm$ 143               | 0.378        | 85 $\pm$ 82                | 257 $\pm$ 121               | 0.129        |
| International, $n$ (%)          | $n=2/13$ ; 15%             | $n=9/32$ ; 28%              | 0.219        | $n=1/16$ ; 6%              | $n=9/25$ ; 36%              | 0.017        |
| Female (%)                      | 49.9% $\pm$ 12.8           | 54.4% $\pm$ 9.4             | 0.395        | 49.8% $\pm$ 14.1           | 54.7% $\pm$ 6.3             | 0.376        |
| Mean age (years)                | 11.9 $\pm$ 1.5             | 13.9 $\pm$ 1.9              | 0.213        | 12.7 $\pm$ 2.2             | 13.6 $\pm$ 1.4              | 0.370        |
| Age range (years)               | 5–17                       | 6–19                        |              |                            |                             |              |

Enrollment per site, screening size, and participants randomized are reported as means  $\pm$  standard deviation.

fluoxetine ( $\kappa=7$ ) being the most common (Supplementary Table S1). Regarding medication class, selective SSRIs and TCAs were the most widely studied, representing more than half of the included studies (Supplementary Table S2).

### Study characteristics

Federal (33%) and industry (57%)-sponsored randomized DBPCTs of anxiety (27%) and MDD (73%) in children and adolescents were included. In addition, trial dosing was fixed in 33% of trials ( $n=16$ ) and flexible in 57% ( $n=28$ ) and 33% had a placebo lead-in. Five trials included a comparison medication (fluoxetine,  $n=4$ ; imipramine,  $n=1$ ) and 11 studies (22%) were performed internationally. The mean enrollment per study was 176 patients (medication,  $n=94$ ; placebo,  $n=81$ ; comparator,  $n=89$ ) with a range of 20–529 patients. On average, 23 patients were randomized per study site. The median treatment duration was 8 weeks (interquartile range 8–10). Approximately half of the sample was female (53%) and the mean age was 13 years (drug mean age: 13; placebo mean age: 13; comparator mean age: 14). Descriptive statistics for anxiety ( $\kappa=13$ ) and depression ( $\kappa=36$ ) studies and for federally-funded and industry-funded trials are provided in Table 2.

Across trials, there was no evidence of heterogeneity for placebo response ( $Q$ 's 0.024 to 3.348, all  $p$ 's  $>0.9$ ) or medication response ( $Q$ 's 0.038 to 1.322, all  $p$ 's  $>0.9$ ; see Supplementary Table S3, available online). Similarly, other heterogeneity statistics (i.e.,  $I^2$  and  $\tau^2$ ) did not suggest heterogeneity (see Supplementary Table S3, available online).

### Medication response

Medication response (based on clinical global impression-improvement response criteria or other categorical response criteria) did not differ between industry- ( $0.607 \pm 0.018$ , CrI 0.571 to 0.642) and federally funded trials ( $0.653 \pm 0.038$ , CrI 0.575 to 0.728,  $p=0.252$ ; Table 2; Fig. 2). Furthermore, medication response did not differ between industry- and federally funded studies of youth with anxiety disorders ( $p=0.335$ ) or MDD ( $p=0.741$ ). In industry-funded trials, treatment response was associated with average patient enrollment ( $p=0.039$ ) and number of sites ( $p=0.023$ ), but not with depression ( $p=0.705$ ) or the percentage of female participants ( $p=0.812$ ). In federally funded trials, medication response rate was not associated with average patient enrollment ( $p=0.574$ ), depression ( $p=0.141$ ), percentage of girls ( $p=0.411$ ), or the number of sites ( $p=0.639$ ).

Regarding design and implementation characteristics, flexible dosing ( $p=0.099$ ), the study of an SSRI ( $p=0.110$ ), completion rate ( $p=0.341$ ), and the proportion of international sites ( $p=0.94$ ) did not predict treatment response.

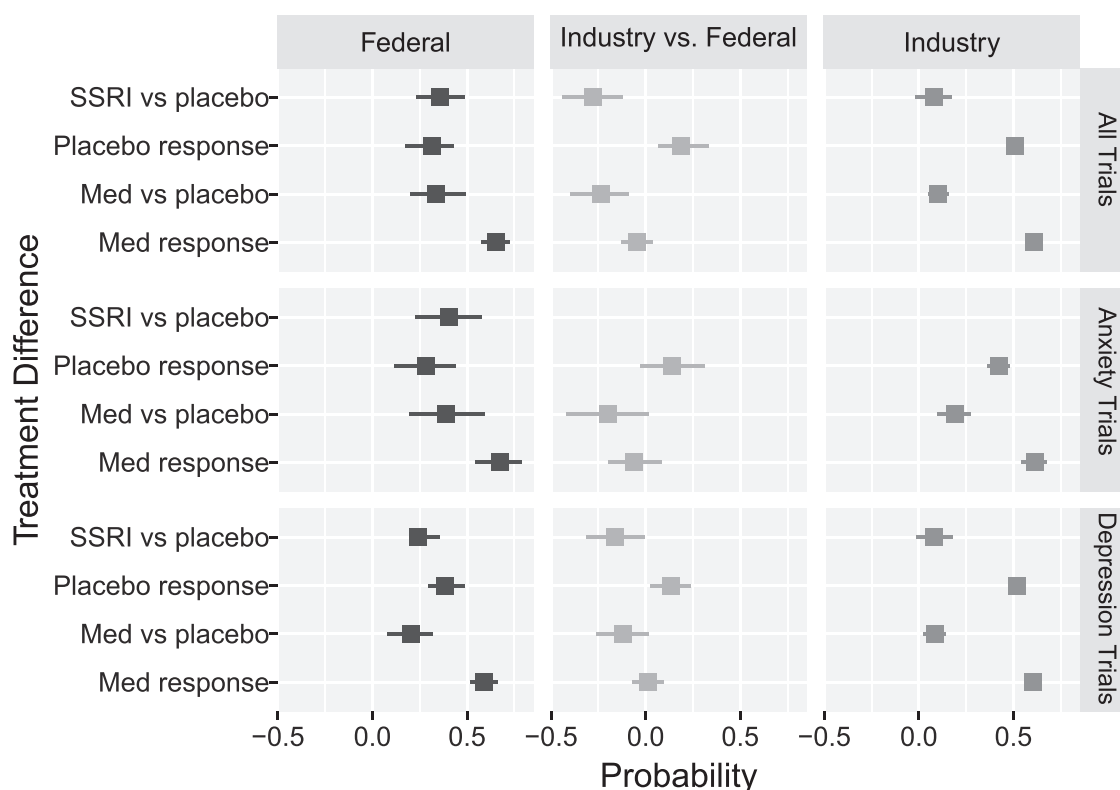
### Placebo response

Placebo response was significantly larger in industry-funded ( $0.506 \pm 0.002$ , CrI 0.462 to 0.549) compared with federally funded MDD trials ( $0.317 \pm 0.062$ , CrI 0.175 to 0.430) ( $p=0.0008$ ; Table 2; Fig. 2). There were insufficient trials to compare placebo response between industry- and federally funded trials of anxiety disorders. Across all trials, greater placebo response was associated with a higher proportion of females ( $p=0.005$ ), depression ( $p=0.017$ ), and number of sites ( $p=0.027$ ). For each 10% increase in female enrollment placebo response rate increased by  $5.47 \pm 1.77\%$ , and for every increase in the number of sites by 10 placebo response increased by  $2.74 \pm 1.17\%$ . Finally, no statistically significant

TABLE 2. MEDICATION AND PLACEBO RESPONSE IN ANTIDEPRESSANT CLINICAL TRIALS OF YOUTH WITH DEPRESSIVE AND ANXIETY DISORDERS, ALLOWING FOR HETEROGENEITY

|                        | Industry vs. federal |                    |                    | Federal           |                   |                   | Industry          |                   |                   |
|------------------------|----------------------|--------------------|--------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
|                        | Anxiety              | MDD                | Anxiety + MDD      | Anxiety           | MDD               | Anxiety + MDD     | Anxiety           | MDD               | Anxiety + MDD     |
| Medication vs. placebo |                      |                    |                    |                   |                   |                   |                   |                   |                   |
| Mean                   | -0.200 $\pm$ 0.111   | -0.121 $\pm$ 0.069 | -0.235 $\pm$ 0.078 | 0.390 $\pm$ 0.099 | 0.205 $\pm$ 0.063 | 0.336 $\pm$ 0.072 | 0.189 $\pm$ 0.046 | 0.084 $\pm$ 0.030 | 0.101 $\pm$ 0.028 |
| PDR                    | 6.93                 | 5.94               | 66.35              | 218.6             | 129.7             | 6432.6            | 461.8             | 62.97             | 362.0             |
| $p$ -Value             | 0.069                | 0.078              | 0.005              | 0.005             | 0.006             | 0.0007            | 0.003             | 0.008             | 0.0008            |
| CrI                    | -0.417 to 0.018      | -0.259 to 0.018    | -0.397 to -0.086   | 0.193 to 0.590    | 0.080 to 0.321    | 0.199 to 0.492    | 0.097 to 0.275    | 0.026 to 0.144    | 0.047 to 0.156    |
| Placebo response       |                      |                    |                    |                   |                   |                   |                   |                   |                   |
| Mean                   | 0.140 $\pm$ 0.085    | 0.136 $\pm$ 0.056  | 0.189 $\pm$ 0.066  | 0.282 $\pm$ 0.079 | 0.386 $\pm$ 0.051 | 0.317 $\pm$ 0.062 | 0.422 $\pm$ 0.030 | 0.521 $\pm$ 0.023 | 0.506 $\pm$ 0.022 |
| PDR                    | 6.21                 | 23.11              | 79.53              | 456.3             | 9757.8            | >10,000           | >10,000           | >10,000           | >10,000           |
| $p$ -Value             | 0.086                | 0.022              | 0.0008             | <0.0001           | <0.0001           | <0.0001           | <0.0001           | <0.0001           | <0.0001           |
| CrI                    | -0.028 to 0.313      | 0.026 to 0.241     | 0.067 to 0.333     | 0.113 to 0.438    | 0.292 to 0.487    | 0.175 to 0.430    | 0.364 to 0.482    | 0.477 to 0.565    | 0.462 to 0.549    |
| Medication response    |                      |                    |                    |                   |                   |                   |                   |                   |                   |
| Mean                   | -0.061 $\pm$ 0.072   | 0.014 $\pm$ 0.042  | -0.046 $\pm$ 0.042 | 0.672 $\pm$ 0.062 | 0.591 $\pm$ 0.038 | 0.653 $\pm$ 0.038 | 0.612 $\pm$ 0.036 | 0.605 $\pm$ 0.019 | 0.607 $\pm$ 0.018 |
| PDR                    | 1.84                 | 1.04               | 2.03               | >10,000           | >10,000           | >10,000           | >10,000           | >10,000           | >10,000           |
| $p$ -Value             | 0.335                | 0.741              | 0.252              | <0.0001           | <0.0001           | <0.0001           | <0.0001           | <0.0001           | <0.0001           |
| CrI                    | -0.200 to 0.083      | -0.069 to 0.096    | -0.130 to 0.038    | 0.546 to 0.791    | 0.518 to 0.666    | 0.575 to 0.728    | 0.542 to 0.678    | 0.567 to 0.643    | 0.571 to 0.642    |

CrI, credible interval; MDD, major depressive disorder; PDR, posterior density ratio.



**FIG. 2.** Probability of events in all studies, anxiety studies and depression studies as a function of funding source. SSRI, serotonin norepinephrine reuptake inhibitor.

differences were observed between average completion rates for medication-treated patients (77.3%) and those who received placebo (76.6%).

Regarding design and implementation characteristics, flexible dosing ( $p=0.434$ ), the study of an SSRI ( $p=0.337$ ), and the percentage of placebo completion ( $p=0.606$ ) did not predict placebo response. However, a greater percentage of girls enrolled in the trial was significantly associated with placebo response ( $p=0.039$ ), whereas the number of international sites ( $p=0.094$ ) was associated with placebo response at a 10% threshold.

In industry-funded trials, placebo response was predicted by depression and trended toward being associated with the number of sites (2.107,  $p=0.059$ , 95% CrI  $-0.0148$  to  $0.676$ ) but was not associated with average patient enrollment ( $p=0.279$ ) or the percentage of female participants ( $p=0.211$ ). In federally funded studies, average patient enrollment ( $p=0.450$ ), depression ( $p=0.235$ ), percentage of female participants ( $p=0.078$ ), and number of sites ( $p=0.331$ ) were not associated with placebo response. There is high correlation between the funding indicator variable and the number of study sites ( $r=0.77$ ), so both cannot be included in the same model due to multicollinearity.

#### Medication-placebo differences

Differences in medication and placebo response were significantly larger in federally funded ( $0.336 \pm 0.072$ , CrI  $0.199$  to  $0.492$ ) compared with industry-funded trials ( $0.101 \pm 0.028$ , CrI  $0.047$  to  $0.156$ ) ( $p=0.005$ ; Table 2; Fig. 2). There were an insufficient number of anxiety SSRI trials to compare industry versus federal funding; however, for depression trials involving SSRIs, medication-placebo differences were smaller in industry-funded ( $0.079 \pm$

$0.036$ , CrI  $-0.015$  to  $0.177$ ) compared with federally funded trials ( $0.242 \pm 0.062$ , CrI:  $0.227$  to  $0.359$ ) ( $p=0.042$ ; Supplementary Table S4).

#### Discussion

This meta-analysis is the largest examination of medication and placebo response in children and adolescents with anxiety and depressive disorders. Whether these trials demonstrate the value of antidepressant medication as compared with placebo is primarily related to the funding source; drug response appears relatively consistent, whereas industry funding drives placebo response upward. This analysis replicates and extends the findings of prior analyses that find that trial outcomes depends on placebo response rates—rather than medication response rates and that whether a trial demonstrates efficacy is dependent on funding source (Bridge et al. 2009; Locher et al. 2017). These findings call into question policies, practice guidelines, and meta-analyses that include the large number of industry-sponsored antidepressant trials when examining the efficacy of antidepressant medications in children and adolescents. In addition, these results are directly relevant to regulatory agencies and indicate that clinical trial design and implementation ultimately determine whether interventions effective for adults with anxiety and depression can be identified as effective in industry trials in children and adolescents.

Across disorders, federally funded trials have nearly a 25% greater chance of demonstrating efficacy compared with industry-funded trials ( $p=0.005$ ). This is driven by an almost 20% greater chance of elevated placebo response in industry-funded trials ( $p<0.001$ ), whereas medication response is nearly identical ( $p=0.252$ ). Differences in implementation may account for the pattern of failed industry-sponsored depression studies, including

trials conducted for regulatory purposes with large number of sites, few patients per site and prepubertal depressed patients, and pressure to complete trials under the FDAMA incentive before their respective patents expired (US FDA 2002; Bridge et al. 2007).

Given that the response rates of medication are similar across randomized controlled trials and placebo response varies based on sponsorship, it is possible that the “superiority” of an intervention reflects differences in study sponsorship, rather than the medication itself (Emslie et al. 1997). As an example, fluoxetine was the first SSRI to be marketed in the United States and the only SSRI ever studied in federally funded pediatric major depression studies (Emslie et al. 1997). Under blinded conditions, the two federally funded studies were able to keep the placebo response rate low in a range considered ideal (33%–35%), substantially contributing to finding fluoxetine effective (Iovieno and Papakostas 2012). Given that the other SSRIs were only studied in industry-sponsored depression trials with high placebo response rates and small drug placebo differences, we cannot draw conclusions about their relative efficacy in teen depression.

Thus, the frequent finding of fluoxetine’s “superiority” in many meta-analyses of antidepressants in youth may be related to rigor required by federal funding (Cipriani et al. 2016; Strawn and Walkup 2020). We would suggest that although the SSRIs may differ in structure, they have similar mechanism of action and in properly selected patients, they should also be effective. We would posit that it is not accurate to build policy or practice guidelines or even suggest that fluoxetine is the only effective antidepressant for adolescent depression. Thus, clinicians considering an antidepressant for a depressed or anxious child or adolescent should *not* presume that fluoxetine, as identified in meta-analyses, is singularly effective. Rather, clinicians should consider SSRIs as a class and match key differences among antidepressants such as tolerability and pharmacokinetics to their patient’s specific needs.

The incentive structure that subtends the findings of this meta-analysis warrants additional discussion. Initially, the pharmaceutical industry received incentives and later civil penalties for failing to conduct postmarketing studies in children and adolescents. This resulted in a *de facto* shift in the study goal away from evaluating efficacy, as in federally funded trials, to satisfying a regulatory requirement. In essence, the study outcome becomes moot. For example, to satisfy FDAMA requirements, two studies of serotonin modulators in children and adolescents with GAD were conducted for a mere 6 months and involved nearly 600 patients who were described as without any “comorbidity” (Strawn et al. 2017). In contrast, federally funded trials that demonstrated the efficacy of SSRIs in pediatric anxiety disorders enrolled a similar sample size ( $N=488$ ), required >5 years to be completed and included the expected, high rates of anxiety comorbidity (Walkup et al. 2008). The differences between these studies are striking not just in terms of funding, but in terms of placebo response rates (~50% vs. 24%), number of sites (82 vs. 6 sites), and investigator expertise. As illustrated by this juxtaposition, to satisfy requirements while benefiting from patent extensions, the industry sponsor of the serotonin modulator needed to identify large groups of investigators who could quickly recruit, engage, treat, and retain a large number of participants (Walkup 2017). Not unexpectedly, the assembly of such a study team often required study investigators who were not child and adolescent psychiatrists (note: during this era, there were very few child and adolescent psychiatrists specializing in anxiety, who were capable of running such a study), but rather an assemblage of unidentified clinicians from a variety of backgrounds that included adult and child psychiatrists from academia, private practice, as well as for-profit research-dedicated clinics (Walkup 2017).

FDAMA and PREA have, at best, guaranteed that large clinical trials will be conducted in children and adolescents. However, there remains no impetus for the pharmaceutical industry to conduct high-quality trials demonstrating efficacy in children and adolescents. More carefully developed and implemented federal policies could create the incentive structure to conduct trials that are poised to demonstrate efficacy and safety. These policies and the regulatory agencies that implement them could (1) improve clinical trials methodology, (2) incentivize clinician-scientists with the necessary expertise to conduct these trials (Walkup 2017), (3) leverage contemporary statistical approaches (Suresh et al. 2020), or even (4) include synthetic placebo data from high-quality prior trials to control the burgeoning placebo response rates.

Even though this is the largest meta-analysis of the factors that contribute to antidepressant and placebo response in children and adolescents with depressive and anxiety disorders, there are several important limitations that warrant additional discussion. First, these analyses rely on the beta-binomial distributional assumptions of the model for placebo and medication response rates. This assumption is standard and can be justified as a reasonable approximation to the data generating process resulting from a randomizing controlled clinical trial. The availability of the raw data would allow for testing of this assumption and evaluation of continuous outcome measures. Second, despite the general similarity of studies and our use of BHM to address the influence of exchangeability assumptions, unobserved factors may still affect the likelihood of medication or placebo response. Third, trial duration varied and could influence the probability of placebo or medication response.

What then is the value of an industry-sponsored study? Clearly, even “failed” trials can provide data on the potentially unique safety profile of antidepressants in children and adolescents. That said, if implementation affects study efficacy outcomes, one might have to consider safety signals in “failed trials” to also be suspect. Working to develop a strong methodology for adverse event monitoring could make even “failed trials” in children and adolescents more informative. Adjusting expectations for industry-sponsored trials would also be beneficial to the field. If one presumed that, even with the “gold standard” approach of a double-blind placebo-controlled RCT, that implementation challenges might occur and lead studies to be failed, one could interpret the literature more accurately. It is not uncommon for registration trials to have mixed outcomes—some trials demonstrate efficacy and some do not; that inconsistency could be properly interpreted if there was greater transparency regarding why some studies succeed and some fail, especially if the monitoring of implementation was more focused on the quality of the clinical assessment. There is great pressure on clinical trialists to recruit, treat, and retain. One method of sample maintenance is a positive attitude of the trialist and substantial support and encouragement for the participant—all of which would enhance placebo response especially in depression trials. It would also be valuable to collect data on the clinical trialists involved in industry-sponsored studies to identify those trialists capable, under blinded conditions, to identify drug from placebo response. Such a registry of trials completed, and site outcome could encourage trialists to develop true expertise in clinical assessment. In addition, if high placebo response rates due to implementation problems were to be expected, it would also be easier to see the value of small drug-placebo differences as reflecting a potential signal of efficacy but not necessarily a reflection of the public health value of the medication. In short, one could consider the medication to have some value based on trial outcome, but that outcome should not be considered when evaluating how that

medication would perform in the hands of a well-trained clinician and a properly selected patient or be included in meta-analyses or drive policy or treatment guidelines.

## Summary

The differences in placebo response and the subsequent variability in medication-placebo differences for industry- and federally funded trials underscore the scientific and clinical factors that may have coalesced in the post-FDAMA regulatory environment and resulted in a number of “failed trials” in children and adolescents (Walkup 2017). This raises the question of whether these trials should be included in reviews or meta-analyses that seek to determine the public health value of antidepressants in the treatment of children and adolescents with anxiety and depression. Ultimately, identifying the proper role and value of industry-funded trials completed for regulatory purposes is critical to establishing the true evidence base for antidepressants in child and adolescent anxiety and depression. Finally, these results remind clinicians of the importance of context in evaluating the evidence base for interventions in children and adolescents (Strawn and Walkup 2020).

## Disclosures

J.R.S. has received research support from the National Institutes of Health as well as Abbvie, Shire, Lundbeck, Otsuka, and Neuronetics. He has received material support from Genesight/Assurex Health; has received royalties from the publication of two texts (Springer); and has served as an author for *UpToDate* and an associate editor for *Current Psychiatry*. He has received honoraria from Genomind, CMEology, Neuroscience Education Institute and consulted to Intracellular. J.T.W. has served on the advisory board of the Tourette Syndrome Association, Trichotillomania Learning Center, Anxiety and Depression Association of America, and has received royalties from Oxford University Press, Guilford Press, and Wolters Kluwer. J.A.M. and S.A.M. report no biomedical financial interests or conflicts of interest.

## Supplementary Material

Supplementary Table S1  
Supplementary Table S2  
Supplementary Table S3  
Supplementary Table S4

## References

- Bezanson J, Edelman A, Karpinski S, Shah VB: Julia: A fresh approach to numerical computing. *SIAM Rev* 59:65–98, 2014.
- Bothwell LE, Podolsky SH: The emergence of the randomized, controlled trial. *N Engl J Med* 375:501–504, 2016.
- Bridge JA, Birmaher B, Iyengar S, Barbe RP, Brent DA: Placebo response in randomized controlled trials of antidepressants for pediatric major depressive disorder. *Am J Psychiatry* 166:42–49, 2009.
- Bridge JA, Iyengar S, Salary CB, Barbe RP, Birmaher B, Pincus HA, Ren L, Brent DA: Clinical response and risk for reported suicidal ideation and suicide attempts in pediatric antidepressant treatment. *JAMA* 297:1683–1696, 2007.
- Cipriani A, Zhou X, Del Giovane C, Hetrick SE, Qin B, Whittington C, Coghill D, Zhang Y, Hazell P, Leucht S, Cuijpers P, Pu J, Cohen D, Ravindran AV, Liu Y, Michael KD, Yang L, Liu L, Xie P: Comparative efficacy and tolerability of antidepressants for major depressive disorder in children and adolescents: A network meta-analysis. *Lancet* 388:881–890, 2016.
- Cohen D, Consoli A, Bodeau N, Purper-Ouakil D, Deniau E, Guile JM, Donnelly C: Predictors of placebo response in randomized controlled trials of psychotropic drugs for children and adolescents with internalizing disorders. *J Child Adolesc Psychopharmacol* 20:39–47, 2010.
- Dobson ET, Strawn JR: Placebo response in pediatric anxiety disorders: Implications for clinical trial design and interpretation. *J Child Adolesc Psychopharmacol* 26:686–693, 2016.
- Dunlop BW, Thase ME, Wun CC, Fayyad R, Guico-Pabia CJ, Musgnung J, Ninan PT: A meta-analysis of factors impacting detection of antidepressant efficacy in clinical trials: The importance of academic sites. *Neuropsychopharmacology* 37:2830–2836, 2012.
- Enck P, Klosterhalfen S, Weimer K, Horing B, Zipfel S: The placebo response in clinical trials: More questions than answers. *Philos Trans R Soc B Biol Sci* 366:1889–1895, 2011.
- Emslie GJ, Rush J, Weinberg WA, Kowatch RA, Hughes CW, Carmody T, Rintelmann J: A double-blind, randomized, placebo-controlled trial of fluoxetine in children and adolescents with depression. *Arch Gen Psychiatry* 54:1031–1037, 1997.
- Field MJ, Boat TF: Safe and Effective Medicines for Children: Pediatric Studies Conducted under the Best Pharmaceuticals for Children Act and the Pediatric Research Equity Act. Washington (DC): National Academies Press, 2012.
- Furukawa TA, Cipriani A, Atkinson LZ, Leucht S, Ogawa Y, Takeshima N, Hayasaka Y, Chaimani A, Salanti G: Placebo response rates in antidepressant trials: A systematic review of published and unpublished double-blind randomised controlled studies. *Lancet Psychiatry* 3:1059–1066, 2016.
- Ge H, Xu K, Ghahramani Z: Turing: A language for flexible probabilistic inference. *Proc Twenty-First Int Conf Artif Intell Stat* 84: 1682–1690, 2018.
- Higgins JPT: Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. *Int J Epidemiol* 37:1158–1160, 2008.
- Higgins JPT, Thompson SG, Deeks JJ, Altman DG: Measuring inconsistency in meta-analyses. *BMJ* 327:557–560, 2003.
- Iovieno N, Papakostas GI: Correlation between different levels of placebo response rate and clinical trial outcome in major depressive disorder: A meta-analysis. *J Clin Psychiatry* 73:1300–1306, 2012.
- Khin NA, Chen YF, Yang Y, Yang P, Laughren TP: Exploratory analyses of efficacy data from major depressive disorder trials submitted to the US food and drug administration in support of new drug applications. *J Clin Psychiatry* 72:464–472, 2011.
- Kruschke JK: *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*, Second Edition. San Diego, CA: Elsevier, 2014.
- Lewis RJ, Angus DC: Time for Clinicians to Embrace Their Inner Bayesian? *J Am Med Assoc* 320:2208–2210, 2018.
- Locher C, Koechlin H, Zion SR, Werner C, Pine DS, Kirsch I, Kessler RC, Kossowsky J: Efficacy and safety of selective serotonin reuptake inhibitors, serotonin-norepinephrine reuptake inhibitors, and placebo in common psychiatric disorders: a meta-analysis in children and adolescents. *JAMA Psychiatry* 74: 1011–1020, 2017.
- Mancini M, Wade AG, Perugi G, Lenox-Smith A, Schacht A: Impact of patient selection and study characteristics on signal detection in placebo-controlled trials with antidepressants. *J Psychiatr Res* 51: 21–29, 2014.

- McGlothlin AE, Viele K: Bayesian hierarchical models. *J Am Med Assoc* 320:2365–2366, 2018.
- Mills JA, Strawn JR: Antidepressant tolerability in pediatric anxiety and obsessive-compulsive disorders: A Bayesian hierarchical modeling meta-analysis. *J Am Acad Child Adolesc Psychiatry* 59: 1240–1251, 2020.
- Papakostas GI, Østergaard SD, Iovieno N: The nature of placebo response in clinical studies of major depressive disorder. *J Clin Psychiatry* 76:456–466, 2015.
- Quintana M, Viele K, Lewis RJ: Bayesian analysis: Using prior information to interpret the results of clinical trials. *J Am Med Assoc* 318:1605–1606, 2017.
- Rutherford BR, Roose SP: A model of placebo response in antidepressant clinical trials. *Am J Psychiatry* 170:723–733, 2013.
- Sinyor M, Levitt AJ, Cheung AH, Schaffer A, Kiss A, Dowlati Y, Lanctot KL: Does inclusion of a placebo arm influence response to active antidepressant treatment in randomized controlled trials? Results from pooled and meta-analyses. *J Clin Psychiatry* 71:270–279, 2010.
- Strawn JR, Croarkin PE: Desvenlafaxine versus placebo in a fluoxetine-referenced study of children and adolescents with major depressive disorder: Design, definitions, and ongoing challenges for child and adolescent psychopharmacology research. *J Child Adolesc Psychopharmacol* 28:363, 2018.
- Strawn JR, Mills JA, Cornwall GJ, Mossman SA, Varney ST, Keeshin BR, Croarkin PE: Buspirone in children and adolescents with anxiety: A review and Bayesian analysis of abandoned randomized controlled trials. *J Child Adolesc Psychopharmacol* 28:2–9, 2017.
- Strawn JR, Walkup JT: Identifying the best treatment for young people with depression. *Lancet Psychiatry* 7:562–563, 2020.
- Suresh V, Mills JA, Croarkin PE, Strawn JR: What next? A Bayesian hierarchical modeling re-examination of treatments for adolescents with selective serotonin reuptake inhibitor-resistant depression. *Depress Anxiety* 37:926–934, 2020.
- Timothy Walsh B, Seidman SN, Sysko R, Gould M: Placebo response in studies of major depression: Variable, substantial, and growing. *J Am Med Assoc* 287:1840–1847, 2002.
- US Food and Drug Administration: Review and Evaluation of Clinical Data (NDA 20-152; Supplement S-032; Sponsor: Bristol-Myers Squibb; Drug: Nefazodone Hydrochloride [Serzone]; Material Submitted: Pediatric Exclusivity Supplement. 2002. [https://www.accessdata.fda.gov/drugsatfda\\_docs/pediatric/020152s032\\_nefazodone\\_Serzone\\_Clinical\\_BPCA.pdf](https://www.accessdata.fda.gov/drugsatfda_docs/pediatric/020152s032_nefazodone_Serzone_Clinical_BPCA.pdf) (accessed March 23, 2021).
- Varigonda AL, Jakubovski E, Bloch MH: Systematic review and meta-analysis: Early treatment responses of selective serotonin reuptake inhibitors and clomipramine in pediatric obsessive-compulsive disorder. *J Am Acad Child Adolesc Psychiatry* 55: 851.e2–859.e2, 2016.
- Walkup JT: Antidepressant efficacy for depression in children and adolescents: Industry- and NIMH-funded studies. *Am J Psychiatry* 174:430–437, 2017.
- Walkup JT, Albano AM, Piacentini J, Birmaher B, Compton SN, Sherrill JT, Ginsburg GS, Rynn MA, McCracken J, Waslick B, Iyengar S, March JS, Kendall PC: Cognitive behavioral therapy, sertraline, or a combination in childhood anxiety. *N Engl J Med* 359:2753–2766, 2008.

Address correspondence to:

*Jeffrey R. Strawn, MD*

*The Anxiety Disorders Research Program*

*Department of Psychiatry and Behavioral Neuroscience*

*University of Cincinnati College of Medicine*

*260 Stetson Street*

*Cincinnati, OH 45219*

*USA*

*E-mail: strawnjr@ucmail.uc.edu*