

RESEARCH

Open Access



# Placebo response in pharmacological and dietary supplement trials of autism spectrum disorder (ASD): systematic review and meta-regression analysis

Spyridon Siafis<sup>1\*</sup>, Oğulcan Çıray<sup>2</sup>, Johannes Schneider-Thoma<sup>1</sup>, Irene Bighelli<sup>1</sup>, Marc Krause<sup>1</sup>, Alessandro Rodolico<sup>3</sup>, Anna Ceraso<sup>4</sup>, Giacomo Deste<sup>4</sup>, Maximilian Huhn<sup>1,5</sup>, David Fraguas<sup>6,7</sup>, Dimitris Mavridis<sup>8,9</sup>, Tony Charman<sup>10</sup>, Declan G. Murphy<sup>11</sup>, Mara Parellada<sup>8,9</sup>, Celso Arango<sup>8,9</sup> and Stefan Leucht<sup>1</sup>

## Abstract

**Background:** Placebo response in autism spectrum disorder (ASD) might dilute drug-placebo differences and hinder drug development. Therefore, this meta-analysis investigated placebo response in core symptoms.

**Methods:** We searched [ClinicalTrials.gov](https://www.clinicaltrials.gov/), CENTRAL, EMBASE, MEDLINE, PsycINFO, WHO-ICTRP (up to July 8, 2018), and PubMed (up to July 4, 2019) for randomized pharmacological and dietary supplement placebo-controlled trials (RCTs) with a minimum of seven days of treatment. Single-group meta-analyses were conducted using a random-effects model. Standardized mean changes (SMC) of core symptoms in placebo arms were the primary outcomes and placebo positive response rates were a secondary outcome. Predictors of placebo response were investigated with meta-regression analyses. The protocol was registered with PROSPERO ID [CRD42019125317](https://www.crd42019125317).

**Results:** Eighty-six RCTs with 2360 participants on placebo were included in our analysis (87% in children/adolescents). The majority of trials were small, single-center with a duration of 8–12 weeks and published after 2009. Placebo response in social-communication difficulties was  $SMC = -0.32$ , 95% CI [-0.39, -0.25], in repetitive behaviors  $-0.23$  [-0.32, -0.15] and in scales measuring overall core symptoms  $-0.36$  [-0.46, -0.26]. Overall, 19%, 95% CI [16–22%] of participants were at least much improved with placebo. Caregiver (vs. clinician) ratings, lower risk of bias, flexible-dosing, larger sample sizes and number of sites, less recent publication year, baseline levels of irritability, and the use of a threshold of core symptoms at inclusion were associated with larger placebo response in at least a core symptom domain.

**Limitations:** About 40% of the trials had an apparent focus on core symptoms. Investigation of the differential impact of predictors on placebo and drug response was impeded by the use of diverse experimental interventions with essentially different mechanisms of action. An individual-participant-data meta-analysis could allow for a more fine-grained analysis and provide more informative answers.

(Continued on next page)

\* Correspondence: [spyridon.siafis@tum.de](mailto:spyridon.siafis@tum.de)

<sup>1</sup>Department of Psychiatry and Psychotherapy, School of Medicine, Technical University of Munich, Ismaningerstr. 22, 81675 Munich, Germany  
Full list of author information is available at the end of the article



© The Author(s). 2020, corrected publication 2020. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

**Conclusions:** Placebo response in ASD was substantial and predicted by design- and participant-related factors, which could inform the design of future trials in order to improve the detection of efficacy in core symptoms. Potential solutions could be the minimization and careful selection of study sites as well as rigorous participant enrollment and the use of measurements of change not solely dependent on caregivers.

**Keywords:** Autism spectrum disorder, Placebo, Trials

## Background

Autism spectrum disorder (ASD) is a group of heterogeneous neurodevelopmental conditions, characterized by social-communication difficulties as well as repetitive-restricted behaviors and sensory abnormalities [1]. The prevalence is about 1–2% [2, 3], and lifetime costs are substantial (at US \$1.4–2.44 million per individual) [4]. Behavioral interventions are the cornerstone of treatment and there is still no approved medication for the core symptoms [5]. Despite that, about half of the individuals with ASD, who might be more susceptible to side effects than neurotypical populations [5], receive psychotropic drugs [6]. Currently approved medications target associated symptoms, e.g., aripiprazole and risperidone for irritability [5]. Therefore, there is an unmet need to develop effective and safe treatments that target causal pathophysiological pathways, improve core symptoms and quality of life.

In spite of the recent advances in “translational” research, late-stage clinical trials for neurodevelopmental disorders have failed [7]. The low success rate could be explained by several factors, such as poor translational validity of preclinical models, true lack of drug efficacy, and suboptimal trial design [8]. One concern is also that placebo effects might dilute effect sizes. However, the magnitude and predictors of placebo response in core symptoms of ASD are still unknown; only investigated in post-hoc analyses of single trials [9, 10] and meta-analyses using aggregated outcome measures, potentially confounded by associated symptoms [11, 12]. In summary, placebo response may play an important role in the failure of clinical trials and the subsequent lack of approved medications for core symptoms. In order to improve the design and sensitivity of future trials, we meta-analyzed placebo response of core symptoms in pharmacological and dietary supplement ASD trials.

## Methods

This systematic review and meta-regression analysis was conducted according to PRISMA [13] (Additional file 3: eAppendix-1) with PROSPERO registration ID CRD42019125317 (Additional file 3: eAppendix-2).

### Participants and interventions

#### Participants

Participants with a diagnosis of ASD using standardized diagnostic criteria (e.g., DSM-III, ICD-10, or more recent

versions) and/or validated diagnostic tools (e.g., ADI-R) [5]. There were no restrictions in terms of age, sex, ethnicity, setting, severity, or the presence of co-occurring conditions.

#### Interventions

Any pharmacological treatment or dietary supplement was eligible, when compared with placebo. We excluded psychological/behavioral and combination interventions (since placebo response might be confounded by the active component of the combination) as well as other interventions (e.g., elimination diets, milk formulations, or homeopathy). The minimum duration of treatment was 7 days, since we aimed to investigate a broad range of data but to exclude trials with a clearly very short duration, e.g., single-dose interventions. There was no restriction in terms of route of administration and dosing-schedule.

#### Type of studies

Blinded and unblinded randomized placebo-controlled trials (RCTs) were eligible. In case of cross-over studies, we used only data from the first phase of the crossover to avoid carry-over effects [14]. We excluded studies with placebo-controlled discontinuation or cluster randomization [15], published before 1980 or smaller than ten participants [16]. Risk of bias of included studies was evaluated by at least two independent reviewers (SS, OC, AR) using the Cochrane Collaboration risk-of-bias tool [17]. Disagreements were resolved by discussion, and if needed, a third author was involved (SL, JST). Studies with a high risk of bias in sequence generation or allocation concealment were excluded (e.g., allocation by alternation or by an unblinded investigator). Studies were further classified as having an overall low, moderate, or high risk of bias [18].

#### Search strategy and selection criteria

We searched (July 8, 2018) [ClinicalTrials.gov](http://ClinicalTrials.gov), CENTRAL, EMBASE, MEDLINE, PsycINFO, PubMed (update on July 4, 2019), and [WHO ICTRP](http://WHO ICTRP). There was no date/time, language, document type, and publication status limitations (Additional file 3: eAppendix-3). Reference lists of included studies and previous reviews [5, 11, 12, 19–27] were also inspected.

### Outcome measures and data extraction

We investigated placebo response in core symptoms. The following primary outcomes, as measured by published scales, were analyzed: (1) social-communication difficulties (e.g., ABC-L/SW [28] or VABS-Socialization [29]), (2) repetitive-restricted behaviors (e.g., ABC-S [28] or CYBOC-PDD [30]), and (3) overall measures of core symptoms (e.g., SRS [31] or CARS [32]). There is no agreement on the optimal outcome measures to use in clinical trials of ASD and so preference was given to the aforementioned most frequently used scales (Additional file 3: eAppendix-5.3) [5, 33–36]. A higher score indicated more difficulties and when necessary, scores were minus-transformed. In the primary analysis, we pooled all studies by preferring ratings by clinicians (observations or interviews) to caregivers/teachers. Separate analyses by type of raters and positive response to treatment defined as at least much improvement in CGI-I, preferably anchored to global autism or core symptoms (when more than one CGI-I evaluations were reported), were analyzed as secondary outcomes. When the number of participants with a positive response was not reported, it was imputed from mean and standard deviation (SD) of CGI-I using a validated method (Additional file 3: eAppendix-2.2) [37, 38].

At least two independent reviewers/contributors selected relevant records and extracted data from eligible studies in an Access database (SS, OC, IB, AR, AC, GD, MK, YZ, and TF). Intention-to-treat data were preferred when available, and for a positive response to treatment, if the original authors presented only the results of completer population, we assumed that participants lost to follow-up did not have a positive response to treatment. Missing SDs were calculated according to the following hierarchy from available statistics (e.g., SE,  $p$  values,  $t$  tests) [39], median/range [40], pooling subscales (e.g., SRS subscales, assuming a correlation of 0.5) [41], or using a validated imputation method [39, 42]. Corresponding authors were contacted by e-mail for additional data, with a reminder e-mail in case of no response (complete list in Additional file 3: eAppendix-4).

### Statistical analysis

#### Synthesis of the results

Single-group meta-analyses of placebo arms were conducted using a random effects model [43]. The effect size for continuous outcomes (core symptoms) was the standardized mean change (SMC) with raw score standardization using the baseline SD of the placebo arm [44, 45]. When baseline SDs were not reported, change or follow-up SDs were used. In the primary analysis, a common pre-post correlation of 0.5 [41] was used for the calculation of variance of SMC [44]. Positive

response rates were logit transformed, and back-transformed for presentation [46]. Heterogeneity was evaluated by visual inspection of forest plots and with the  $\chi^2$  ( $p$  value < 0.1) and  $I^2$  statistics (considerable heterogeneity when > 50%);  $\chi^2$  might detect small amounts of clinically unimportant heterogeneity; therefore, we based our evaluation on  $I^2$  [17].

#### Sensitivity analyses and publication bias

Predefined sensitivity analyses of the primary outcomes were conducted using a fixed-effects model or by exclusion of studies with genetic syndrome as inclusion criteria, using only diagnostic tools, single-blind, shorter than 4 weeks, presenting only completers data, with at least moderate overall risk of bias, with estimated SD (imputed, from medians/range, or pooled subscales). Post-hoc, we excluded studies without baseline SDs and we used the correlations of 0.25/0.75 for the calculation of variance of SMC [41]. Regarding responder rates, we post-hoc excluded studies with imputed responder rates [38]. We explored small study effects as proxy for publication bias with contour-enhanced funnel plots, Egger's test [47], and trim-and-fill [48].

#### Meta-regression analyses

The dependent variable was SMC and the independent variable was selected from a list of covariates from the literature [9, 11, 12, 49–51]. First, we conducted univariable and then multivariable meta-regressions similar to our previous analyses in schizophrenia [51]: we used the factors that were significant in the univariable analysis and then a formal backward stepwise algorithm with a removal criterion of  $p = 0.15$ . Meta-regressions were not performed for categorical covariates with less than five data points per level. Spearman's  $\rho$  were calculated post-hoc between SMCs of placebo and experimental intervention as well as between covariates.

**Intervention-related factors** Intervention-related factors were route of administration (oral versus others) [52], type of experimental intervention (pharmacological versus dietary supplement), and dose-schedule (fixed versus flexible).

**Study-related factors** Study-related factors were duration of treatment (weeks), publication year, washout from psychotropic medications (coded post-hoc as the presence of washout or not, because definitions varied), placebo lead-in with exclusion of those showing a positive response, type of rater (clinicians versus caregivers), total sample size, number of sites, %academic sites, number of arms and medications, %participants on placebo, sponsorship (industry-funded/patent application

versus industry-independent), country of origin (US versus not only US), and risk of bias domains.

**Participant-related factors** Participant-related factors were the presence of any associated conditions by inclusion criteria (i.e., irritability, ADHD, and other conditions apart from intellectual disability or genetic syndrome), mean age and age group (children/adolescent versus adults/mixed, post-hoc), %participants with intellectual disability (at least mild or  $IQ < 70$ ), %female (post-hoc), ethnicity (%Caucasian/Hispanic, post-hoc), and baseline BMI (post-hoc) [9]. Due to inconsistent reporting of baseline severity [11, 12], we used CGI-Severity (ranging 1–7) as a measure of global severity and ABC-Irritability (ranging 0–45) as a measure of serious behavioral problems [53]. Baseline severity in core symptoms could not be investigated as a potential predictor due to the large diversity of scales and standardization methods (such as using the lower and upper limits of the measurement scale [54]) could not be utilized (trials reported raw and standard scores such as VABS or T-scores of SRS). We also examined the use of a threshold of core symptom severity for inclusion (not only for the confirmation of diagnosis).

Analyses were conducted using metafor (v2.1-0) [45] and meta (v4.9-9) [55] in R (v3.6.2) [56]. Statistical threshold was set at two-sided alpha 5%. Due to the limited statistical power and exploratory (observational)

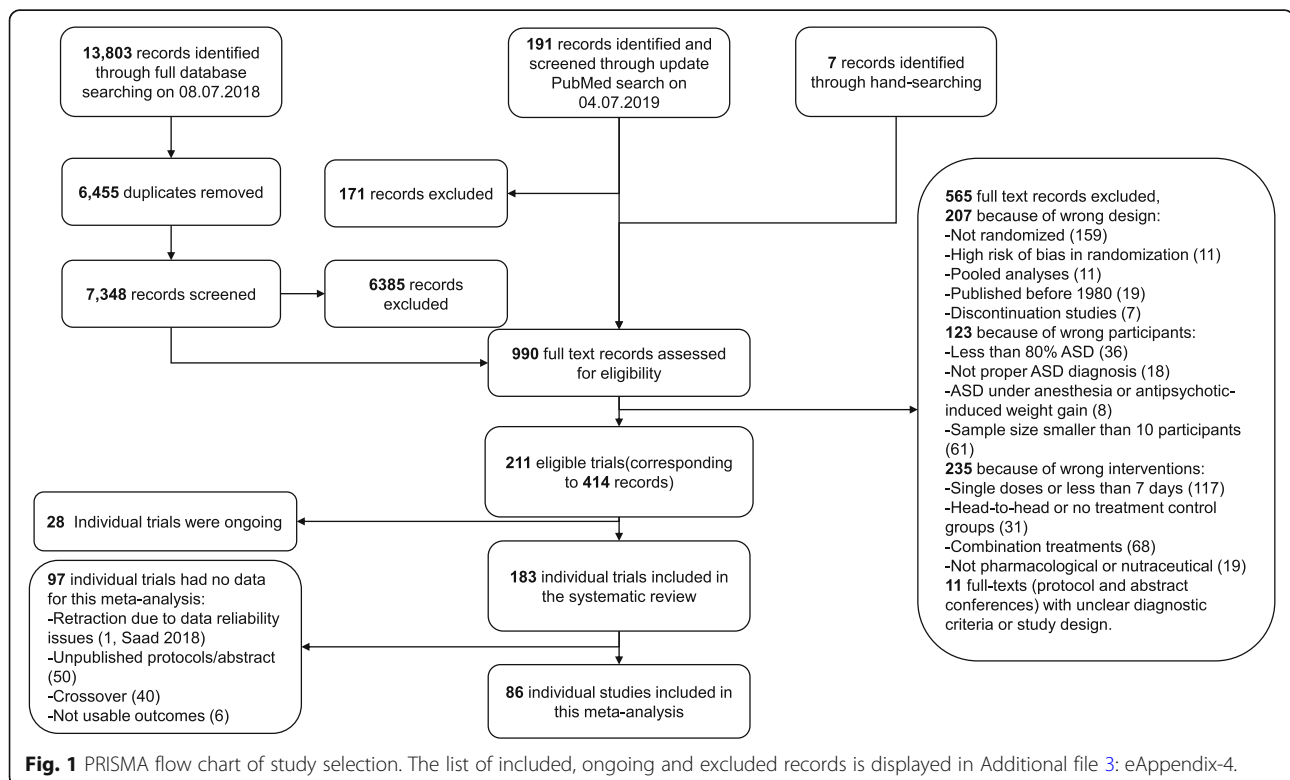
nature of meta-regression analyses, alpha was not adjusted for multiple testing. Correction for multiple testing is not generally recommended by the Cochrane Handbook [39].

## Results

### Description of included studies

The PRISMA flow diagram is presented in Fig. 1. In this analysis, 86 ( $k$ ) studies were included, 71% comparing pharmacological treatments and 29% dietary supplements with placebo (eAppendix-5.1 and Table-S). Of the 86 studies, 75 were conducted in children/adolescents, eight in adults, and three included both age groups. The overall sample size ( $n$ ) was 5365, 44% on placebo. The majority of studies were parallel (85%), single-center (60%, indicated in  $k = 78$ ), and double-blind (only one single-blind [57] and none was open) with two arms (88%) and small sample sizes (median 45, interquartile-range [30–91]). About half of the studies (48%) had a duration of 12 weeks or more and three less than 4 weeks [57–59] (median 10 weeks [8–12]) as well as half used a fixed dose schedule (51%,  $k = 84$ ) and had a washout from psychotropic drugs (55%,  $k = 75$ ), yet definitions and duration varied. Placebo lead-in with exclusion of those with a positive response was used in five studies.

All of the studies used standardized diagnostic criteria, except five that used only diagnostic tools [60–64]. Associated conditions were the focus of and they were





required for inclusion in 29 studies (irritability in 52% of them), and a genetic syndrome in one (neurofibromatosis-type-1) [60]. Core symptoms was the primary focus in 34 trials, while the focus was unclear in 23 studies. Nine studies included participants using a threshold of core symptom severity, using ABC-L/SW, CARS, RLRS, SRS, and YBOCS-versions (in five out of eight). Participants on placebo had a median age of 8.63 years ([6.55–10.16],  $k = 82$ ), 17.16% were female ([0–20.9%],  $k = 80$ ), and 54.2% had comorbid intellectual disability ([0–75%],  $k = 30$ ). The median of baseline CGI-Severity was 4.73 ([4.39–5.00],  $k = 38$ ) and ABC-Irritability was 17.18 ([13.71–22.70],  $k = 36$ ).

Overall, 40% of the studies had an overall low risk of bias, 52% moderate and 8% high. Description of the methods was adequately reported in more than half of the studies for sequence generation (63%), allocation concealment (54%), and blinding (72%). Missing outcomes were adequately addressed in 60%, with a median overall dropout rate from placebo of 12.93% ([6.25–22.6%] and  $k = 70$  trials out of 86 reported attrition rates). Of the studies, 23% had a high risk of selective reporting, and 13% high risk in other biases, mainly due to imbalances between groups (Additional file 3: eAppendix-5.2). Finally, 38% of the studies were industry-sponsored (including five in which investigators applied for a patent on the experimental intervention), and sponsorship was unclear in three studies.

### Primary outcomes

#### *Social-communication difficulties*

**Primary analysis** In the primary analysis, 52 studies with 1497 participants on placebo were included. Most of the scales were filled by caregivers (77%). ABC-L/SW was the most used scale (56%) followed by VABS-Socialization (13%). Pooled placebo response was SMC =  $-0.32$  [95%CI  $-0.39, -0.25$ ], with moderate levels of heterogeneity ( $I^2 = 31.88\%$ ,  $\chi^2 = 74.87$ ,  $p = 0.02$ ) (Fig. 2).

**Sensitivity analysis and publication bias** The results did not change materially in sensitivity analyses (Additional file 3: eAppendix-6.1). There was no indication of small-study effects from a visual inspection of the funnel plot and Egger's test or publication bias ( $z = -0.38$ ,  $p = 0.70$ ) (Additional file 3: eAppendix-6.2). Also, fixed and random effects summaries were identical, an indication that smaller and larger studies give similar results.

**Meta-regressions** The results of the univariable and multivariable meta-regression analyses are presented in Table 1, Figure-S, and eAppendix-6.3. Placebo response in social-communication difficulties decreased over years (by 0.016 [0.003, 0.030] SMC units per publication year).

Larger placebo response was associated with caregiver ratings ( $-0.164$  [ $-0.315, -0.012$ ]), low risk of bias in other bias ( $-0.160$  [ $-0.299, -0.021$ ]), and higher baseline ABC-Irritability ( $-0.017$  [ $-0.028, -0.006$ ] per point). In the multivariable meta-regression, using the backward selection procedure, other bias, baseline ABC-Irritability, and the type of rater remained as covariates in the model, but the latter two were not significant due to their interaction with the other covariates. In a model without ABC-Irritability (available in 31 studies), publication year, other bias, and type of rater remained, the latter was not significant.

### *Repetitive behaviors*

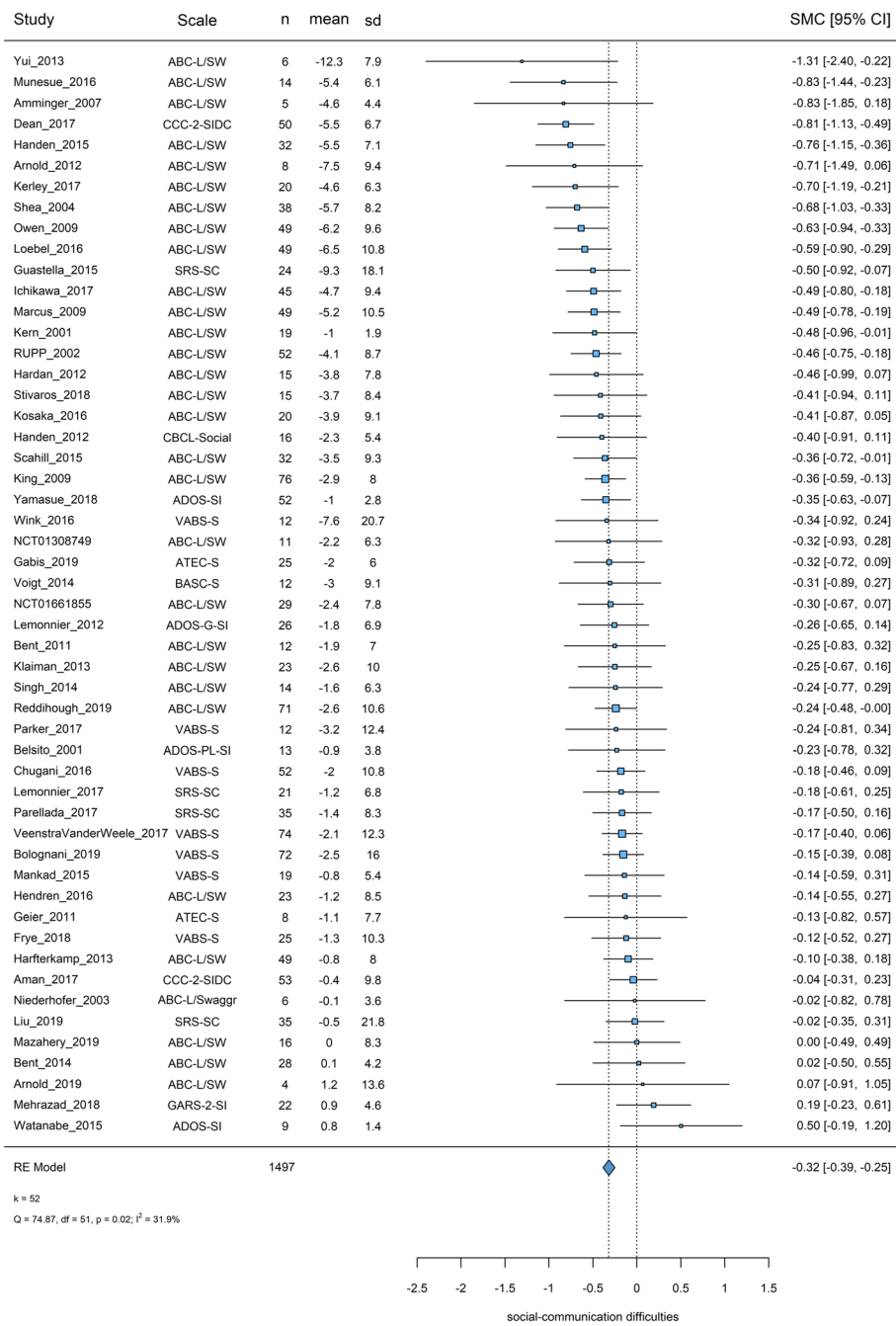
**Primary analysis** Fifty-two studies were included in the primary analysis with 1492 participants. Caregivers filled about half of the scales (56%). The most frequently used scales were ABC-S (44%) and YBOCS-versions (33%). Overall placebo response was SMC =  $-0.23$  [ $-0.32, -0.15$ ], and there was some heterogeneity ( $I^2 = 55\%$ ,  $\chi^2 = 113.32$ ,  $p < 0.001$ ) (Fig. 3).

**Sensitivity analysis and publication bias** Sensitivity analyses did not change the results materially, though there was a small difference between fixed and random-effects summary estimates, indicating possible small-study effects. Egger's test was not significant and yielded a marginal  $p$  value ( $z = 1.71$ ,  $p = 0.09$ ); it has been suggested that for this test, a threshold of 0.1 should be employed. By visual inspection of the funnel plot, we detected a possible asymmetry (Additional file 3: eAppendix-6.2) and the trim-and-fill adjusted placebo response was  $-0.33$  [ $-0.41, -0.25$ ].

**Meta-regressions** Higher placebo response was associated with larger sample sizes ( $-0.02$  [ $-0.03, -0.01$ ] SMC units per ten participants), flexible-dosing ( $-0.195$  [ $-0.351, -0.038$ ]), and the use of a threshold of core symptoms at inclusion ( $-0.346$  [ $-0.516, -0.175$ ]). These covariates remained in the multivariable model, but the use of a threshold of core symptoms at inclusion was not significant. Nevertheless, the findings might have been driven by three antidepressant trials in children/adolescents [65–67], with larger sample sizes ( $\sim 150$ ) and multiple sites (3, 6, and 18), as well as using flexible-dosing and a threshold of CYBOCS-PDD for inclusion (Table 1).

### *Overall core symptoms*

**Primary analysis** Forty-five studies with 1063 participants were included in the primary analysis. Caregivers filled about half of the scales (51%). The most frequently



**Fig. 2** Placebo response in scales measuring social-communication difficulties. Squares and bars represent standardized mean changes (SMC) and 95% confidence intervals for each study. The size of the square is proportional to the weight of the study in the meta-analysis. The diamond represent the pooled SMC. Heterogeneity is quantified with a  $\chi^2$  test ( $Q$ ) and  $I^2$ . \*In Chugani 2016, standard errors might have been reported as SDs. Therefore, we calculated SDs from the reported values (no reply from the corresponding author). It should be noted that in Niederhofer 2003, an aggregated score of ABC-L/SW rated by both caregivers and teachers were reported, in Amminger 2007, ABC-L/SW was rated by clinicians of the day care center. Scale: the scale used (clinician rated scales based on observation or interviews were preferred in the primary analysis);  $n$ : the number of participants on placebo; mean: mean change from baseline to endpoint (negative values for improvement); sd: the standard deviation used for the standardization (baseline standard deviations were preferred); SMC: standardized mean changes, 95% CI: 95% confidence intervals,  $k$  = total number of studies included in the analysis

**Table 1** Univariable meta-regression analyses

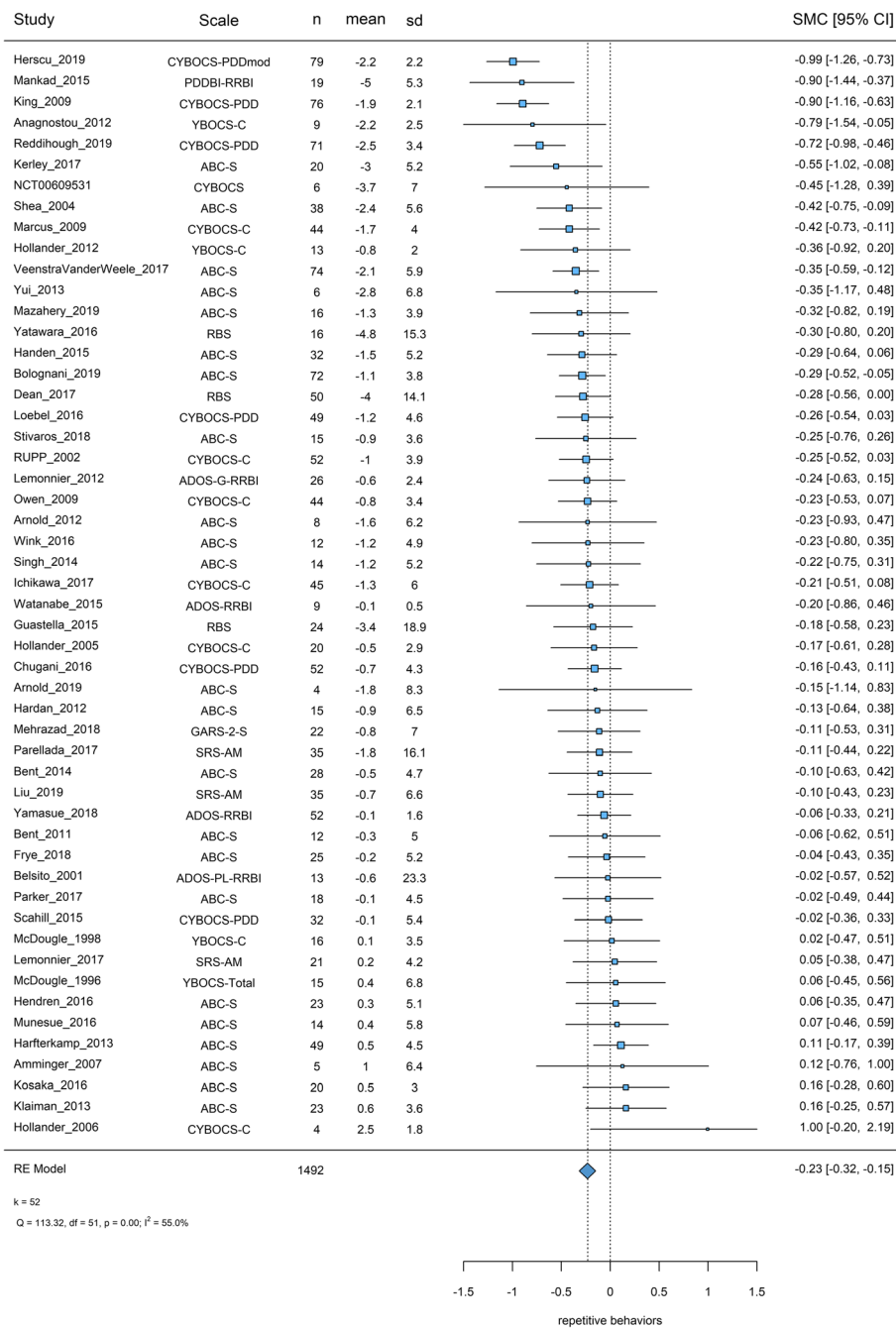
Covariate	Social-communication difficulties			Repetitive behaviors and restricted interests			Overall core symptoms		
	k, n	β [95% CI]	R <sup>2</sup> (%)	k, n	β [95% CI]	R <sup>2</sup> (%)	k, n	β [95% CI]	R <sup>2</sup> (%)
Intervention									
Route (ref. not oral)	52, 1497	0 [-0.205, 0.206]	0.998	52, 1492	-0.166 [-0.391, 0.059]	0.148	45, 1063	-0.042 [-0.301, 0.217]	0.749
Experimental intervention (ref. dietary supplement)	52, 1497	-0.082 [-0.229, 0.064]	0.268	52, 1492	-0.081 [-0.262, 0.1]	0.382	45, 1063	-0.023 [-0.228, 0.181]	0.822
Dose schedule (ref. fixed)**	51, 1491	-0.095 [-0.229, 0.038]	0.163	<b>51, 1486</b>	<b>-0.195 [-0.351, -0.038]</b>	<b>0.015</b>	44, 1057	-0.044 [-0.276, 0.189]	0.713
Study design									
Publication year	<b>52, 1497</b>	<b>0.016 [0.003, 0.03]</b>	<b>0.017</b>	52, 1492	-0.006 [-0.022, 0.009]	0.416	45, 1063	-0.004 [-0.021, 0.013]	0.607
Country (ref. outside US/mixed)	52, 1497	-0.002 [-0.139, 0.135]	0.973	52, 1492	-0.024 [-0.194, 0.145]	0.778	45, 1063	-0.054 [-0.257, 0.148]	0.598
Sponsorship (ref. no)	50, 1469	0.078 [-0.054, 0.21]	0.246	51, 1470	0.002 [-0.172, 0.175]	0.984	43, 996	0.057 [-0.152, 0.266]	0.590
No. sites*	48, 1395	-0.003 [-0.008, 0.003]	0.295	48, 1427	-0.004 [-0.011, 0.003]	0.287	<b>41, 957</b>	<b>-0.025 [-0.045, -0.005]</b>	<b>0.015</b>
% academic sites	46, 1340	-0.196 [-0.435, 0.044]	0.11	47, 1383	-0.08 [-0.397, 0.236]	0.619	41, 957	-0.117 [-0.461, 0.227]	0.505
No. arms	52, 1497	-0.036 [-0.208, 0.136]	0.684	52, 1492	0.035 [-0.182, 0.252]	0.753	Insufficient data		
Duration (weeks)	52, 1497	0.002 [-0.007, 0.012]	0.669	52, 1492	-0.009 [-0.021, 0.003]	0.124	45, 1063	0.003 [-0.009, 0.015]	0.677
Washout (ref. no)	46, 1404	-0.064 [-0.204, 0.076]	0.369	47, 1339	-0.006 [-0.165, 0.153]	0.943	39, 995	-0.021 [-0.248, 0.205]	0.855
Sample size**	52, 1497	0 [-0.001, 0.001]	0.967	<b>52, 1492</b>	<b>-0.002 [-0.003, -0.001]</b>	<b>0.004</b>	45, 1063	-0.001 [-0.004, 0.001]	0.213
% participants on placebo	52, 1497	0.172 [-0.625, 0.968]	0.673	52, 1492	-0.333 [-1.31, 0.643]	0.503	45, 1063	0.401 [-0.981, 1.783]	0.569
Rater (ref. clinician)***	<b>51, 1491</b>	<b>-0.164 [-0.315, -0.012]</b>	<b>0.034</b>	51, 1483	0.131 [-0.033, 0.294]	0.117	43, 1009	-0.148 [-0.361, 0.065]	0.174
Sequence generation (ref. unclear)	52, 1497	0.147 [-0.047, 0.34]	0.138	52, 1492	-0.043 [-0.302, 0.216]	0.743	45, 1063	0.243 [-0.048, 0.533]	0.102
Allocation concealment (ref. unclear)	52, 1497	-0.045 [-0.228, 0.138]	0.631	52, 1492	-0.012 [-0.233, 0.208]	0.912	<b>45, 1063</b>	<b>-0.252 [-0.485, -0.019]</b>	<b>0.034</b>
Blinding (ref. unclear/high)	52, 1497	0.083 [-0.068, 0.235]	0.28	52, 1492	0.006 [-0.191, 0.202]	0.953	45, 1063	-0.019 [-0.269, 0.231]	0.881
Missing outcome (ref. unclear/high)	52, 1497	-0.141 [-0.293, 0.01]	0.067	52, 1492	-0.052 [-0.247, 0.143]	0.6	45, 1063	-0.162 [-0.362, 0.038]	0.111
Selective reporting (ref. unclear/high)	52, 1497	-0.014 [-0.254, 0.227]	0.912	52, 1492	-0.156 [-0.471, 0.160]	0.333	Insufficient data		12.72

**Table 1** Univariable meta-regression analyses (Continued)

Covariate	Social-communication difficulties			Repetitive behaviors and restricted interests			Overall core symptoms					
	k, n	β [95% CI]	p	R <sup>2</sup> (%)	k, n	β [95% CI]	p	R <sup>2</sup> (%)	k, n	β [95% CI]	p	R <sup>2</sup> (%)
Other bias (ref. unclear/high)	52, 1497	-0.160 [-0.299, -0.021]	0.024	27.38	52, 1492	-0.091 [-0.303, 0.120]	0.398	0	45, 1063	-0.033 [-0.254, 0.188]	0.768	0
Participant (ref. adults/mixed)	52, 1497	-0.057 [-0.27, 0.155]	0.597	0	52, 1492	-0.107 [-0.323, 0.109]	0.33	1.55	45, 1063	0.148 [-0.124, 0.42]	0.287	9.35
Mean age	51, 1478	0.002 [-0.011, 0.014]	0.809	0	51, 1486	0.003 [-0.008, 0.014]	0.547	0	45, 1063	-0.007 [-0.021, 0.007]	0.297	7.37
% female	50, 1453	0.131 [-0.65, 0.911]	0.743	0	50, 1482	-0.282 [-1.114, 0.55]	0.507	0	45, 1063	0.277 [-0.650, 1.203]	0.559	0
% intellectual disability	17, 620	-0.328 [-0.672, 0.017]	0.063	42.21	21, 672	0.128 [-0.335, 0.59]	0.588	0	14, 316	0.382 [-0.105, 0.869]	0.124	20.87
% Caucasian or Hispanic	29, 963	0.036 [-0.351, 0.423]	0.855	0	29, 971	-0.157 [-0.715, 0.402]	0.583	0	21, 545	-0.42 [-1.204, 0.364]	0.294	0
Associated conditions at baseline (ref. no)	52, 1497	-0.068 [-0.212, 0.075]	0.351	3.45	52, 1492	0.111 [-0.065, 0.287]	0.216	5.83	45, 1063	-0.012 [-0.242, 0.218]	0.919	0
Baseline mean BMI	12, 445	-0.063 [-0.156, 0.029]	0.179	19.01	12, 461	-0.02 [-0.132, 0.091]	0.720	0	8, 254	-0.033 [-0.122, 0.057]	0.476	0
Baseline mean CGI-S	26, 917	-0.042 [-0.25, 0.167]	0.694	0	26, 925	0.033 [-0.227, 0.292]	0.805	0	17, 513	-0.077 [-0.442, 0.288]	0.678	0
Baseline mean ABC-Irritability	31, 917	-0.017 [-0.028, -0.006]	0.002	100	30, 884	-0.005 [-0.022, 0.013]	0.608	0	17, 340	-0.006 [-0.035, 0.023]	0.68	0
Minimum threshold of core symptoms for inclusion (ref. no)**	52, 1497	0.085 [-0.094, 0.264]	0.358	0.88	52, 1492	-0.346 [-0.516, -0.175]	<	53.85	Insufficient data			

A negative coefficient represent an increase of placebo response. For dichotomous covariates, the reference level is mentioned. Meta-regression with dichotomous covariates were not performed when there were less than five data points for a level of the covariate (e.g., placebo lead-in or number of medications). For continuous covariates, the covariate refer to a change of 1 point of the variable, e.g., per year for publication year: decrease of placebo response in social-communication difficulties by 0.016 per year, % percentage of intellectual disability: increase of placebo response in social-communication difficulties by -0.328 from 0 to 1 (100%) of the participants had intellectual disability. \*The effect of number of sites on placebo response in overall core symptoms was not significant (k = 40, coefficient 0.0182 [-0.268, 0.0631], p = 0.4287), when one outlier study with 26 sites was excluded (Bolognani 2019). \*\*Placebo response in repetitive behaviors was not found to be predicted by sample size (k = 49, coefficient -0.001 [-0.002, 0.000], p = 0.052), flexible-dosing (k = 48, coefficient -0.073 [-0.189, 0.043], p = 0.218) and using a minimum threshold of core symptoms (k = 49, coefficient -0.083 [-0.24, 0.073], p = 0.2950), when three antidepressant trials (Herscu 2019, King 2009, and Reddihough 2019) were excluded. \*\*\*In the meta-regression of type of rater, Niederhofer 2003 was not included in social-communication difficulties (aggregated caregiver/teacher rating of ABC-L/5W), Saad 2015 was not included in overall core symptoms (it was not clear if CARS was rated only by parents or also filled by clinicians), and Anagnostou\_2012 was not included in overall core symptoms and repetitive behaviors (SRS and RBS-R might have been rated as self-reports)

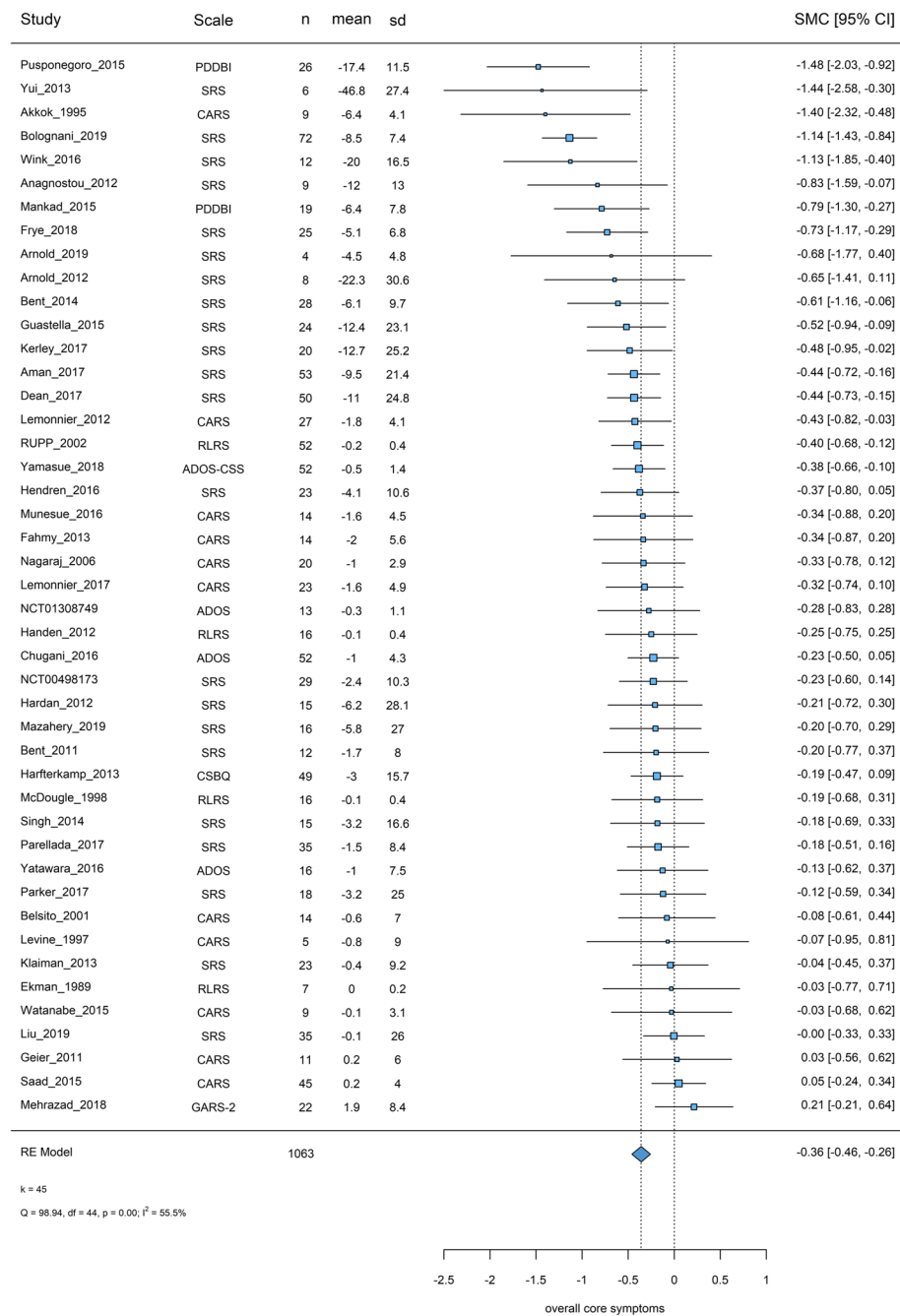




**Fig. 3** Placebo response in scales measuring repetitive behaviors. Squares and bars represent standardized mean changes (SMC) and 95% confidence intervals for each study. The size of the square is proportional to the weight of the study in the meta-analysis. The diamond represent the pooled SMC. Heterogeneity is quantified with a  $\chi^2$  test ( $Q$ ) and  $I^2$ . \*In Chugani 2016, standard errors might have been reported as SDs. Therefore, we calculated SDs from the reported values (no reply from the corresponding author). In Amminger 2007, ABC-S was rated by clinicians of the day care center. Scale: the scale used (clinician rated scales based on observation or interviews were preferred in the primary analysis); n: the number of participants on placebo; mean: mean change from baseline to endpoint (negative values for improvement); sd: the standard deviation used for the standardization (baseline standard deviations were preferred); SMC: standardized mean changes, 95% CI: 95% confidence intervals,  $k$  = total number of studies included in the analysis

used scales were SRS (49%) and CARS (24%). Overall placebo response was SMC = -0.36 [-0.46, -0.26] and heterogeneity was considerable ( $I^2 = 55.53\%$ ,  $\chi^2 = 98.94$ ,  $p < 0.001$ ) (Fig. 4).

**Sensitivity analysis and publication bias** Sensitivity analyses did not change the results materially, no asymmetry was detected in the funnel plot, and Egger’s test yielded a marginal  $p$  value ( $z = -1.82$ ,  $p = 0.07$ ).



**Fig. 4** Placebo response in scales measuring overall core symptoms. Squares and bars represent standardized mean changes (SMC) and 95% confidence intervals for each study. The size of the square is proportional to the weight of the study in the meta-analysis. The diamond represent the pooled SMC. Heterogeneity is quantified with a  $\chi^2$  test ( $Q$ ) and  $I^2$ . \*In Anagnostou 2012, we reversed baseline and endpoint values of SRS: in the manuscript, original baseline values were lower than endpoint in both placebo and oxytocin arms, meaning an increase of severity of symptoms during the study, which is not consistent with the reported positive effect size and the other outcomes (no reply from the corresponding author), in Saad 2015, CARS was rated by caregivers but it was unclear if also filled by clinicians (no reply from the corresponding author), as well as in RUPP 2002 and Handen 2012 the Ritvo-Freeman Life Rating Scale was rated by caregivers. Scale: the scale used (clinician rated scales based on observation or interviews were preferred in the primary analysis); n: the number of participants on placebo; mean: mean change from baseline to endpoint (negative values for improvement); sd: the standard deviation used for the standardization (baseline standard deviations were preferred); SMC: standardized mean changes, 95% CI: 95% confidence intervals, k= total number of studies included in the analysis

**Meta-regressions** Larger placebo response was associated with more trial sites ( $-0.025$  [ $-0.045, -0.005$ ] SMC units per site, not significant when an outlier was removed) [68], and low risk of bias in allocation concealment ( $-0.252$  [ $-0.485, -0.019$ ]). In the multivariable model, allocation concealment and number of sites were both significant.

Number of medications and the use of placebo lead-in had not sufficient data for all outcomes, while number of arms, selective reporting, and the use of the threshold of core symptoms did not have sufficient data for meta-regressions in overall core symptoms.

## Secondary outcomes

### Placebo response by type of rater

Results based on scales filled by different type of raters (Additional file 3: eAppendix-6.4) were similar to those of meta-regressions by type of rater (one effect size per study, clinician ratings were preferred whenever available).

### CGI-I positive response rates

The overall positive response rate as defined by at least much improvement in the CGI-I was 19% [16–22%] ( $k = 57, n = 1686, I^2 = 53\%$ ) (Fig. 5). The anchoring system of CGI was unclear in 35 studies, while seven considered both core and associated symptoms (three used OACIS [69]), three reported separate evaluations for global autism symptoms and for the trial target symptom, and three considered mainly core symptoms and nine associated symptoms (two reported the RUPP-framework [70]) (Table-S).

### Post-hoc correlation analyses

**Between covariates** Exploratory correlations between covariates are presented in Additional file 3: eAppendix-5.4. Significant correlations with a Spearman's  $|\rho| > 0.5$  were found between sample size and number of sites ( $\rho = 0.77$ ), percentage of academic sites and sponsorship ( $\rho = -0.52$ ), as well as number of sites ( $\rho = -0.51$ ), risk of bias domain of sequence generation and the domain of selective reporting ( $\rho = 0.57$ ), number of arms and percentage of participants on placebo ( $\rho = -0.54$ ), and between other covariates with a large proportion of missing data, e.g., baseline ABC-Irritability, BML, CGI-S, and percentage of participants with intellectual disability (Additional file 3: eAppendix-5.3).

**Between placebo and drug response** SMCs of placebo and experimental intervention were correlated in social-communication difficulties (Spearman's  $\rho = 0.525, p < 0.001$ ) and overall core symptoms ( $\rho = 0.539, p < 0.001$ ), but no correlation was found in repetitive behaviors ( $\rho = 0.233, p = 0.096$ ) (Additional file 3: eAppendix-6.5).

## Discussion

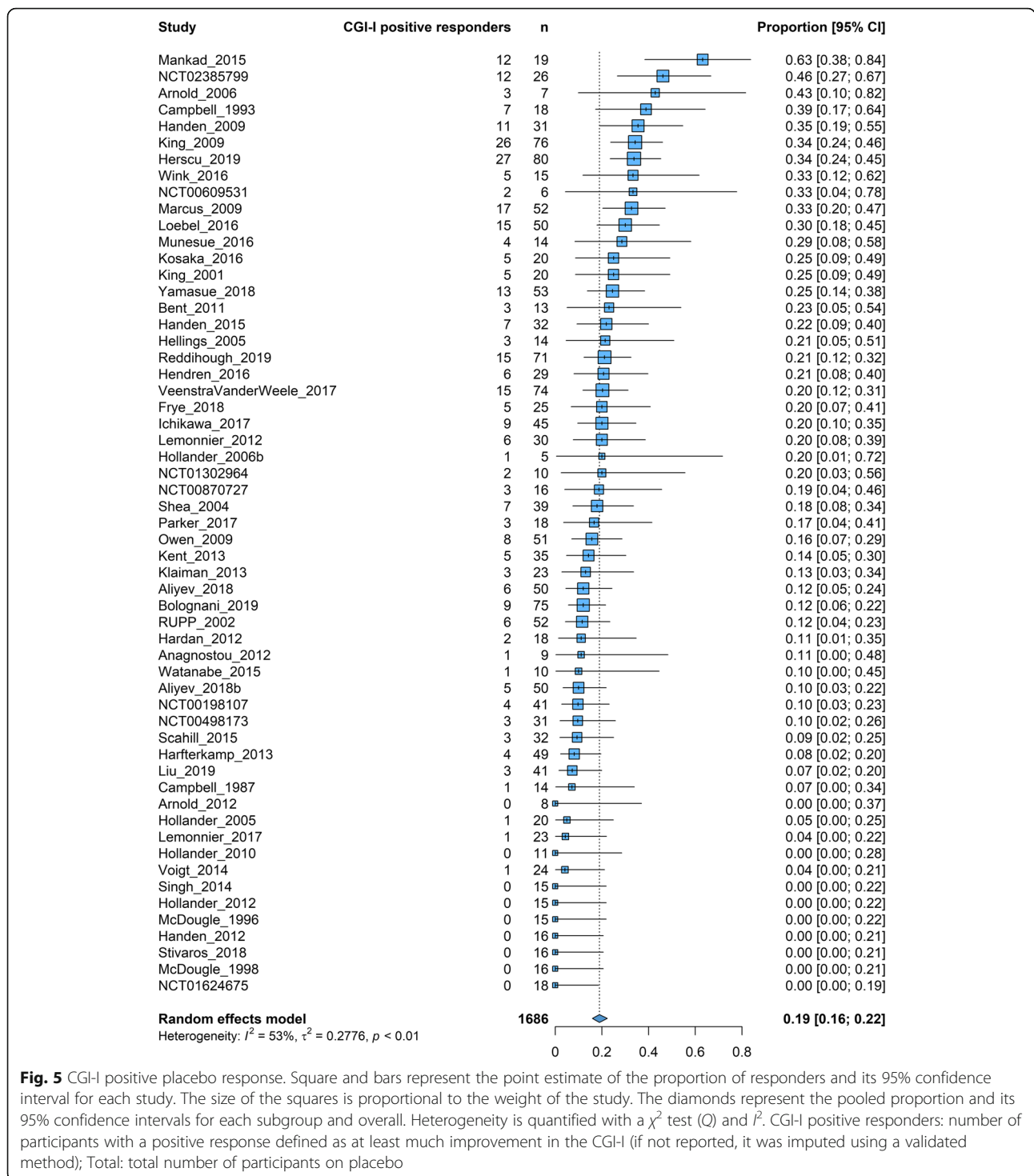
In pharmacological and dietary supplement ASD trials, placebo response was substantial and comparable among core symptoms; about 20% of the participants were at least much improved with placebo. We found potential predictors of larger placebo response in at least one symptom domain, i.e., baseline irritability, the use of a threshold of core symptoms at inclusion, caregiver ratings, larger sample size and number of sites, lower risk of bias, flexible-dosing, and less recent publication year.

### Predictors of placebo response

#### Participant-related factors

It has been argued that placebo response might be larger in children/adolescents than adults [71]. We did not find a difference between age groups or an effect of mean age. Nonetheless, extrapolations between age groups should be interpreted with caution because the majority of studies were in pediatric populations (87%). Other participant characteristics did not predict placebo response (e.g., sex, ethnicity, BMI, intellectual disability).

Low baseline severity has been found to predict placebo response in most psychiatric conditions [50]. We did not find an effect of baseline global severity (CGI-S), yet available data were sparse (baseline CGI-S was reported in less than half studies,  $k = 38, 44\%$ ) and narrowly ranged between 3.88 and 6 (Additional file 3: eAppendix-5.1); also because most of the studies required participants to be at least moderately ill (i.e., CGI-S  $\geq 4$ ). Baseline severity in core symptoms could not be analyzed as a potential predictor due to the large diversity of scales. On the other hand, we found that trials using a cut-off of core symptoms for inclusion might have a larger placebo response in repetitive behaviors, yet this association was not significant in a multivariable meta-regression and it might have been driven by three antidepressant trials that used a cut-off of the clinician-administered scale CYBOCS-PDD [65–67]. Trials that utilize a baseline score cut-off could be prone to regression to the mean effects as well as baseline score inflation, especially for clinician-administered scales and under participant recruitment pressure [72]. These effects could be partially avoided by using different scales at assessing participants for inclusion and as primary outcomes [73], yet this might be challenging given the lack of optimal scales in ASD. Centralized raters blind to inclusion criteria might also reduce baseline inflation and increase inter-rater reliability, yet the execution of the trial could become complicated [72]. Since inflated scores are usually very close to the inclusion cut-off, a potential solution could be that the primary analysis is conducted by including participants with a higher cut-off (that is blinded to the investigators) than the inclusion cut-off [74].



**Fig. 5** CGI-I positive placebo response. Square and bars represent the point estimate of the proportion of responders and its 95% confidence interval for each study. The size of the squares is proportional to the weight of the study. The diamonds represent the pooled proportion and its 95% confidence intervals for each subgroup and overall. Heterogeneity is quantified with a  $\chi^2$  test ( $Q$ ) and  $I^2$ . CGI-I positive responders: number of participants with a positive response defined as at least much improvement in the CGI-I (if not reported, it was imputed using a validated method); Total: total number of participants on placebo

The presence of an associated condition was required as inclusion criteria in about one-third of the trials (29 out of 86), and it was not found to predict placebo response. Nevertheless and since co-occurring symptoms and diagnoses are highly prevalent in participants with ASD [5], it can be expected that participants in other

studies had also associated symptoms of varying levels. Accordingly, the median of baseline ABC-Irritability was 17.18 IQR [13.71–22.70], while normative data suggested a mean of 12.8 [75]. Thus, our sample in general could be consisted of participants with somewhat higher levels of irritability. Indeed, the most frequently investigated

associated condition in our sample was irritability ( $k = 15$ ) and the presence of an associated condition was correlated to baseline ABC-Irritability ( $\rho = 0.49$ ,  $p < 0.001$ , Additional file 3: eAppendix-5.4). We found that baseline ABC-Irritability was associated with a larger placebo response in social-communication difficulties, yet this association was not significant in a multivariable meta-regression. The contrary was found in a quite large trial ( $n = 149$ ) investigating citalopram for repetitive behaviors, yet participants had lower levels of irritability (mean ABC-Irritability = 11.2) [9]. Additionally, a small 8-week observational study investigating the effects of participation in a study protocol suggested that placebo-effects may be mainly observed in children with higher levels of irritability [76]. Such participation effects could be decreased by a screening phase with adequate duration, which could also investigate the stability of symptoms and incorporate a potential washout of psychotropic drugs. However, no effect was found for the use of a washout phase and there were not enough data to investigate the use of a placebo lead-in phase, which is in general not recommended [72].

#### Design- and intervention-related factors

Caregiver ratings seemed to be more prone to placebo response in social-communication difficulties, but the effect was not consistent in multivariable meta-regressions. It has been argued that placebo-by-proxy effects are important components of placebo response in child/adolescent psychiatry, since they can alter caregiver perception of symptoms (thus improving directly scores in caregiver scales), and/or modify caregiver behaviors toward children and subsequently improving symptoms (thus improving scores also in non-caregiver scales) [71, 77]. In addition, many of the existing scales were not designed to measure change but rather as screening (e.g., SRS [31]) or diagnostic tools (CARS [32] and ADOS [78]), and efforts have been made for their improvement and adaptation, such as the ADOS calibrated severity score [79]. Given the lack of optimal scales, CGI has been extensively used and it is recommended for all trials irrespective of their target in order to investigate global autism symptoms and incorporate both core and associated symptoms [80, 81]. However, the anchoring system of CGI should be clearly reported, since it could vary materially among trials with different target symptoms (Table-S).

Therefore, there is a critical need to develop standardized and sensitive measures of core symptoms, which do not solely depend on caregivers [82, 83]. The semi-structured interview of VABS might be a promising measure of change in social-communication difficulties [33], with potential sensitivity to detect efficacy [68, 84] and empirically derived cut-offs of minimal-clinical-

important differences [85]. Recent instruments have also been developed, among others the Brief Observation of Social Communication Change (BOSCC) [86, 87], the Autism Behavior Inventory [88], and the Autism Impact Measure (AIM) [89], but their utilization is yet to be determined. Patient- (or parent-) reported outcomes have also gained recently greater attention [90], yet they should not be considered immune to placebo-effects [91]. The utilization of scales that require more extensive training and experience (e.g., ADOS, BOSCC, and VABS) might be challenging in larger scale trials, and thus a low inter-rater reliability could increase the variance of measurements and subsequently decrease drug-placebo differences. A notable example is the multi-center arbaclofen trial [84], in which VABS should have been completed by the same clinician and caregiver for each participant. However, there was quite low adherence to the protocol (rater change in about 25% of the participants), potentially because VABS-Socialization was a secondary outcome, not expected to be sensitive in the context of the trial. A post-hoc per-protocol analysis of no rater change found a significant improvement of arbaclofen in comparison to placebo, in contrast to the non-significant difference of the primary analysis [84]. Therefore, proper training of the raters and inter-rater reliability of the measurements as well as guidance and adherence to the protocol should be ensured, especially in multi-site trials.

Sample size and number of sites have been suggested as predictors of placebo response [50, 51, 92]. We also found that a larger sample size was associated with a larger placebo response in repetitive behaviors, yet the results might be driven by three antidepressant trials [65–67]. This association could also be explained by a potential publication bias and the small-study effects found in the funnel plot (see Additional file 3: eAppendix-6.2), since the results of less precise trials with larger placebo response in repetitive behaviors might have been not published. Additionally, sample size was closely related to the number of sites (Spearman's  $\rho = 0.77$ ,  $p < 0.001$ , see Additional file 3: eAppendix-6), which predicted placebo response in overall core symptoms, yet the latter was driven by another outlier study with 26 sites [68]. Trials with more sites were more frequently industry-sponsored ( $\rho = 0.27$ ,  $p = 0.04$ ) and consisted of less academic sites ( $\rho = -0.51$ ,  $p < 0.001$ ). It should be noted though that the majority of included studies were single-center (median number of sites 1 IQR [1–4]), had academic sites (about 83% consisted only of academic sites), and small sample sizes (median 45 IQR [30–91]); therefore, the results could not be extrapolated to a wider range of potential values. Nevertheless, more sites and the recruitment of non-academic professional sites, which could have less experience and enroll



competitively, might increase variability, be prone to less rigorous participant selection and baseline score inflation [73, 74, 92]. Therefore, trials should be well powered, yet extremely large sample sizes could be avoided, as well as sites should be carefully selected and their number should be kept at the minimum feasible.

Studies with low risk of bias in other biases (mainly baseline imbalance) and allocation concealment were associated with larger placebo response in social-communication and overall core symptoms, respectively. It is intriguing that studies with a better quality in terms of risk of bias might have a larger placebo response. However, the above risk of bias domains evaluate the randomization process, and in inadequately randomized trials, control groups might have a poorer prognosis [93].

The association between dosing schedule and placebo response can be puzzling, e.g., both flexible [94] and fixed-dosing schedules [95] have been associated with larger placebo responses in depression. We found an association between flexible-dosing and larger placebo response in repetitive behaviors, yet it was driven by three antidepressant trials [65–67]. Flexible-dosing could allow dose optimization guided by clinical response and/or the occurrence of side effects. The dose titration schedule and criteria as well as the starting dose and dose ranges should be carefully selected in the context of large placebo responses. For example, in one the aforementioned antidepressant trials, large placebo responses (> 25% reduction from baseline in CYBOCS-PDD) might have impeded dose escalation from a low starting dose (2 mg of fluoxetine) to a stable appropriate dose (> 10 mg) for sufficient duration of treatment (> 4 weeks) [67]. On the other hand, dose-response studies are a special type of fixed-dosing studies that might be prone to larger placebo responses. They are multi-arm and participants have an increased chance to receive active medication, as well as larger sample sizes and multiple sites are usually required. These factors have been associated with a larger placebo response in psychiatry [50], yet not all of them were replicated in our analysis, probably due to the limited number of studies with those characteristics. A notable example is the dose-response study of aripiprazole [96], which had a placebo positive response rate of 33% in comparison to 16% in the similarly designed but flexible-dosing study [97] (Fig. 3). However, this has not always been observed, such as in risperidone trials, i.e., 14% in the dose-response study [98] in comparison to 12% [53] and 18% [99] in the flexible-dose studies.

Country of origin and type of experimental intervention (pharmacological or dietary supplement) was not found to predict placebo response, in contrast to a

previous meta-analysis [11], which included also many Iranian trials with risperidone-combined treatments that were excluded from our review (combination treatments such as risperidone + placebo were excluded, see Additional file 3: eAppendix-4). Therefore, the findings in the previous meta-analysis could have been confounded by larger responses in combined placebo groups, i.e., response of risperidone + placebo.

There is no clear consensus about the adequate trial duration and half of the included studies had a duration of at least 12 weeks, yet the duration of the trial should be based on the mechanism of action of the experimental intervention and a longer duration could be required in order to observe sustained changes in core symptoms [100]. We did not find an effect of trial duration, yet shorter-term trials have been associated with larger placebo response in psychiatry [50]. However, in longer-term trials including young children, anticipated developmental trajectories could also explain placebo effects and subsequently mask drug-placebo differences [101]. Therefore, developmentally based scales might be necessary to overcome this challenge [82] as well as trial designs could include additional follow-up assessments in order to confirm stability of improvement [101].

In most psychiatric disorders, placebo response has increased over a period of 60 years [49, 50, 102], but this trend was not replicated in ASD trials, which were more recent, mainly published between 2009 and 2017. Even, placebo response in social-communication difficulties might have decreased over years. However, this effect was not found when ABC-Irritability was included in multivariable meta-regression. Temporal changes in the definition of ASD and research practices might play an important role per se, as differences between ASD and neurotypical populations might have been decreased over the years [103].

### Limitations

Our analysis has limitations. First, our analysis focused on placebo response in core symptoms of pharmacological and dietary supplement interventions. Therefore, we did not investigate placebo response in associated symptoms or of psychological/behavioral or multimodal interventions, which could also be of interest. However, core symptoms was the apparent focus in about 40% of the included trials, while many trials focused on associated symptoms, mainly irritability or ADHD symptoms. Second, there was a large diversity of scales used as well as a wide variability of their use, e.g., different CGI-I anchoring systems. Third, moderators of drug-placebo differences were not investigated and efforts to minimize placebo response could also affect drug response, since they were correlated in social-communication difficulties and overall core symptoms, but not in repetitive

behaviors (Additional file 3: eAppendix-6.5). In addition, some predictors might have a different impact on placebo and drug response [51]. Nevertheless, a more fine-grained analysis was impeded by the use of diverse experimental interventions with essentially different mechanisms of action (Additional file 3: eAppendix-5.1), e.g., contrary to schizophrenia [49, 51, 102], for which antipsychotics are the cornerstone of treatment [104].

Fourth, a common estimated pre-post correlation was used, but effect sizes were not materially changed in sensitivity analyses (Additional file 3: eAppendix-6.1). Fifth, despite the large number of eligible studies, about half did not provide data in spite of our efforts (authors of 85% of included studies published after 1990 could be contacted, with a reminder e-mail in case of no response, and 17% of them provided additional data/clarifications, Additional file 3: eAppendix-4), and a priori we did not use data from the whole crossover period (in forty trials), in order to avoid carry-over effects [14]. Sixth, due to the fact that information for many predictors, especially for participant-related factors (Additional file 3: eAppendix-5), was missing in many studies, we could not employ a full multivariable meta-regression and we focused on a series of univariable meta-regressions. Therefore, we cannot exclude the possibility of omitted variable bias in the results, i.e., the fact that the effect of the omitted variables may be added to the predictor considered in the univariable meta-regression. It should be noted that meta-regressions of aggregate data have an observational nature and they are prone to ecological fallacy, thus our findings should be considered exploratory and hypothesis-generating, considering also that there was no adjustment to multiple testing. Accordingly, individual-participant-data meta-analysis could allow for a more fine-grained analysis and further elucidate the impact of participant-level factors, such as age, sex, as well as baseline severity of core/associated symptoms.

## Conclusions

In order to increase the detection of efficacy of experimental interventions for ASD, high-quality and adequately powered trials are required, and predictors of placebo response should be considered. Extremely large sample sizes could be avoided and when multiple sites are needed, they should be carefully selected, trained, and monitored as well as their number should be kept at the minimum feasible. This would also facilitate a more rigorous selection of participants and a higher inter-rater reliability of measurements. Furthermore, scales that do not solely depend on caregiver reports could be selected as primary outcomes, since placebo-by-proxy effects are expected. Nevertheless, our findings highlight the urgent need for optimal and developmentally-based measures of change in core symptoms [82, 83]. The mechanism of

action of the experimental intervention could guide the selection of an appropriate, yet sufficiently long, trial duration as well as of the dose schedule and dose ranges. Participant-related factors, such as age, sex, and baseline severity of core/associated symptoms as well as factors that could differentially moderate drug response warrant further investigation. Last, in order to facilitate comparability between studies and synthesis of evidence, trials should better characterize their participants and improve their reporting, including the CGI anchoring system.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s13229-020-00372-z>.

**Additional file 1.**

**Additional file 2.**

**Additional file 3.**

## Abbreviations

ABC-L/SW: Aberrant Behavior Checklist-Lethargy/Social Withdrawal; ABC-S: Aberrant Behavior Checklist-Stereotypic Behavior; ADHD: Attention-deficit/hyperactivity disorder; ADI-R: Autism Diagnostic Interview-Revised; ADOS: Autism Diagnostic Observation Scale; ASD: Autism spectrum disorders; BOSCC: Brief Observation of Social Communication Change; BMI: Body mass index; CARS: Childhood Autism Rating Scale; CGI: Clinical global impression; (C)YBOCS-PDD: (Children) Yale Obsessive Compulsive Scale-Pervasive Developmental Disorders; DSM: Diagnostic and Statistical Manual of Mental Disorders; ICD: International Classification of Diseases; RLRs: Ritvo-Freeman Real Life Rating Scale; SD: Standard deviation; SMC: Standardized mean change; SRS: Social Responsiveness Scale; VABS: Vineland Adaptive Behavior Scale

## Acknowledgements

We would like to thank Farhad Sokrane, information specialist of the Cochrane Schizophrenia Group, who conducted the first search in electronic databases, Dr. Yikang Zhu for the translation of a Chinese study and Prof. Toshi Furukawa for the translation of a Japanese study. We would like to thank the following authors that the kindly contribution to the review by provided additional data and/or clarifications about their studies: Adi Aran, Kaat Alaerts, Nadir Aliyev, Eugene Arnold, Haim Belmaker, Yéhékel Ben-Ari, Leventhal Bennet, Stephen Bent, Helena Brentani, Jan Buitelaar, Ana Maria Castejon, Michael Chez, Torsten Danfors, Paulo Fontoura, Robert Grimaldi, Paul Gringras, Alexander Häge, Antonio Hardan, Robert Hendren, Janet Kern, Bruno Leheup, Wenn Liu, Raquel Martinez, James McCracken, Tali Nir, Deborah Pearson, Jeanette Ramer, Dan Rossignol, Kevin Sanders, Elisa Santocchi, Renato Scifo, Sarah Shea, Lawrence Scahill, Jeremy Veenstra-Vanderweele Paul Wang, David Wilensky, Hidenori Yamasue, and Lingli Zhang.

## Authors' contributions

SS (study design, study selection, data extraction, contacting authors for additional data, statistical analysis, interpretation of the data, drafting the first version of the manuscript, study supervision), OC (study selection, data extraction, contacting authors for additional data), JST (data extraction, technical support with Access database), IB (study selection), MK (study selection), AR (data extraction), AC (study selection), GD (study selection), MH (technical support with Access database), DF (study design, interpretation of the data), DM (statistical analysis, interpretation of the data), TC (interpretation of the data), DGM (interpretation of the data), MP (study design, interpretation of the data), CA (study design, interpretation of the data), SL (study design, data extraction, statistical analysis, interpretation of the data, drafting the first version of the manuscript, study supervision). All authors critically reviewed the manuscript for important intellectual content. The author(s) read and approved the final manuscript.

### Funding

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 777394 for the project AIMS-2-TRIALS. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and AUTISM SPEAKS, Autistica, SFARI. CA, MP and DF were supported by the Spanish Ministry of Science, Innovation and Universities. Instituto de Salud Carlos III, co-financed by ERDF Funds from the European Commission, "A way of making Europe", CIBERSAM, Madrid Regional Government (B2017/BMD-3740 AGES-CM-2), European Union Structural Funds and European Union Seventh Framework Program and H2020 Program; Fundación Familia Alonso, Fundación Alicia Koplowitz and Fundación Mutua Madrileña. The funding source had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. Open access funding provided by Projekt DEAL.

### Availability of data and materials

All data generated during this study are included in this published article (and its supplementary information files). The datasets analyzed during the current study are available from the corresponding author on reasonable request.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

In the last 3 years, Stefan Leucht has received honoraria as a consultant/ advisor and/or for lectures from LB Pharma, Otsuka, Lundbeck, Boehringer Ingelheim, LTS Lohmann, Janssen, Johnson&Johnson, TEVA, MSD, Sandoz, SanofiAventis, Angelini, Recordati, Sunovion, and Geodon Richter. David Fraguas has been a consultant and/or has received fees from Angelini, Eisai, IE4Lab, Janssen, Lundbeck, and Otsuka. He has also received grant support from Instituto de Salud Carlos III (Spanish Ministry of Science, Innovation and Universities) and from Fundación Alicia Koplowitz. Mara Parellada has received educational honoraria from Otsuka, research grants from FAK and Fundación Mutua Madrileña (FMM), Instituto de Salud Carlos III (Spanish Ministry of Science, Innovation and Universities) and European ERANET and H2020 calls, travel grants from Otsuka and Janssen. Consultant for Exeltis and Servier. Celso Arango has been a consultant to or has received honoraria or grants from Acadia, Angelini, Gedeon Richter, Janssen Cilag, Lundbeck, Otsuka, Roche, Sage, Sanofi, Servier, Shire, Schering Plough, Sumitomo Dainippon Pharma, Sunovion and Takeda. In the last 3 years, Maximilian Huhn has received speakers honoraria from Janssen. Declan Murphy has received consulting fees from Roche. The other authors have nothing to disclose.

### Author details

<sup>1</sup>Department of Psychiatry and Psychotherapy, School of Medicine, Technical University of Munich, Ismaningerstr. 22, 81675 Munich, Germany.

<sup>2</sup>Department of Child and Adolescent Psychiatry, School of Medicine, Balçova Dokuz Eylul University, İzmir, Turkey. <sup>3</sup>Department of Experimental and Clinical Medicine, Psychiatric Clinic University Hospital 'Gaspere Rodolico', University of Catania, Catania, Italy. <sup>4</sup>Department of Psychiatry, Spedali Civili Hospital, Brescia, Italy. <sup>5</sup>Department of Psychiatry, Psychosomatic Medicine and Psychotherapy Social Foundation Bamberg, Teaching Hospital of the University of Erlangen, Erlangen, Germany.

<sup>6</sup>Department of Child and Adolescent Psychiatry, Hospital General Universitario Gregorio Marañón, Institute of Psychiatry and Mental Health, IISGM, CIBERSAM, Madrid, Spain. <sup>7</sup>School of Medicine, Universidad Complutense, Madrid, Spain. <sup>8</sup>Department of Primary Education, University of Ioannina, Ioannina, Greece. <sup>9</sup>Faculté de Médecine, Université Paris Descartes, Paris, France. <sup>10</sup>Department of Psychology, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK. <sup>11</sup>Department of Forensic and Neurodevelopmental Sciences, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK.

Received: 1 April 2020 Accepted: 14 August 2020

Published online: 26 August 2020

### References

- American Psychiatric A. Diagnostic and statistical manual of mental disorders (DSM-5®): American Psychiatric Pub; 2013.
- Baio J, Wiggins L, Christensen DL, Maenner MJ, Daniels J, Warren Z, Kurzius-Spencer M, Zahorodny W, Robinson Rosenberg C, White T, et al. Prevalence of Autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 Sites, United States, 2014. *MMWR Surveill Summ.* 2018;67(6):1–23.
- Baxter AJ, Brugha TS, Erskine HE, Scheurer RW, Vos T, Scott JG. The epidemiology and global burden of autism spectrum disorders. *Psychol Med.* 2015;45(3):601–13.
- Buescher AVS, Cidav Z, Knapp M, Mandell DS. Costs of Autism Spectrum disorders in the United Kingdom and the United States costs of ASD in the UK and US costs of ASD in the UK and US. *JAMA Pediatrics.* 2014;168(8):721–8.
- Howes OD, Rogdaki M, Findon JL, Wichers RH, Charman T, King BH, Loth E, McAlonan GM, McCracken JT, Parr JR, et al. Autism spectrum disorder: consensus guidelines on assessment, treatment and research from the British Association for Psychopharmacology. *J Psychopharmacol.* 2018;32(1):3–29.
- Jobski K, Hofer J, Hoffmann F, Bachmann C. Use of psychotropic drugs in patients with autism spectrum disorders: a systematic review. *Acta Psychiatr Scand.* 2017;135(1):8–28.
- Berry-Kravis EM, Lindemann L, Jonch AE, Apostol G, Bear MF, Carpenter RL, Crawley JN, Curie A, Des Portes V, Hossain F, et al. Drug development for neurodevelopmental disorders: lessons learned from fragile X syndrome. *Nat Rev Drug Discov.* 2018;17(4):280–99.
- Gribkoff VK, Kaczmarek LK. The need for new approaches in CNS drug discovery: Why drugs have failed, and what can be done to improve outcomes. *Neuropharmacology.* 2017;120:11–9.
- King BH, Dukes K, Donnelly CL, Sikich L, McCracken JT, Scahill L, Holl e E, Bregman JD, Anagnostou E, et al. Baseline factors predicting placebo response to treatment in children and adolescents with autism spectrum disorders: a multisite randomized clinical trial. *JAMA Pediatrics.* 2013;167(11):1045–52.
- Arnold LE, Farmer C, Kraemer HC, Davies M, Witwer A, Chuang S, Disilvestro R, McDougle CJ, McCracken J, Vitiello B et al: Moderators, mediators, and other predictors of risperidone response in children with autistic disorder and irritability. *J Child Adolesc Psychopharmacol.* 2010;20(2):83–93..
- Masi A, Lampit A, Glozier N, Hickie IB, Guastella AJ. Predictors of placebo response in pharmacological and dietary supplement treatment trials in pediatric autism spectrum disorder: a meta-analysis. *Transl Psychiatry.* 2015;5:e640.
- Masi A, Lampit A, DeMayo MM, Glozier N, Hickie IB, Guastella AJ. A comprehensive systematic review and meta-analysis of pharmacological and dietary supplement interventions in paediatric autism: moderators of treatment response and recommendations for future research. *Psychol Med.* 2017;47(7):1323–34.
- Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ.* 2009;339:b2700.
- Elbourne DR, Altman DG, Higgins JPT, Curtin F, Worthington HV, Vail A. Meta-analyses involving cross-over trials: methodological issues. *Int J Epidemiol.* 2002;31(1):140–9.
- Whiting-O'Keefe QE, Henke C, Simborg DW. Choosing the correct unit of analysis in Medical Care experiments. *Med Care.* 1984;22(12):1101–14.
- Reichow B, Volkmar FR, Cicchetti DV. Development of the evaluative method for evaluating and determining evidence-based practices in autism. *J Autism Dev Disord.* 2008;38(7):1311–9.
- Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, Savovic J, Schulz KF, Weeks L, Sterne JA. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ.* 2011;343:d5928.
- Furukawa TA, Salanti G, Atkinson LZ, Leucht S, Ruhe HG, Turner EH, Chaimani A, Ogawa Y, Takeshima N, Hayasaka Y, et al. Comparative efficacy and acceptability of first-generation and second-generation antidepressants in the acute treatment of major depression: protocol for a network meta-analysis. *BMJ Open.* 2016;6(7):e010919.

19. Carrasco M, Volkmar FR, Bloch MH. Pharmacologic treatment of repetitive behaviors in autism spectrum disorders: evidence of publication bias. *Pediatrics*. 2012;129(5):e1301–10.
20. Sturman N, Deckx L, van Driel ML. Methylphenidate for children and adolescents with autism spectrum disorder. *Cochrane Database Syst Rev*. 2017;11:Cd011144.
21. Fung LK, Mahajan R, Nozzolillo A, Bernal P, Krasner A, Jo B, Coury D, Whitaker A, Veenstra-Vanderweele J, Hardan AY. Pharmacologic treatment of severe irritability and problem behaviors in autism: a systematic review and meta-analysis. *Pediatrics*. 2016;137(Suppl 2):S124–35.
22. Hirsch LE, Pringsheim T. Aripiprazole for autism spectrum disorders (ASD). *Cochrane Database Syst Rev*. 2016;2016(6):Cd009043.
23. Williamson E, Sathé NA, Andrews JC, Krishnaswami S, McPheeters ML, Fannesbeck C, Sanders K, Weitlauf A, Warren Z. AHRQ comparative effectiveness reviews. In: *Medical Therapies for Children With Autism Spectrum Disorder-An Update*. Rockville (MD): Agency for Healthcare Research and Quality (US); 2017.
24. Williams K, Brignell A, Randall M, Silove N, Hazell P. Selective serotonin reuptake inhibitors (SSRIs) for autism spectrum disorders (ASD). *Cochrane Database Syst Rev*. 2013;(8):Cd004677. <https://community.cochrane.org/style-manual/references/reference-types/cochrane-publications>.
25. Hurwitz R, Blackmore R, Hazell P, Williams K, Woolfenden S. Tricyclic antidepressants for autism spectrum disorders (ASD) in children and adolescents. *Cochrane Database Syst Rev*. 2012;(3):Cd008372. <https://community.cochrane.org/style-manual/references/reference-types/cochrane-publications>.
26. Nye C, Brice A. Combined vitamin B6-magnesium treatment in autism spectrum disorder. *Cochrane Database Syst Rev*. 2005;(4):Cd003497. <https://community.cochrane.org/style-manual/references/reference-types/cochrane-publications>.
27. Fraguas D, Díaz-Caneja CM, Pina-Camacho L, Moreno C, Durán-Cutilla M, Ayora M, González-Vioque E, de Matteis M, Hendren RL, Arango C. Dietary interventions for autism spectrum disorder: a meta-analysis. *Pediatrics*. 2019;144(5):e20183218.
28. Aman MG, Singh NN, Stewart AW, Field CJ. Psychometric characteristics of the aberrant behavior checklist. *Am J Ment Defic*. 1985;89(5):492–502.
29. Sparrow SS. Vineland Adaptive Behavior Scales. In: Kreutzer JS, DeLuca J, Caplan B, editors. *Encyclopedia of Clinical Neuropsychology*. New York, NY: Springer New York; 2011. p. 2618–21.
30. Scahill L, McDougle CJ, Williams SK, Dimitropoulos A, Aman MG, McCracken JT, Tierney E, Arnold LE, Cronin P, Grados M, et al. Children's Yale-Brown obsessive compulsive scale modified for pervasive developmental disorders. *J Am Acad Child Adolesc Psychiatry*. 2006;45(9):1114–23.
31. Constantino JN, Davis SA, Todd RD, Schindler MK, Gross MM, Brophy SL, Metzger LM, Shoushtari CS, Splinter R, Reich W. Validation of a brief quantitative measure of autistic traits: comparison of the social responsiveness scale with the autism diagnostic interview-revised. *J Autism Dev Disord*. 2003;33(4):427–33.
32. Schopler E, Reichler RJ, DeVellis RF, Daly K. Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *J Autism Dev Disord*. 1980;10(1):91–103.
33. Anagnostou E, Jones N, Huerta M, Halladay AK, Wang P, Scahill L, Horrigan JP, Kasari C, Lord C, Choi D, et al. Measuring social communication behaviors as a treatment endpoint in individuals with autism spectrum disorder. *Autism*. 2014;19(5):622–36.
34. Scahill L, Aman MG, Lecavalier L, Halladay AK, Bishop SL, Bodfish JW, Grondhuis S, Jones N, Horrigan JP, Cook EH, et al. Measuring repetitive behaviors as a treatment endpoint in youth with autism spectrum disorder. *Autism*. 2013;19(1):38–52.
35. European Medicines Agency: Guideline on the clinical development of medicinal products for the treatment of Autism Spectrum Disorder (ASD). In: Edited by CHMP; 2018: EMA/CHMP/598082/592013.
36. Bolte EE, Diehl JJ. Measurement tools and target symptoms/skills used to assess treatment response for individuals with autism spectrum disorder. 2013, 43(11):2491–2501.
37. Furukawa TA, Cipriani A, Barbui C, Brambilla P, Watanabe N. Imputing response rates from means and standard deviations in meta-analyses. *Int Clin Psychopharmacol*. 2005;20(1):49–52.
38. Samara MT, Spinelli LM, Furukawa TA, Engel RR, Davis JM, Salanti G, Leucht S. Imputation of response rates from means and standard deviations in schizophrenia. *Schizophr Res*. 2013;151(1-3):209–14.
39. Higgins JPT, Green S (editors). *Cochrane handbook for systematic reviews of interventions* version 5.1.0 [updated March 2011]. The Cochrane Collaboration; 2011. Available from <https://www.cochrane-handbook.org>. (see <https://community.cochrane.org/style-manual/references/reference-types/cochrane-publications>).
40. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol*. 2005;5(1):13.
41. Balk EM, Earley A, Patel K, Trikalinos TA, Dahabreh IJ. Empirical assessment of within-arm correlation imputation in trials of continuous outcomes. In: *Rockville: Agency for Healthcare Research and Quality (US); 2012: EHC141-EF*.
42. Furukawa TA, Barbui C, Cipriani A, Brambilla P, Watanabe N. Imputing missing standard deviations in meta-analyses can provide accurate results. *J Clin Epidemiol*. 2006;59(1):7–10.
43. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177–88.
44. Becker BJ. Synthesizing standardized mean-change measures. *Br J Math Stat Psychol*. 1988;41(2):257–78.
45. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010;36(3):1–48.
46. Schwarzer G, Chaimaitel H, Abu-Raddad LJ, Rücker G. Seriously misleading results using inverse of Freeman-Tukey double arcsine transformation in meta-analysis of single proportions. *Res Synth Methods*. 2019;10(3):476–83.
47. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629–34.
48. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*. 2000;56(2):455–63.
49. Leucht S, Chaimani A, Leucht C, Huhn M, Mavridis D, Helfer B, Samara M, Cipriani A, Geddes JR, Salanti G, et al. 60years of placebo-controlled antipsychotic drug trials in acute schizophrenia: Meta-regression of predictors of placebo response. *Schizophr Res*. 2018;201:315–23.
50. Weimer K, Colloca L, Enck P. Placebo effects in psychiatry: mediators and moderators. *Lancet Psychiatry*. 2015;2(3):246–57.
51. Leucht S, Chaimani A, Mavridis D, Leucht C, Huhn M, Helfer B, Samara M, Cipriani A, Geddes JR, Davis JM. Disconnection of drug-response and placebo-response in acute-phase antipsychotic drug trials on schizophrenia? Meta-regression analysis. *Neuropsychopharmacology*. 2019;44(11):1955–66.
52. Sandler A. Placebo effects in developmental disabilities: implications for research and practice. *Ment Retard Dev Disabil Res Rev*. 2005;11(2):164–70.
53. RUPP. Risperidone in Children with Autism and Serious Behavioral Problems. *N Engl J Med*. 2002;347(5):314–21.
54. Agid O, Kapur S, Arenovich T, Zipursky RB. Delayed-onset hypothesis of antipsychotic action: a hypothesis tested and rejected. *Arch Gen Psychiatry*. 2003;60(12):1228–35.
55. Schwarzer G. meta: An R package for meta-analysis. *R news*. 2007;7(3):40–5.
56. R Core Team. *R: A language and Environment for statistical computing*. Vienna: R Foundation for Statistical Computing; 2019. <https://www.R-project.org/>.
57. Akkoc FG, Bahar, Oktem, Ferhunde; Reid, Larry D.; Sucuoglu, Bulbin: Otizm'de naltrekson sagaltiminin davranissal ve biyokimyasal boyutlari. *Turk Psikiyatri Dergisi*. 1995;6(4):251–62.
58. Pusponegoro HD, Ismael S, Firmansyah A, Sastroasmoro S, Vandenplas Y. Gluten and casein supplementation does not increase symptoms in children with autism spectrum disorder. *Acta Paediatr*. 2015;104(11):e500–5.
59. Campbell M, Anderson LT, Small AM, Adams P, Gonzalez NM, Ernst M. Naltrexone in autistic children: behavioral symptoms and attentional learning. *J Am Acad Child Adolesc Psychiatry*. 1993;32(6):1283–91.
60. Stivaros S, Garg S, Tziraki M, Cai Y, Thomas O, Mellor J, Morris AA, Jim C, Szumanska-Ryt K, Parkes LM, et al. Randomised controlled trial of simvastatin treatment for autism in young children with neurofibromatosis type 1 (SANTA). *Mol Autism*. 2018;9:12.
61. Bent S, Hendren RL, Zandi T, Law K, Choi JE, Widjaja F, Kalb L, Nestle J, Law P. Internet-based, randomized, controlled trial of omega-3 fatty acids for hyperactivity in autism. *J Am Acad Child Adolesc Psychiatry*. 2014;53(6):658–66.
62. Fahmy SF, El-Hamamsy M, Zaki O, Badary OA. Effect of L-carnitine on behavioral disorder in autistic children. *Value Health*. 2013;16(3):A15.
63. Geier DA, Kern JK, Davis G, King PG, Adams JB, Young JL, Geier MR. A prospective double-blind, randomized clinical trial of levocarnitine to treat autism spectrum disorders. *Med Sci Monit*. 2011;17(6):Pi15–23.



64. Belsito KM, Law PA, Kirk KS, Landa RJ, Zimmerman AW. Lamotrigine therapy for autistic disorder: a randomized, double-blind, placebo-controlled trial. *J Autism Dev Disord.* 2001;31(2):175–81.
65. King BH, Hollander E, Sikich L, McCracken JT, Scahill L, Bregman JD, Donnelly CL, Anagnostou E, Dukas K, Sullivan L, et al. Lack of efficacy of citalopram in children with autism spectrum disorders and high levels of repetitive behavior: citalopram ineffective in children with autism. *Arch Gen Psychiatry.* 2009;66(6):583–90.
66. Reddihough DS, Marraffa C, Mouti A, O'Sullivan M, Lee KJ, Orsini F, Hazell P, Granich J, Whitehouse AJO, Wray J, et al. Effect of fluoxetine on obsessive-compulsive behaviors in children and adolescents with autism spectrum disorders: a randomized clinical trial. *JAMA.* 2019;322(16):1561–9.
67. Herscu P, Handen BL, Arnold LE, Snape MF, Bregman JD, Ginsberg L, Hendren R, Kolevzon A, Melmed R, Mintz M, et al. The SOFIA study: negative multicenter study of low dose fluoxetine on repetitive behaviors in children and adolescents with autistic disorder. *J Autism Dev Disord.* 2019;50(9):3233–44.
68. Bolognani F, Del Valle RM, Squassante L, Wandel C, Derks M, Murtagh L, Sevigny J, Khwaja O, Umbricht D, Fontoura P. A phase 2 clinical trial of a vasopressin V1a receptor antagonist shows improved adaptive behaviors in men with autism spectrum disorder. *Sci Transl Med.* 2019;11(491):eaat7838.
69. Butter E, Mulick J. The Ohio autism clinical impressions scale (OACIS). Columbus, OH: Children's Research Institute; 2006.
70. Arnold LE, Aman MG, Martin A, Collier-Crespin A, Vitiello B, Tierney E, Asarnow R, Bell-Bradshaw F, Freeman BJ, Gates-Ulanet P, et al. Assessment in multisite randomized clinical trials of patients with autistic disorder: the Autism RUPP Network. Research Units on Pediatric Psychopharmacology. *J Autism Dev Disord.* 2000;30(2):99–111.
71. Weimer K, Gulewitsch MD, Schlarb AA, Schwillie-Kiuntke J, Klosterhalfen S, Enck P. Placebo effects in children: a review. *Pediatr Res.* 2013;74(1):96–102.
72. Rutherford BR, Roose SP. A model of placebo response in antidepressant clinical trials. *Am J Psychiatry.* 2013;170(7):723–33.
73. Parellada M, Moreno C, Moreno M, Espieiro A, de Portugal E, Arango C. Placebo effect in child and adolescent psychiatric trials. *Eur Neuropsychopharmacol.* 2012;22(11):787–99.
74. 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.* 2014;51:21–9.
75. Kaat AJ, Lecavalier L, Aman MG. Validity of the aberrant behavior checklist in children with autism spectrum disorder. *J Autism Dev Disord.* 2014;44(5):1103–16.
76. Jones RM, Carberry C, Hamo A, Lord C. Placebo-like response in absence of treatment in children with Autism. *Autism Res.* 2017;10(9):1567–72.
77. Whalley B, Hyland ME. Placebo by proxy: the effect of parents' beliefs on therapy for children's temper tantrums. *J Behav Med.* 2013;36(4):341–6.
78. Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC, Pickles A, Rutter M. The autism diagnostic observation schedule—Generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord.* 2000;30(3):205–23.
79. Shumway S, Farmer C, Thurm A, Joseph L, Black D, Golden C. The ADOS calibrated severity score: relationship to phenotypic variables and stability over time. *Autism Res.* 2012;5(4):267–76.
80. Provenzano U, Fusar-Poli L, Brondino N, Damiani S, Vercesi M, Meyer N, Rocchetti M, Politi P. What are we targeting when we treat autism spectrum disorder? A systematic review of 406 clinical trials. *Autism.* 2020;24(2):274–84.
81. Aman MG, Novotny S, Samango-Sprouse C, Lecavalier L, Leonard E, Gadow KD, King BH, Pearson DA, Gernsbacher MA, Chez M. Outcome measures for clinical drug trials in autism. *CNS spectrums.* 2004;9(1):36–47.
82. Bishop S, Farmer C, Kaat A, Georgiades S, Kanne S, Thurm A. The need for a Developmentally Based measure of social communication skills. *J Am Acad Child Adolesc Psychiatry.* 2019;58(6):555–60.
83. Grzadzinski R, Janvier D, Kim SH. Recent developments in treatment outcome measures for young children with autism spectrum disorder (ASD). *Semin Pediatr Neurol.* 2020;34:100806.
84. Veenstra-VanderWeele J, Cook EH, King BH, Zarevics P, Cherubini M, Walton-Bowen K, Bear MF, Wang PP, Carpenter RL. Arbaclofen in children and adolescents with autism spectrum disorder: a randomized, controlled, phase 2 trial. *Neuropsychopharmacology.* 2017;42(7):1390–8.
85. Chatham CH, Taylor KI, Charman T, Liogier D'ardhuy X, Eule E, Fedele A, Hardan AY, Loth E, Murtagh L, Del Valle RM, et al. Adaptive behavior in autism: Minimal clinically important differences on the Vineland-II. *Autism Res.* 2018;11(2):270–83.
86. Grzadzinski R, Carr T, Colombi C, McGuire K, Dufek S, Pickles A, Lord C. Measuring changes in Social communication behaviors: preliminary development of the brief observation of social communication change (BOSSC). *J Autism Dev Disord.* 2016;46:2464–79.
87. Gengoux GW, Abrams DA, Schuck R, Millan ME, Libove R, Ardel CM, Phillips JM, Fox M, Frazier TW, Hardan AY. A Pivotal response treatment package for children with autism spectrum disorder: An RCT. *Pediatrics.* 2019;144(3):e20190178.
88. Bangerter A, Ness S, Aman MG, Esbensen AJ, Goodwin MS, Dawson G, Hendren R, Leventhal B, Khan A, Opler M, et al. Autism behavior inventory: a novel tool for assessing core and associated symptoms of autism spectrum disorder. *J Child Adolesc Psychopharmacol.* 2017;27(9):814–22.
89. Kanne SM, Mazurek MO, Sikora D, Bellando J, Branum-Martin L, Handen B, Katz T, Freedman B, Powell MP, Warren Z. The autism impact measure (AIM): initial development of a new tool for treatment outcome measurement. *J Autism Dev Disord.* 2014;44(1):168–79.
90. Graham SM, Zampella CJ, Clements C, JP MC, Maddox BB, Parish-Morris J, Udhmani MD, Schultz RT, Miller JS. A Lifespan approach to patient-reported outcomes and quality of life for people on the autism spectrum. *Autism Res.* 2020;13(6):970–87.
91. Estevinho MM, Afonso J, Rosa I, Lago P, Trindade E, Correia L, Dias CC, Magro F. Gedi: Placebo effect on the health-related quality of life of inflammatory bowel disease patients: a systematic review with meta-analysis. *J Crohns Colitis.* 2018;12(10):1232–44.
92. 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.* 2009;166(1):42–9.
93. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ.* 1998;317(7167):1185–90.
94. 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.* 2016;3(11):1059–66.
95. Khan A, Khan SR, Walens G, Kolts R, Giller EL. Frequency of positive studies among fixed and flexible dose antidepressant clinical trials: an analysis of the food and drug administration summary basis of approval reports. *Neuropsychopharmacology.* 2003;28(3):552–7.
96. Marcus RN, Owen R, Kamen L, Manos G, McQuade RD, Carson WH, Aman MG. A placebo-controlled, fixed-dose study of aripiprazole in children and adolescents with irritability associated with autistic disorder. *J Am Acad Child Adolesc Psychiatry.* 2009;48(11):1110–9.
97. Owen R, Sikich L, Marcus RN, Corey-Lisle P, Manos G, McQuade RD, Carson WH, Findling RL. Aripiprazole in the treatment of irritability in children and adolescents with autistic disorder. *Pediatrics.* 2009;124(6):1533–40.
98. Kent JM, Kushner S, Ning X, Karcher K, Ness S, Aman M, Singh J, Hough D. Risperidone dosing in children and adolescents with autistic disorder: a double-blind, placebo-controlled study. *J Autism Dev Disord.* 2013;43(8):1773–83.
99. Shea S, Turgay A, Carroll A, Schulz M, Orlik H, Smith I, Dunbar F. Risperidone in the treatment of disruptive behavioral symptoms in children with autistic and other pervasive developmental disorders. *Pediatrics.* 2004;114(5):e634–41.
100. European Medicines Agency (EMA). Guideline on the clinical development of medicinal products for the treatment of Autism Spectrum Disorder (ASD). Committee: Committee for Medicinal Products for Human Use (CHMP), Reference number: EMA/CHMP/598082/2013. Published on: 21/11/2017. Available at: [https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-development-medicinal-products-treatment-autism-spectrum-disorder-asd\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-development-medicinal-products-treatment-autism-spectrum-disorder-asd_en.pdf).
101. Potter LA, Scholze DA, Biag HMB, Schneider A, Chen Y, Nguyen DV, Rajaratnam A, Rivera SM, Dwyer PS, Tassone F, et al. A Randomized controlled trial of sertraline in young Children with autism spectrum disorder. *Front Psychiatry.* 2019;10:810.
102. Leucht S, Leucht C, Huhn M, Chaimani A, Mavridis D, Helfer B, Samara M, Rabaioli M, Bacher S, Cipriani A, et al. Sixty years of placebo-controlled antipsychotic drug trials in acute schizophrenia: systematic review, bayesian meta-analysis, and meta-regression of efficacy predictors. *Am J Psychiatry.* 2017;174(10):927–42.



103. Rødgaard E-M, Jensen K, Vergnes J-N, Soulières I, Mottron L. Temporal changes in effect sizes of studies comparing individuals with and without autism: a meta-analysis. *JAMA Psychiatry*. 2019;76(11):1124–32.
104. Huhn M, Nikolakopoulou A, Schneider-Thoma J, Krause M, Samara M, Peter N, Arndt T, Bäckers L, Rothe P, Cipriani A, et al. Comparative efficacy and tolerability of 32 oral antipsychotics for the acute treatment of adults with multi-episode schizophrenia: a systematic review and network meta-analysis. *Lancet*. 2019;394(10202):939–51.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

