Sixty Years of Placebo-Controlled Antipsychotic Drug Trials in Acute Schizophrenia: Systematic Review, Bayesian Meta-Analysis, and Meta-Regression of Efficacy Predictors

Stefan Leucht, M.D., Claudia Leucht, M.D., Maximilian Huhn, M.D., Anna Chaimani, Ph.D., Dimitris Mavridis, Ph.D., Bartosz Helfer, M.Sc., Myrto Samara, M.D., Matteo Rabaioli, Susanne Bächer, Andrea Cipriani, M.D., Ph.D., John R. Geddes, M.D., Georgia Salanti, Ph.D., John M. Davis, M.D.

Objective: Antipsychotic drug efficacy may have decreased over recent decades. The authors present a meta-analysis of all placebo-controlled trials in patients with acute exacerbations of schizophrenia, and they investigate which trial characteristics have changed over the years and which are moderators of drug-placebo efficacy differences.

Method: The search included multiple electronic databases. The outcomes were overall efficacy (primary outcome); responder and dropout rates; positive, negative, and depressive symptoms; quality of life; functioning; and major side effects. Potential moderators of efficacy were analyzed by meta-regression.

Results: The analysis included 167 double-blind randomized controlled trials with 28,102 mainly chronic participants. The standardized mean difference (SMD) for overall efficacy was 0.47 (95% credible interval 0.42, 0.51), but accounting for small-trial effects and publication bias reduced the SMD to 0.38. At least a "minimal" response occurred in 51% of the antipsychotic group versus 30% in the placebo group, and 23% versus 14% had a "good" response. Positive symptoms (SMD 0.45) improved more than negative symptoms (SMD 0.35) and depression (SMD 0.27). Quality of life (SMD 0.35) and functioning (SMD 0.34) improved even in the short term. Antipsychotics differed substantially in side effects. Of the response predictors analyzed, 16 trial characteristics changed over the decades. However, in a multivariable meta-regression, only industry sponsorship and increasing placebo response were significant moderators of effect sizes. Drug response remained stable over time.

Conclusions: Approximately twice as many patients improved with antipsychotics as with placebo, but only a minority experienced a good response. Effect sizes were reduced by industry sponsorship and increasing placebo response, not decreasing drug response. Drug development may benefit from smaller samples but better-selected patients.

Am J Psychiatry 2017; 174:927–942; doi: 10.1176/appi.ajp.2017.16121358

Antipsychotics are criticized by distinguished physicians (1, 2). This criticism makes the lay public (3), patients (4), general physicians, and policy makers skeptical. The skepticism is driven by the fact that most placebo-controlled studies are conducted by the pharmaceutical industry, which is not trusted (5, 6), an issue that we examine here. The essence of the argument is that antipsychotic drugs have multiple side effects, but only little efficacy, and that therefore their use should be restricted to a minimum (1).

Indeed, an early, large (463 participants) trial from 1964 sponsored by the National Institute of Mental Health (7), which is often used as a reference for antipsychotic drug efficacy, showed a substantial difference between antipsychotics and placebo (61% versus 22% of patients were much improved). In contrast, in recent years there have been a number of failed trials in which even standard drugs such as haloperidol did not outperform placebo (8). A systematic review and meta-regression suggested that an increasing placebo response rate could explain this phenomenon. However, it analyzed only predictors of placebo response and had little on the improvement with drug relative to placebo, which is crucial for patients (9). Drug response could well have increased in parallel to placebo response, so that the net effect would be the same. Therefore, there is a need to identify predictors of *drug-placebo differences* beyond the predictors of *placebo response* identified in this previous review (9). Another analysis suggested a parallel decrease in drug response, but it included only a small number of placebo-controlled trials, so that drug response was dominated by active-drug-controlled trials, which are very different from placebo-controlled trials. It is not plausible that drug response decreases when placebo response increases, making a

See related feature: **Clinical Guidance** *(Table of Contents)*

reassessment important (see Discussion section) (10). Finally, a previous network meta-analysis we conducted (11) did not provide a meta-regression of efficacy predictors. Moreover, it primarily examined newer second-generation antipsychotics (and only two of 52 old ones listed by the World Health Organization; http://www.whocc.no/), thus offering limited information on the first 40 years of antipsychotic drugs.

In this context, we present a comprehensive systematic review of all acute-phase placebo-controlled antipsychotic drug trials in schizophrenia since the introduction of chlorpromazine in 1953, addressing efficacy, tolerability, and quality of life and functioning, on which information is increasingly asked for. We explored with meta-regression which trial characteristics have changed over the years and which ones are moderators of drug-placebo differences. The results of this broad summary of the first 60 years of antipsychotic trials should inform clinicians, should provide clues for the future design of antipsychotic drug trials, and should help to put the debate about antipsychotic drugs on a rational basis.

METHOD

We followed the PRISMA guidelines (12) (checklist in Table S1 of the data supplement accompanying the online version of this article) and initially published a protocol in PROSPERO, an international prospective register of systematic reviews (CRD42013003342; Table S2 in the online data supplement).

Inclusion/Exclusion Criteria

Participants. Adults with acute exacerbations of schizophrenia or related disorders (following the Cochrane Schizophrenia Group, we accepted all diagnostic criteria and we also included schizoaffective, schizophreniform, and delusional disorder, because these do not require generally different treatment [13]). We excluded relapse prevention studies in stable patients receiving maintenance medication (14), studies of patients with predominant negative symptoms, and studies of patients with major concomitant somatic or psychiatric illness.

Interventions. We included all antipsychotics licensed in at least one country, except clozapine; it may be a more efficacious drug (11), and so pooling it with the other compounds would not have been appropriate (only one clozapine arm with nine patients had to be excluded on this basis [15], making the impact of this decision negligible). We excluded intramuscular formulations because these are used primarily either for emergency use (short-acting intramuscular drugs) or for relapse prevention (long-acting depot drugs). Both flexible- and fixed-dose studies were included. All flexible-dose studies were included because they allow investigators to titrate to an adequate dose. Of the fixed-dose studies we included only target-to-maximum doses according to the International Consensus Study of Antipsychotic Dosing (16). We averaged the results for eligible fixed doses in single studies with appropriate formulas before entering the studies in the meta-analysis (17).

Types of studies. We included published and unpublished double-blind, placebo-controlled, randomized controlled trials of at least 3 weeks' duration (18). Studies with a high risk in sequence generation or allocation concealment were excluded (17). We a priori excluded Chinese studies due to serious quality concerns (19, 20). Risk of bias was independently assessed by at least two reviewers (among C.L., S.L., M.H., B.H.) with the Cochrane Collaboration's risk-of-bias tool (17).

Search Strategy

We searched the Cochrane Schizophrenia Group Controlled Trials Register (compiled by regular systematic hand searches and searches of more than 15 databases, clinical trial registers, the Food and Drug Administration web site, and conference proceedings [21] without language restrictions; available to us until version August 2009) with the term "placebo," and we searched MEDLINE, EMBASE, PsychINFO, Cochrane CENTRAL, and ClinicalTrials.gov (last search October 2016; search terms presented in Table S3 in the online data supplement). These searches were supplemented by screening previous reviews (9, 11, 22–30).

Outcomes

1. Primary outcome: We examined the mean overall change in symptoms, as represented by the change in the total score on the Positive and Negative Syndrome Scale (PANSS; 31) or, if the PANSS change was not available, change in the score on the Brief Psychiatric Rating Scale (BPRS; 32). If the change value was not available for either scale, then the endpoint value of one of these scales was used, and if neither of these scales was used, then the score on another published schizophrenia rating scale was accepted (33).
2. Responders: We analyzed how many patients achieved a) at least a "minimal response," defined as either at least a 20% PANSS/BPRS reduction from baseline or a Clinical Global Impressions Scale (CGI) score indicating at least "slightly improved" (34–36), and b) a "good response"—either at least a 50% PANSS/BPRS reduction or a CGI rating of at least "much improved" (34–36). Results of the single definitions were consistent and were presented separately, as well. We also analyzed c) any study-defined definition.
3. Discontinuation: We calculated dropout rates related to discontinuation due to a) any cause and b) inefficacy.
4. Positive, negative, and depressive symptoms, quality of life, and social functioning: These outcomes were measured by means of published rating scales (33).
5. Major side effects: We recorded the presence of extrapyramidal side effects (antiparkinson medication use at least once), weight gain, sedation, prolactin increase, and QTc prolongation.

Study Selection and Data Extraction

At least two reviewers (among authors M.H., M.S., and S.L. and contributor M.T. [cited in acknowledgments]) independently selected potentially relevant publications from

the abstracts found by our search and made decisions about whether to include studies. At least two reviewers (among authors S.L., C.L., M.H., B.H., M.S., M.R., and S.B. and contributors M.K., P.R., T.A., and N.P. [cited in acknowledgments]) extracted data in duplicate in a spreadsheet. Disagreement was resolved by discussion. Missing data were requested from authors or the sponsoring pharmaceutical companies for all studies published in the last 30 years. When possible, we extracted intention-to-treat data, and we preferred data based on mixed-effect models of repeated measurements over last-observation-carried-forward data. For dichotomous data we assumed that participants lost to follow-up would not have responded (conservative approach). Missing standard deviations were estimated from test statistics or by using the mean standard deviation of the remaining studies (37, 38).

Statistical Synthesis of Study Results

We used a Bayesian hierarchical random-effects model in OpenBUGS 3.2.3 (http://www.openbugs.net/w/FrontPage) to estimate summary effect sizes for each outcome, as heterogeneity was expected. We primarily examined all antipsychotics as a group because efficacy differences between drugs are small (11, 39–41), but results of individual drugs are presented as well. For the primary analysis we merged the different antipsychotic arms within multiarm trials (17) but properly accounted for the inherent correlation in the drug-specific analyses. We estimated standardized mean differences (SMDs) for continuous outcomes and risk ratios for dichotomous outcomes, together with the 95% credible interval (CrI) for each. The numbers needed to treat to benefit and to harm were estimated by using the meta-analytic summary of an outcome in all placebo arms. Heterogeneity was assessed by visual inspection, the between-study standard deviation, and the I-square statistic (values >50% were considered as indicating considerable heterogeneity [42]).

Meta-Regression Analyses

We meta-regressed publication year and the frequency of the moderators to explore which trial characteristics have changed over time. Then, in meta-regressions of the primary outcome we were particularly interested in exploring whether the drug-placebo difference became smaller over the decades, and we systematically examined all possible moderators, reported previously (9, 10, 43–45), that might explain this phenomenon. We categorized the moderators into patient-, drug-, and study-design-related factors, although there were expected overlaps. Moderators that were significant in univariable analyses were included in a multivariable meta-regression model. To identify the most important moderators from this model we used the stochastic search variable selection algorithm to estimate the probability that each variable should be included in the meta-regression model (see protocol, Table S2 in the online data supplement [46]). To measure the strength of a moderator we compared the meta-regression models with the meta-analysis without covariates and estimated the percentage of heterogeneity explained by a moderator. Meta-regressions were not performed on individual drugs, because statistical power would have been insufficient for most of them. Post hoc analyses following recent research (e.g., by Agid et al. [9]) are noted in parentheses.

Patient-related factors. The patient-related factors were chronicity (9) as measured by the patient's age, duration of illness, duration of the current episode, and first-episode status (9, 44); the percentage of men (44); U.S. population versus not or mixed countries (47); degree of placebo response (9) and degree of drug response (10) (post hoc analyses following recent research), as measured by the PANSS change or by the BPRS change converted to PANSS by a validated method (48); severity at baseline, as measured by the PANSS total score (49); in- versus outpatient (9); and operationalized criteria (e.g., ICD-10 or DSM-III to DSM-IV-R) versus nonspecific "clinical diagnoses."

Drug-related factors. We classified the antipsychotics by their mechanisms according to the Neuroscience-Based Nomenclature (50); antipsychotic doses in chlorpromazine equivalents according to the International Consensus Study of Antipsychotic Dosing (10, 16); and fixed versus flexible doses (45).

Design-related factors. We analyzed the impact of risk of bias (appropriate versus unclear randomization [51] and allocation concealment methods [52], blinding [52], and missing outcome data [17, 53]); study duration (9); duration of washout (9); requirement of a scale-derived minimum level of symptoms at baseline (49); use of PANSS versus BPRS; sample size (54); number of sites (9); percentage of academic sites (9) (post hoc analysis following recent research); number of medications and arms (9); percentage of participants assigned to placebo (45); and drug company sponsorship of at least one study arm (medication donation alone was not considered company sponsorship [6]).

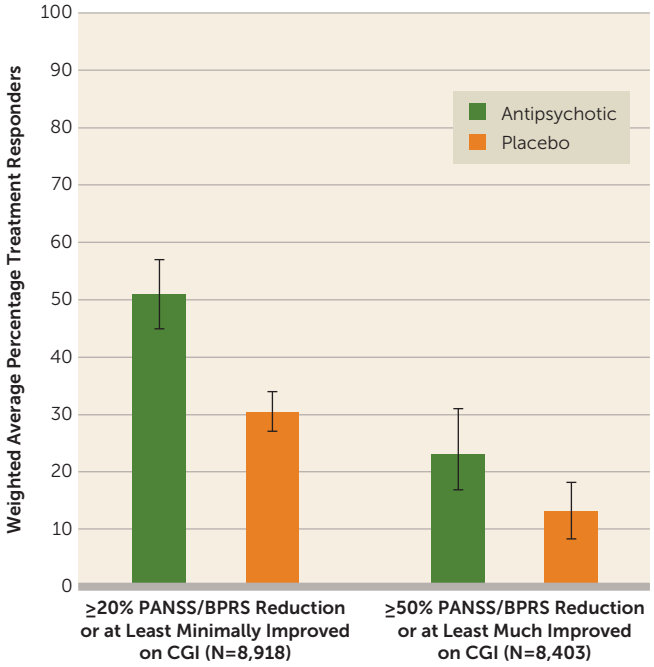
Sensitivity Analyses of Primary Outcome

We applied a fixed-effects instead of a random-effects model, we calculated odds ratios, and we excluded studies based on study completers. We explored whether the effect sizes of haloperidol, the only drug for which both early and recent studies were available, had decreased over the years as well (post hoc analysis).

Publication Bias

We used contour-enhanced funnel plots (55), a selection model (software OpenBUGS) (56), and the trim-and-fill method (57) to assess whether eventual small-trial effects were likely due to publication bias (Table S2 in the online data supplement presents details).

FIGURE 1. Proportions of Patients Taking Antipsychotics and Placebo Who Were at Least Minimally Improved and at Least Much Improved After Treatment[a]



[a] PANSS, Positive and Negative Syndrome Scale; BPRS, Brief Psychiatric Rating Scale; CGI, Clinical Global Impressions Scale improvement rating. Error bars represent 95% credible intervals.

RESULTS

Description of Included Studies

Figure S1 in the online data supplement presents the PRISMA (12) flow diagram. Overall, 167 studies published from 1955 to 2016 with 28,102 participants were included (see Table S4 in the online data supplement). The mean duration of illness was 13.4 (SD 4.7) years, the mean age was 38.7 (SD 5.5) years, and the median duration of studies with useable outcomes was 6 weeks (range 3–28 weeks; for the primary outcome all but one study lasted ≤12 weeks; one study without any useable outcomes lasted 156 weeks). There were no studies exclusively examining first-episode patients or treatment-resistant patients. The most frequently used drugs with data for at least one outcome were chlorpromazine (36 studies), haloperidol (28 studies), olanzapine (20 studies), risperidone (15 studies), quetiapine (eight studies), paliperidone (eight studies), aripiprazole (nine studies), thioridazine (seven studies), lurasidone (seven studies), asenapine (six studies), and loxapine (six studies); for all other drugs, fewer than five studies were available. Risk of bias is presented in supplemental Table S5. We included only randomized, double-blind trials, but the reports often did not indicate full details about sequence generation or allocation concealment. Descriptions of the methods and success of blinding were frequently insufficient as well. The data confirmed the high dropout rates in current schizophrenia studies (mean 37.2%, SD 20.5). Older studies were poorly reported, making it often impossible to extract outcome data (50% of the studies had a high risk of selective reporting). Finally, 70 studies (42%) were sponsored by the manufacturers of one antipsychotic included, 72 (43%) were not primarily industry sponsored, and in 25 (15%) of the studies the sponsor was unclear.

Outcome Results

The mean effect size of all studies combined was 0.47 (95% CrI 0.42, 0.51; I^2 52%; 105 studies with 22,741 participants). Patients treated with antipsychotics were twice as likely to respond as those on placebo when any response criterion was accepted (97 studies, N=20,690; response ratio 1.93, 95% CrI 1.72, 2.19); the number needed to treat to benefit (NNT) was 6 (95% CrI 5, 8; I^2 61%). At least a "minimal" response was experienced by 51% (95% CrI 45%–57%) of the antipsychotic-treated patients compared with 30% (27%–34%) on placebo (46 studies, N=8,918; response ratio 1.75, 95% CrI 1.59, 1.97; NNT 5, 95% CrI 4, 5), while 23% (95% CrI 16%–32%) versus 14% (95% CrI 10%–19%) had a "good" response (38 studies, N=8,403; response ratio 1.96, 95% CrI 1.65, 2.44; NNT 8, 95% CrI 6, 11) (Figure 1). Similar results were obtained when responder rates based on the PANSS/BPRS or CGI were analyzed separately (see online Table S6, which also presents odd ratios).

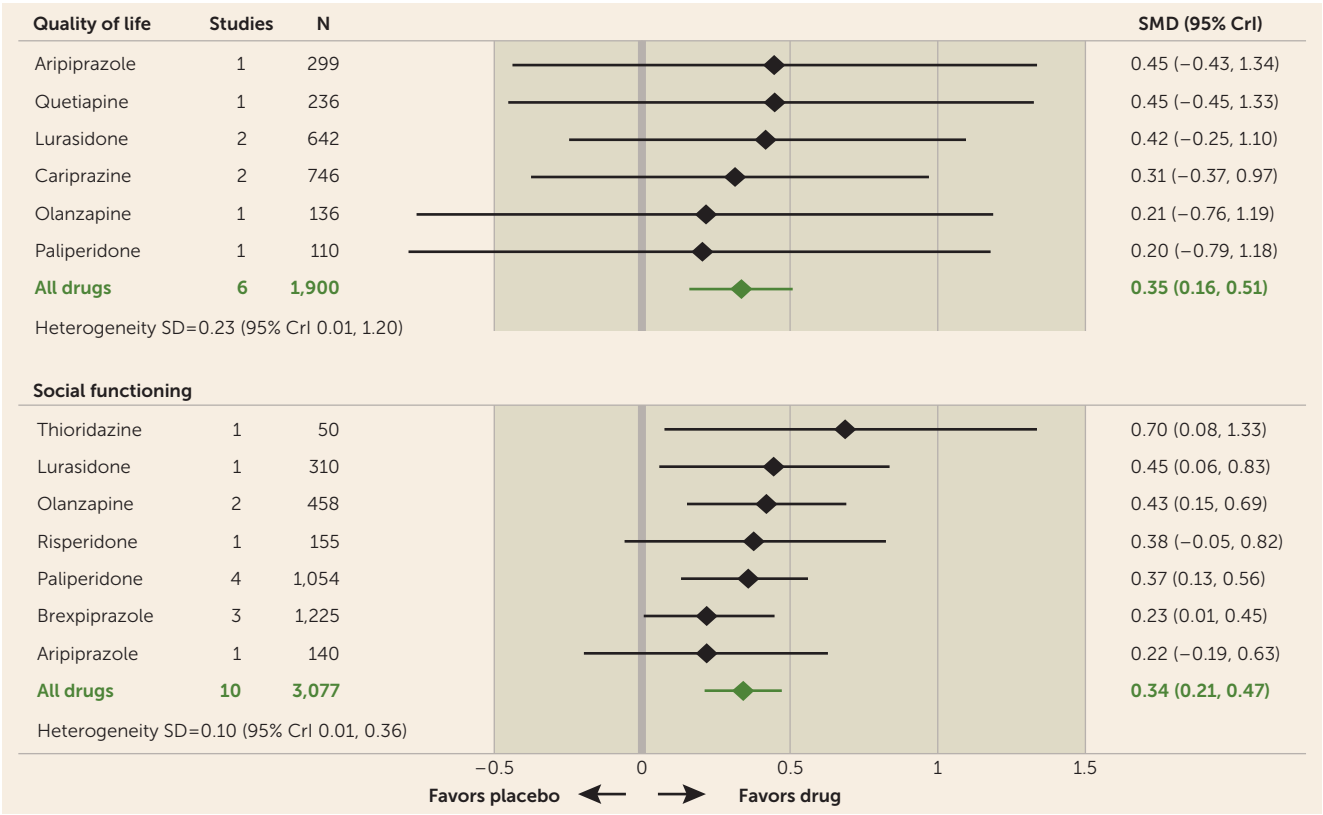
Participants receiving placebo were more likely to discontinue the studies prematurely, both for any reason (38% drug, 56% placebo; 105 studies, N=22,851; risk ratio 1.25, 95% CrI 1.20, 1.31; NNT 11, 95% CrI 9, 14; I^2 19%) and for inefficacy of treatment (13% drug, 26% placebo; 94 studies, N=23,017; risk ratio 2.09, 95% CrI 1.90, 2.32; NNT 7, 95% CrI 6, 9; I^2 46%).

The effect size for positive symptoms (64 studies, N=18,174; SMD 0.45, 95% CrI 0.40, 0.50; I^2 56%) was similar to that for overall symptoms, while effects on negative symptoms (69 studies, N=18,632; SMD 0.35, 95% CrI 0.31, 0.40; I^2 42%) and depression (33 studies, N=9,658; SMD=0.27, 95% CrI 0.20, 0.34; I^2 50%) were smaller.

As shown by six trials, the quality of life of participants in the antipsychotic groups was better than that in the placebo group (N=1,900; SMD 0.35, 95% CrI 0.16, 0.51; I^2 43%), and so were improvements in social functioning, as shown in 10 trials (N=3,077; SMD 0.34, 95% CrI 0.21, 0.47; I^2 46%) (Figure 2).

Antipsychotic drugs produced more movement disorders, as judged by use of antiparkinson medications (19% drug, 10% placebo; 63 studies, N=14,942; risk ratio 1.93, 95% CrI 1.65, 2.29; number needed to treat to harm [NNH] 12, 95% CrI 9, 16; I^2 51%); they were more sedating (14% drug versus 6% placebo; 86 studies, N=18,574; risk ratio 2.80, 95% CrI 2.30, 3.55; I^2 54%), led to more weight gain (59 studies, N=17,076; SMD −0.40, 95% CrI −0.47, −0.33; I^2 73%), led to more prolactin increase (51 studies; N=15,219; SMD −0.43, 95% CrI −0.55, −0.30; I^2 91%), and led to more QTc prolongation (29 studies, N=9,883; SMD −0.19, 95% CrI −0.29, −0.08; I^2 80%) than placebo. Results for individual drugs are presented in Figures 3 and 4 and in Figure S2 in the online data supplement.

FIGURE 2. Posttreatment Quality of Life and Social Functioning of Patients Taking Antipsychotics and Placebo[a]



[a] SMD, standardized mean difference; SD=standard deviation; CrI, credible interval. Some studies compared two antipsychotics with placebo. SMDs were obtained from a random-effects model assuming a common heterogeneity across all drugs.

Change of Trial Characteristics Over Time

Table 1 shows that several trial characteristics changed significantly over the years: the numbers of participants and sites, the use of minimum baseline severity as an inclusion criterion, fixed-dose designs, use of operationalized criteria, use of the PANSS, percentage of men, studies outside the United States, and placebo response increased, while the duration of the washout period, use of dopamine D_2 antagonists (50), study duration, risk of bias in terms of incomplete outcome data, mean doses, and number of academic sites decreased.

Moderators of Antipsychotic Efficacy: Univariable Analysis

Effect sizes have become smaller over the years. The coefficient of −0.08 in Table 2 indicates that a study published 10 years later than another one had, on average, a 0.08-unit lower effect size. Figure 5 demonstrates this effect not only for all antipsychotics as a class (Figure 5A) but also for haloperidol, the only antipsychotic for which both early and recent studies were available (Figure 5B). Moreover, Figures 5C and 5D show that the decrease of effect size was paralleled by an increase in placebo response, while drug response remained quite stable, which contradicts a previous analysis (10).

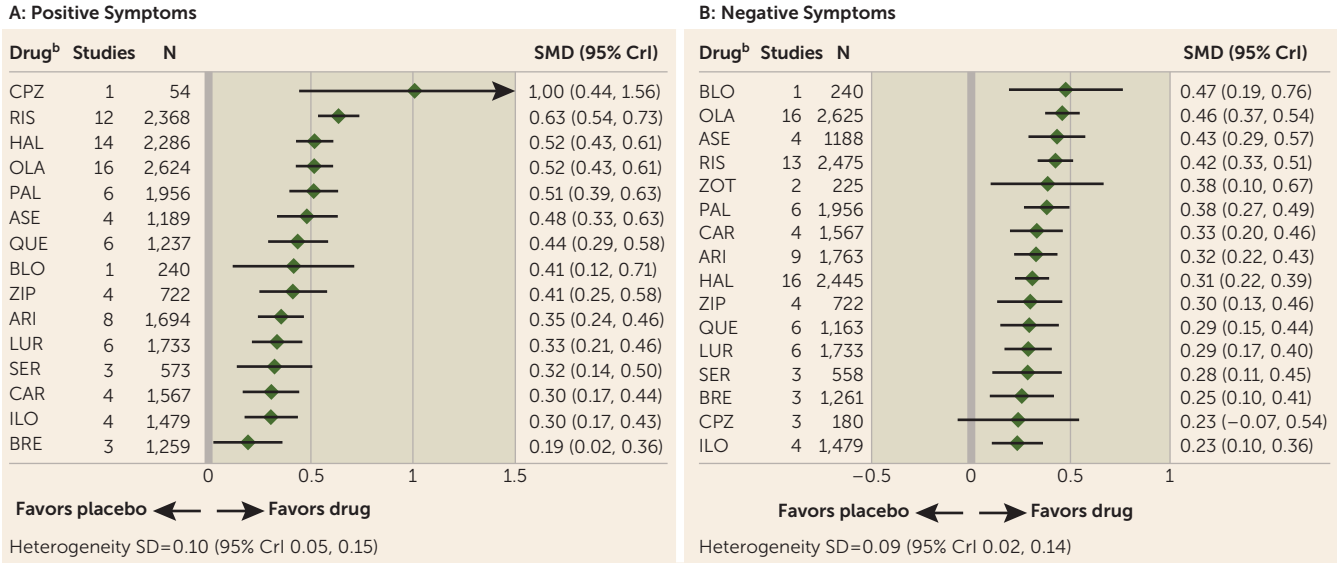
Significant factors involving study design were larger sample size (total numbers of participants and sites), number of drugs, PANSS rather than BPRS, a minimum symptom level as an entry criterion, and industry sponsorship. With the exception of the number of drugs, all these factors were associated with *smaller* effect sizes (Table 2).

Significant patient- or drug-related factors were operationalized rather than clinical diagnostic criteria, higher placebo response rates, lower dose, and D_2 and 5-HT_{1A} receptor partial agonists versus D_2 antagonists (mainly haloperidol), all of which were associated with *smaller* effect sizes (Table 2).

Moderators of Antipsychotic Efficacy: Multivariable Analysis

As several significant predictors are related by nature, we made the following choices for the multivariable model: 1) we chose sample size as representative of the number of sites, number of drugs, and number of arms, and 2) we chose publication year to represent the use of operationalized criteria (such criteria did not exist for early studies), use of the PANSS (introduced in 1987) versus BPRS, and drug mechanisms according to the Neuroscience-Based Nomenclature (50), which also changed over the years. Only pharmaceutical sponsorship and the degree of placebo response remained

FIGURE 3. Effect on Positive and Negative Symptoms of Single Antipsychotics Compared With Placebo[a]



[a] These are raw effect sizes that have not been corrected for the effects of increasing placebo response over the years. The effect sizes of the single drugs have not been compared with each other. Moreover, for some drugs few data were available, making the results unreliable. For example, the results of positive symptoms for chlorpromazine are based on only one study with 54 patients. This caused uncertainty about the true effect, which is expressed by a large 95% CrI. SMD, standardized mean difference; SD, standard deviation; CrI, credible interval.

[b] ARI, aripiprazole; ASE, asenapine; BLO, blonanserine; BRE, brexpiprazole; CAR, cariprazine; CPZ, chlorpromazine; HAL, haloperidol; ILO, iloperidone; LUR, lurasidone; OLA, olanzapine; PAL, paliperidone; QUE, quetiapine; RIS, risperidone; SER, sertindole; ZIP, ziprasidone; ZOT, zotepine.

significant, and they resulted in large probabilities from the variable selection algorithm (82.8% and 81.6%, respectively), implying that they are probably the most important moderators. Studies with a 10-PANSS-point larger placebo response, on average, had a 0.13 smaller effect size, and, surprisingly, industry-sponsored studies, on average, had a 0.16 smaller effect size compared with non-industry-sponsored trials (Table 3).

Both predictors remained the only significant ones in a post hoc sensitivity analysis where all significant moderators were entered. When pharmaceutical sponsorship—which is probably a composite of various factors—was removed from the model in another sensitivity analysis, only degree of placebo response remained significant, demonstrating the strength of this factor. Both sensitivity analyses had less explanatory power than the primary analysis, however (31.3% heterogeneity explained in the primary model versus 18.8% in both sensitivity analyses; see Table S7 in the online data supplement).

Publication Bias

Contour-enhanced funnel plots revealed small-trial effects. As studies were missing in the area of nonsignificant effect sizes (Figure 6) and as the selection model showed a strong correlation between probability of publication and magnitude of effect in various scenarios (range of correlation coefficients, R=0.66–0.85), part of the small-trial effects is likely a result of publication bias. The publication bias "adjusted" SMD ranged between 0.36 and 0.41 in various scenarios of the selection model, corroborated by the trim-and-fill method (adjusted SMD 0.38, 95% CrI 0.33, 0.43; see online Table S2).

Sensitivity Analyses

The use of a fixed-effects rather than a random-effects model (105 studies, N=22,741; SMD=0.44, 95% CrI 0.42, 0.47) and the exclusion of completer analyses (95 studies, N=22,352; SMD=0.46, 95% CrI 0.42, 0.51) did not significantly change the primary outcome, nor did the use of odds ratios (Table S6 in online data supplement).

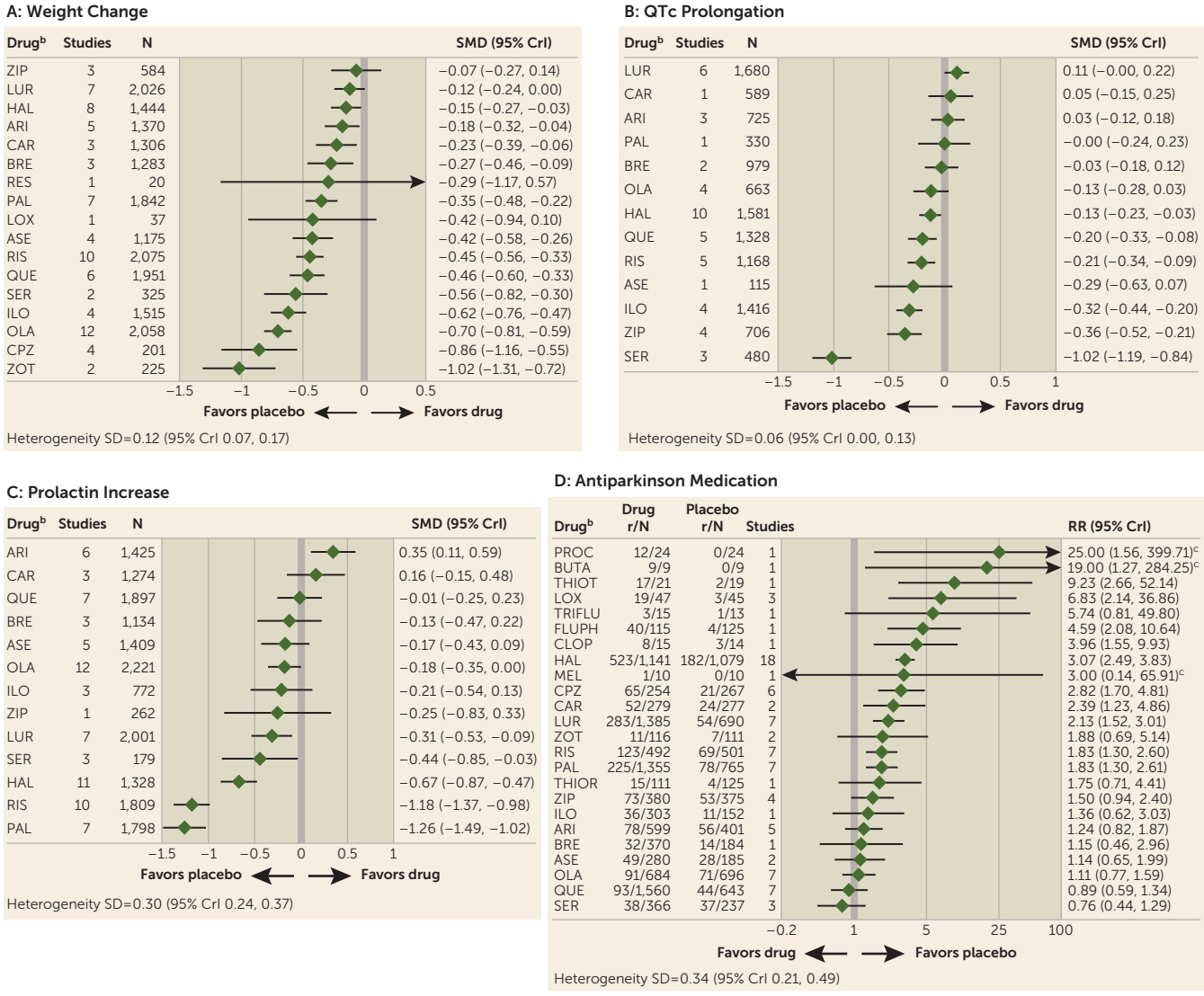
DISCUSSION

We believe this to be the first comprehensive meta-analysis of all acute-phase, placebo-controlled antipsychotic drug trials since the introduction of chlorpromazine. The multiple new analyses and results can be important for clinical practice and trial methodology.

Overall Efficacy

We examined two response criteria—"any" response and a "good" response to antipsychotics. This was important because previous systematic reviews (23–29) analyzed whatever response criterion was presented in the individual studies, leading to a difficult-to-interpret criteria mix. Several analyses showed that a 20% PANSS/BPRS reduction roughly corresponded to minimal improvement on the CGI and a 50% PANSS/BPRS reduction corresponded to much improvement, and they justified analyzing them together (35, 58–60) (results of individual scales were similar, as shown in

FIGURE 4. Effect on Weight Gain, QTc Prolongation, Prolactin, and Antiparkinson Medication of Single Antipsychotics Compared With Placebo[a]



[a] For some drugs few data were available, making the results unreliable. For example, the results for weight gain with reserpine are based on only one study with 20 patients. This caused uncertainty about the true effect, which is expressed by a large 95% CrI. The effect sizes of the single drugs have not been compared with each other, but 95% CrIs that do not overlap with the y-axis mean statistically significant differences compared with placebo. r, number of participants with an event; RR, relative risk; SMD, standardized mean difference; SD, standard deviation; CrI, credible interval.

[b] ARI, aripiprazole; ASE, asenapine; BRE, brexpiprazole; BUTA, butaperazine; CAR, cariprazine; CLOP, clopenthixol; CPZ, chlorpromazine; FLUPH, fluphenazine; HAL, haloperidol; ILO, iloperidone; LOX, loxapine; LUR, lurasidone; MEL, melperone; OLA, olanzapine; PAL, paliperidone; PROC, prochlorpromazine; QUE, quetiapine; RES, reserpine; RIS, risperidone; SER, sertindole; THIOT, thiothixene; THIOR, thioridazine; TRIFLU, trifluperazine; ZIP, ziprasidone; ZOT, zotepine.

[c] This relative risk was obtained after a continuity correction and from a fixed-effect model.

Table S6). Antipsychotic drugs clearly increased the number of patients with "any" response (51% versus 30%), but, importantly, few patients (23% versus 14%) reached a "good" response within the confines of short-term double-blind trials. The mean effect size for overall symptoms (0.47) was only medium according to Cohen (61), and it translates to a difference of 9.6 PANSS points. This contrasts with the large (N=463) early NIMH study from 1964 that has been used frequently as a benchmark for antipsychotic drug efficacy (7). Its impressive difference in response rates (61% under drug versus 22% under placebo showed much improvement) can be explained by the fact that approximately 50% of patients suffered from their first episode or were antipsychotic naive (7). In the current review not a single study was restricted to first-episode patients. Thus, its results are representative only for chronically ill, often previously treated patients, who respond less well to antipsychotics (62). In the future, first-episode trials could provide better signal detection. Moreover, the trials in a meta-analysis are weighted by their sample size. Thus, the mean effect size of 0.47 largely represents the

TABLE 1. Meta-Regressions Showing Which Antipsychotic Trial Characteristics Have Changed Over Time

Explanatory Variable	Weighted Mean Publication Year	Number of Studies	Meta-Regression	
			Coefficient	95% CrI
Study design factors				
Number of total participants[a]		105	79.77[b]	58.50, 101.03
Number of sites[a]		96	12.22[b]	9.57, 14.91
Academic sites (%)[a]		59	−13.75[b]	−19.75, −7.74
Baseline severity entry minimum score[a]				
No (reference)	1988	29		
Yes	2008	73	2.52[c]	1.18, 5.38
Duration of washout period (days)[a]		89	−9.20[b]	−11.78, −6.62
Study duration (weeks)[a]		96	−0.92[b]	−1.33, −0.50
Randomization				
Low risk (reference)	2007	48		
Unclear	2006	57	0.80[c]	0.53, 1.20
Allocation concealment				
Low risk (reference)	2008	33		
Unclear	2006	72	0.76[c]	0.48, 1.21
Intention-to-treat analysis or completers[a]				
Intention-to-treat analysis (reference)	2007	95		
Completers	1981	7	0.21[c]	0.11, 0.39
Risk of bias due to missing outcome data[a]				
Low risk (reference)	2008	73		
Unclear	2002	19	0.50[c]	0.32, 0.81
High risk	2000	13	0.47[c]	0.27, 0.82
Blinding				
Low risk (reference)	2007	57		
Unclear	2006	48	0.80[c]	0.52, 1.23
High risk	—	—	—	
Number of arms				
Two arms (reference)	2006	10		
More than two arms	2007	95	1.11[c]	0.74, 1.65
Number of medications				
Two medications (reference)	2009	33		
More than two medications	2005	72	0.53[c]	0.27, 1.02
Industry-sponsored drug or not				
Nonsponsored drugs (reference)	2006	32		
At least one sponsored drug	2007	65	1.15[c]	0.73, 1.81
Percentage patients randomized to placebo		105	0.00[b]	−0.04, 0.03
Scale[a]				
PANSS (reference)	2009	68		
BPRS	1990	33	0.02[c]	0.00, 0.07
Drug-related factors				
Drug mechanism[a,d]				
M1 (reference)	1998	18		
M2 versus M1	1999	47	0.49[c]	0.26, 0.92
M3 versus M1	2012	12	14.79[c]	2.74, 79.93
M4 versus M1	2008	17	4.40[c]	1.17, 16.51
M5 versus M1	2007	7	4.79[c]	0.65, 37.17
Fixed or flexible dose[a]				
Fixed dose (reference)	2008	79		
Flexible dose	1997	26	0.38[c]	0.24, 0.60
Mean dose (chlorpromazine equivalents)[a]		91	−86.95[b]	−121.18, −52.71

continued

TABLE 1, *continued*

Explanatory Variable	Weighted Mean Publication Year	Number of Studies	Meta-Regression	
			Coefficient	95% CrI
Patient-related factors				
Percentage men[a]		91	6.81[b]	3.81, 9.82
Operationalized criteria or not[a]				
Operationalized (reference)	2007	88		
Not operationalized	1977	16	0.07[c]	0.02, 0.20
Country[a]				
United States (reference)	2002	45		
Other or mixed	2008	60	2.32[c]	1.37, 3.93
Placebo response[a]		99	2.74[b]	1.60, 3.88
Drug response		100	0.27[b]	−0.95, 1.49
Average age		100	0.64[b]	−0.08, 1.37
Duration of illness		60	0.66[b]	−0.21, 1.53
Baseline severity (PANSS total score)		85	−0.48[b]	−1.57, 0.62

[a] This characteristic resulted in a statistically significant association with publication year.
[b] This coefficient shows the average increase or decrease for the respective moderator associated with a 10-year increase in publication year. For example, a study that is 10 years newer would on average have 79.77 more participants.
[c] This coefficient shows the average odds ratio of the respective moderator associated with a 10-year increase in publication year. For example, a study that is 10 years newer would on average have 2.52 times the odds of having a baseline severity entry minimum score.
[d] M1–M5 are drug mechanisms of action according to the Neuroscience-Based Nomenclature (NbN) (50). M1: receptor antagonists (D_2) clopenthixol, fluphenazine, haloperidol, perphenzine, pimozide, pipotiazine, sulpiride, and trifluoperazine. M2: receptor antagonists (D_2, 5-HT_2) chlorpromazine, iloperidone, loxapine, lurasidone, olanzapine, sertindole, thioridazine, ziprasidone, and zotepine. M3: receptor partial agonists (D_2, 5-HT_{1A}) aripiprazole, brexpiprazole, and cariprazine. M4: receptor antagonists (D_2, 5-HT_2, NE, α_2) asenapine, paliperidone, and risperidone. M5: receptor antagonist (D_2, 5-HT_2) and reuptake inhibitor (NET) quetiapine. A few old drugs have not been classified by NbN yet.

effects of the avalanche of trials after the reintroduction of clozapine in 1990. In this context we caution that the efficacy effect sizes of the single drugs in Figure 3 and Figure S2 have not been adjusted for publication year.

Negative Symptoms and Depression

A recent meta-analysis found an almost two times higher superiority compared with placebo (SMD 0.58 versus 0.35 here), but, mistakenly, standard errors rather than standard deviations were often used in the calculation of effect sizes, which artificially inflated them (63). In our analysis of antipsychotics, the effect size for negative symptoms was smaller, but the range was similar to that for positive symptoms. However, as studies with primary negative symptoms were excluded, the effect size might mainly reflect reductions of secondary negative symptoms (64). Similarly, the small effect of antipsychotics on depressive symptoms might also be a consequence of the reduction of positive symptoms and associated psychological distress. Nevertheless, some second-generation antipsychotics have proven efficacy in major depressive disorder (65).

Side Effects

The only purpose of our side effect analysis was to present a brief measure of the efficacy-tolerability trade-off in this class review. Effect sizes across drugs were medium with sometimes very large heterogeneity. This heterogeneity reflects the enormous differences of single antipsychotics in their side effects, and it suggests that careful choices can minimize the side effect burden for individual patients. But we caution that sometimes small participant numbers, which make results unreliable (reflected by large 95% credible intervals), must be considered in interpreting Figure 4 and supplemental Figure S2.

Outcomes Related to Social Integration

The results suggest a small to medium benefit of antipsychotics in quality of life even in the short term. As in our systematic review on maintenance treatment with antipsychotics compared with placebo (14), only six recent trials have investigated this crucial outcome, which combines efficacy and safety and which might be more relevant for patients that the mere reduction of hallucinations and delusions, but with a sample size of 1,900 patients the results are robust (66). The same holds true for social functioning. More time may be needed until antipsychotics develop their full effects, but patients may want to know whether after approximately 6 weeks they are already doing better in this regard. Outcomes that help to understand whether antipsychotics also help social reintegration should become a standard (Figure 2).

Meta-Regression of Response Predictors Including Industry Sponsorship

Table 1 shows that several study characteristics changed over the decades, and some of them were also significant predictors of drug-placebo differences.

In univariable analyses, drug-placebo differences decreased over time, with an average rate of 0.08 effect size units per decade, signifying that a study from 1970 would have

TABLE 2. Univariable Meta-Regressions for Moderators of Antipsychotic Efficacy

Explanatory Variable	Univariable Meta-Regression		Coefficient Corresponds to[a]	Number of Studies	N
	Coefficient	95% CrI			
Study design factors					
Publication year	−0.08[c]	−0.12, −0.04	10-year increase	105	22,741
Number of total participants	−0.04[c]	−0.06, −0.01	100 participants more	105	22,741
Number of sites	−0.02[c]	−0.04, 0.00	10-site increase	96	20,941
Number of medications	0.08[c]	0.02, 0.15	1 drug more	105	22,741
Baseline severity entry minimum score	−0.17[c]	−0.29, −0.04	Minimum entry score	102	22,291
Industry sponsored drug or not	−0.15[c]	−0.25, −0.05	Sponsored	97	22,397
Scale (PANSS or BPRS)	0.18[c]	0.07, 0.30	BPRS	101	22,589
Risk of bias due to missing outcome data	0.05	−0.03, 0.12	Unclear or high risk	105	22,741
Percentage of academic sites	0.01	−0.01, 0.03	10% increase	59	9,379
Number of arms	0.00	−0.04, 0.04	1 arm increase	105	22,741
Minimum duration of washout phase	0.03	−0.01, 0.06	10-day increase	89	18,586
Percentage randomized to placebo	0.01	−0.05, 0.07	10% increase	105	22,741
Study duration	0.10	−0.10, 0.29	10-week increase	96	22,443
Blinding	0.00	−0.09, 0.09	Unclear or high risk	105	22,741
Allocation concealment	0.01	−0.08, 0.11	Unclear risk	105	22,741
Randomization	−0.03	−0.13, 0.06	Unclear risk	105	22,741
Drug-related factors[d]					
Drug mechanism M2 versus M1	−0.13	−0.28, 0.01	M2	101	22,315[e]
Drug mechanism M3 versus M1	−0.26[c]	−0.43, −0.09	M3	101	22,315[e]
Drug mechanism M4 versus M1	−0.11	−0.28, 0.05	M4	101	22,315[e]
Drug mechanism M5 versus M1	−0.18	−0.39, 0.03	M5	101	22,315[e]
Fixed or flexible dose	0.04	−0.08, 0.17	Flexible dose	105	22,741
Mean dose	0.03[c]	0.00, 0.05	100-CPZ-unit increase	91	19,957
Patient-related factors					
Operationalized criteria or not	0.22[c]	0.04, 0.40	No operationalized criteria	103	22,151
Placebo response (mean PANSS change score in placebo arm)	−0.15[c]	−0.21, −0.09	10-unit PANSS increase	99	22,520
Drug response (mean change score in drug arm)	0.05	−0.02, 0.12	10-unit PANSS increase	100	22,564
Average age	−0.08	−0.20, 0.03	10-year increase	100	22,567
Baseline severity score	0.10	−0.01, 0.20	10-unit PANSS increase	85	21,259
Duration of illness	−0.07	−0.23, 0.08	10-year increase	60	14,278
Percentage of men	−0.01	−0.04, 0.02	10% increase	91	21,119
Country	0.02	−0.07, 0.12	Non-U.S. or mixed study	105	22,741
First episode[f]					
Duration of current episode[f]					
In- or outpatients at study start[f]					

[a] For example, a 10-year increase in publication year on average reduces the SMD by 0.08 unit. CPZ, chlorpromazine.
[b] Standard mean difference (SMD) after adjustment for covariate. For example, after adjustment for publication year, a study published in 2000 would on average have a standardized mean difference of 0.50, or nonsponsored studies would have an average SMD of 0.57.
[c] Statistically significant moderator.
[d] M1–M5 are drug mechanisms of action according to the Neuroscience-Based Nomenclature (NbN) (50). M1: receptor antagonists (D_2) clopenthixol, fluphenazine, haloperidol, perphenzine, pimozide, pipotiazine, sulpiride, and trifluoperazine. M2: receptor antagonists (D_2, 5-HT_2) chlorpromazine, iloperidone, loxapine, lurasidone, olanzapine, sertindole, thioridazine, ziprasidone, and zotepine. M3: receptor partial agonists (D_2, 5-HT_{1A}) aripiprazole, brexpiprazole, and cariprazine. M4: receptor antagonists (D_2, 5-HT_2, NE, α_2) asenapine, paliperidone, and risperidone. M5: receptor antagonist (D_2, 5-HT_2) and reuptake inhibitor (NET) quetiapine. A few old drugs have not been classified by NbN yet.
[e] Number for the overall model.
[f] Not enough data were available for the variable number of patients with a first episode, and there were too few data for duration of the current episode. The vast majority of studies included only inpatients. Therefore, these variables could not be analyzed in a meaningful way.

an effect size of approximately 0.74 and a study from 2015 an effect size of 0.38, a trend that has not been stopped by the most recent antipsychotics brexpiprazole and cariprazine (Figure 3 and supplemental Figure S2). Publication year can only be a surrogate for other factors, but it is not surprising that trials without standardized diagnostic criteria or using the BPRS had larger effect sizes than trials with standardized criteria or using the PANSS, nor that the old D_2 antagonists had larger effect sizes than the recent D_2/5-HT_{1a} partial agonists. In the early studies, standardized criteria and the PANSS were simply not available, and D_2 antagonists were primarily examined in the older, smaller trials. When we analyzed haloperidol separately, its superiority compared with placebo had become smaller over the years as well

TABLE 2, *continued*

SMD at Moderator Mean Value or Reference Category[b]		Moderator Mean Value or Reference Category	Heterogeneity SD		% Heterogeneity Explained
SMD	95% CrI		SD	95% CrI	
0.50	0.45, 0.55	2000	0.14	0.10, 0.19	12.5
0.49	0.44, 0.54	225	0.15	0.11, 0.20	6.3
0.49	0.44, 0.55	28	0.17	0.12, 0.22	—
0.40	0.33, 0.47	2 drugs	0.15	0.11, 0.20	6.3
0.61	0.50, 0.72	Without entry score	0.15	0.11, 0.20	6.3
0.57	0.48, 0.66	Nonsponsored	0.14	0.10, 0.19	12.5
0.43	0.38, 0.48	PANSS	0.15	0.11, 0.20	6.3
0.45	0.40, 0.50	Low risk	0.16	0.12, 0.21	0
0.57	0.51, 0.64	58%	0.15	0.08, 0.23	6.3
0.47	0.38, 0.57	2 arms	0.16	0.12, 0.21	0
0.50	0.45, 0.55	9 days	0.15	0.11, 0.20	6.3
0.47	0.42, 0.52	28.3%	0.16	0.12, 0.21	0
0.46	0.42, 0.51	6.5 weeks	0.16	0.11, 0.21	0
0.47	0.41, 0.53	Low risk	0.16	0.12, 0.21	0
0.46	0.39, 0.53	Low risk	0.16	0.12, 0.21	0
0.48	0.42, 0.55	Low risk	0.16	0.12, 0.21	0
0.60	0.48, 0.73	M1	0.16[b]	0.12, 0.21	0
0.60	0.48, 0.73	M1	0.16[b]	0.12, 0.21	0
0.60	0.48, 0.73	M1	0.16[b]	0.12, 0.21	0
0.60	0.48, 0.73	M1	0.16[b]	0.12, 0.21	0
0.46	0.41, 0.51	Fixed dose	0.16[b]	0.12, 0.21	0
0.49	0.45, 0.54	580.6 CPZ units	0.15	0.11, 0.20	6.3
0.45	0.41, 0.50	Operationalized criteria	0.16	0.11, 0.20	0
0.48	0.44, 0.52	6.24 units	0.13	0.08, 0.18	18.8
0.46	0.42, 0.51	17.45	0.16	0.11, 0.21	0
0.47	0.42, 0.51	38	0.16	0.11, 0.20	0
0.45	0.41, 0.50	94.6 units	0.16	0.11, 0.21	0
0.47	0.42, 0.53	14 years	0.15	0.09, 0.21	6.3
0.46	0.41, 0.51	66.3%	0.16	0.12, 0.21	0
0.45	0.38, 0.53	U.S.	0.16	0.12, 0.21	0

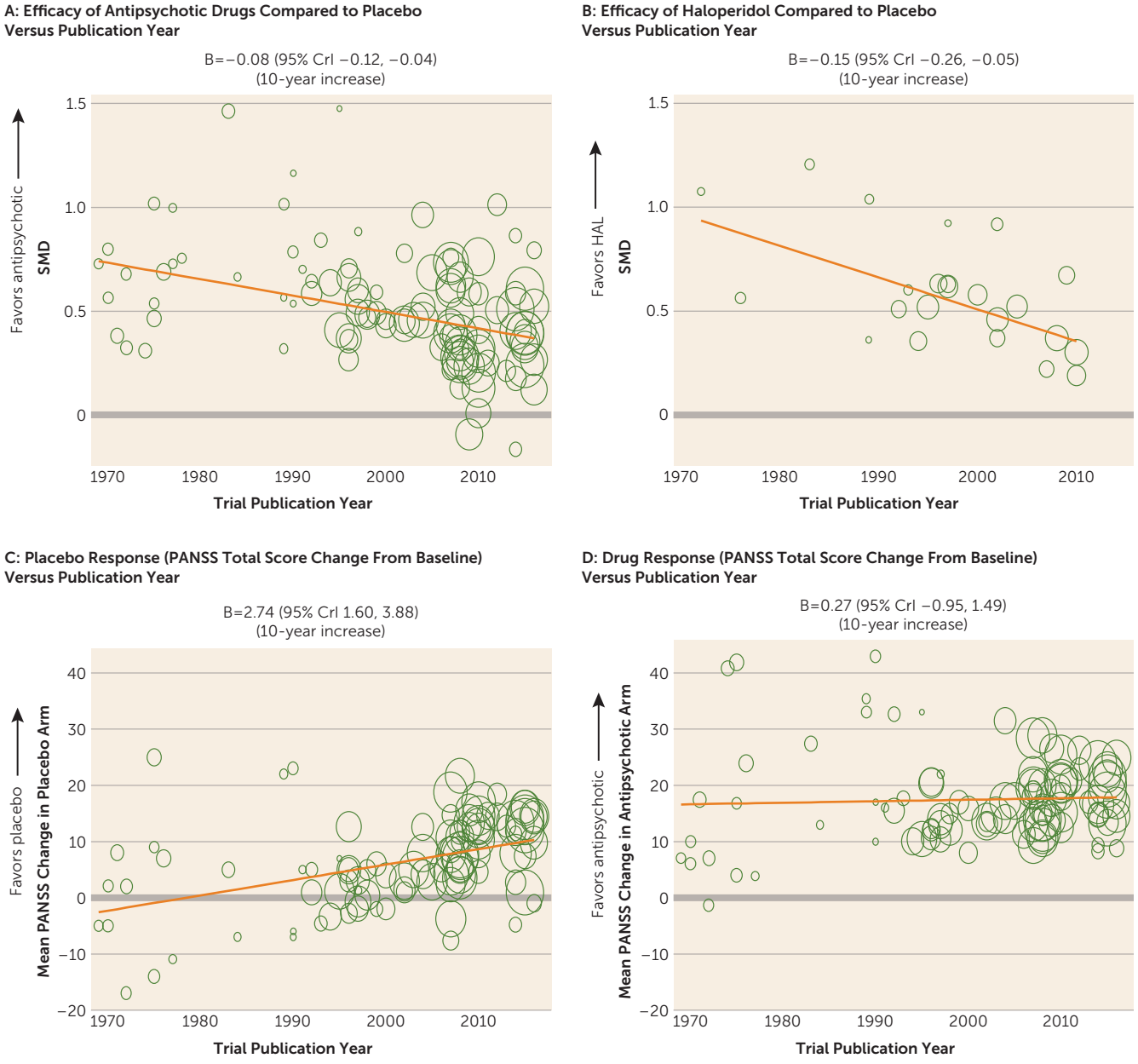
(Figure 5B), demonstrating that decreasing effect sizes over time cannot be explained solely by more recent drugs being less efficacious.

Larger sample sizes and a related moderator, number of sites, were associated with smaller effect sizes, which is consistent with the funnel plot suggesting substantial small-trial effects, which are well known from other medical fields (54, 67). The patients in small trials might be better selected than those in large trials. In contrast, the methodology of the often older, small trials was less stringent. For example, independent monitoring is a relatively recent requirement. However, our specific tests suggest that studies were missing, at least in part, because of publication bias.

The purpose of minimum baseline severity thresholds is to have drug-responsive populations (49), but, counterintuitively, studies with severity thresholds had lower effect sizes, possibly because such criteria invited artificial baseline inflation. Although the direction of the effect changed in the multivariable analysis, suggesting that this moderator might be confounded by another one, alternative ways to have severely ill populations should be considered.

However, in the multivariable meta-regression, the only moderators that remained significant were the degree of placebo response and industry sponsorship, which, contrary to criticisms and our expectations (6), was associated with *smaller* effect sizes. Industry studies are often large and involve multiple countries and sites, leading to problems such as cultural differences in the interpretation of psychopathology, which may increase variability and decrease effect sizes. The "patent clock" is running down, thus patients are recruited quickly by professional centers. As multiple

FIGURE 5. Effect Sizes Over Time for Efficacy of Antipsychotic Drugs[a]



[a] SMD=standardized mean difference; B=regression coefficient.

effective antipsychotics are available, patients think twice before consenting to a placebo-controlled trial and those who do consent can create negative selection; examples include (partial) nonresponders to previous drugs or so-called "professional patients" who benefit from a free trial of medication by answering a newspaper advertisement. These factors may also contribute to high placebo response.

Differences From Previous Analyses

Agid and colleagues (9) focused on predictors of *placebo response*, while our research question was about the *drug-placebo difference*. To explain placebo response was important from a methodological point of view, but for patients and psychiatrists it is the drug-placebo difference that counts. In this context it was by no means self-explanatory that—together with pharmaceutical sponsorship—placebo response was the strongest predictor of efficacy effect sizes in our analysis. Some other factor could have well been more important, and a parallel increase of drug response, which would have attenuated placebo response, was a priori likely. Only a few significant predictors of placebo response in the study of Agid et al. (9) were also significant predictors of drug-placebo differences here, at least in univariable analyses (publication year and the number of sites). In addition to the different research question, another possible explanation is that our database was two times larger

TABLE 3. Multivariable Meta-Regression Model for Moderators of Antipsychotic Efficacy[a]

	Multivariable Meta-Regression				
Moderator	Coefficient	95% CrI	Coefficient Corresponds to	Interpretation	Probability (%)[b]
Placebo response	−0.13[c]	−0.20, −0.06	10-unit increase	10-point higher mean PANSS change score in placebo arm would reduce SMD on average by 0.13 unit	80.6
Industry sponsored or not	−0.16[c]	−0.28, −0.04	Industry sponsored	SMD for studies including at least one sponsored drug would be on average 0.16 unit smaller than nonsponsored studies	82.8
Publication year	−0.02	−0.09, 0.05	10-year increase	Study published 10 years later would have on average 0.02-unit smaller SMD	25.0
Sample size	0.01	−0.02, 0.04	100-participant increase	Study with 100 more participants would have on average 0.01-unit larger SMD	3.3
Mean dose	0.01	−0.03, 0.04	100-CPZ-unit increase	Mean dose 100 CPZ units higher would increase SMD on average by 0.01 unit	3.3
Baseline severity minimum score	0.05	−0.13, 0.21	Baseline severity minimum score	SMD for studies having minimum baseline severity entry score would be on average 0.05 unit larger than that for studies without minimum baseline severity score	48.4

[a] Summary of the model: 78 studies with 19,060 participants; heterogeneity SD 0.11, 95% CrI 0.07, 0.16; the model explained 31.3% of the heterogeneity. SMD, standard mean difference; CPZ, chlorpromazine.
[b] In a simulation process, this is the probability that a model that includes this moderator would have been selected as the preferred model.
[c] Statistically significant moderator.

(105 randomized controlled trials versus 50 in the study by Agid et al.), which made our results more robust.

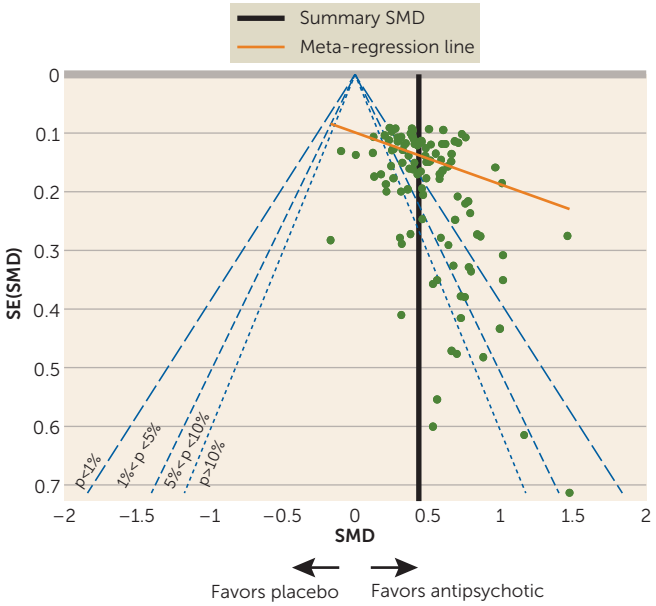
Rutherford and colleagues (10) addressed drug-placebo differences, but four out of six of the findings that they emphasized in their abstract were not confirmed by our multivariable analysis. They used multilevel meta-analysis and hierarchical modeling, which is another valid method.

It is unclear how much this difference in methods accounted for trial duration and severity, findings emphasized in the abstract by Rutherford et al. (10), not being significant moderators in our analysis. But the fact that we had 2.5 times more placebo-controlled studies available could have changed many findings. The major difference in results was that in the Rutherford et al. analysis (10) increasing response in the placebo arms was paralleled by *decreasing* response in the drug arms. In our analysis, drug response remained stable over the years (Figure 5D). In the study by Rutherford et al. (10), many of the data on drug response were from trials that compared drugs head to head (208 arms) and were not placebo-controlled trials (39 arms) (68). Although trial type was statistically controlled for, this is a quite different population of trials. For example, dropout rates are much higher in placebo-controlled trials than in active-controlled trials (69). When we reanalyzed the drug arms in the 39 placebo-controlled studies of Rutherford et al., drug response remained stable (online supplemental Table S8). Thus, drug response has decreased only in *active*-controlled studies, not in placebo-controlled studies. This has major implications for drug development: To improve signal detection in placebo-controlled trials, researchers need to focus on reducing placebo response rather than on increasing drug response. Finally, Agid et al. (9) did not detect publication bias and Rutherford et al. (10) did not explore it.

Limitations

The major limitation is that all antipsychotics were analyzed as a class, because efficacy differences between individual drugs are thought to be small (except clozapine, for which one trial with only nine patients was excluded) (11, 70). The number of drugs involved rendered it impossible to fully control for the resulting heterogeneity. Additionally, many older studies were so poorly reported that it was impossible to extract outcome data. For example, two early large Veterans Affairs studies (312 and 692 patients, respectively) showed a significant superiority of antipsychotics compared with placebo, but an effect size could not be calculated (71, 72). In one of them the difference in response rates was as high as in the NIMH study (7) (51% responded to antipsychotics versus 8% to placebo) (71). Only 46 and 38 studies, respectively, reported on the number of participants with at least "minimal response" and "good response." It is possible that some authors presented response data based on the cutoff showing the best result. Finally, conventional meta-analyses cannot detect subtle moderators of treatment effects. The main reason is ecological fallacy, i.e., conclusions about individuals

FIGURE 6. Contour-Enhanced Funnel Plot for Correlation Between Probability of Publication and Magnitude of Effect for Antipsychotic Drugs[a]



[a] SE=standard error; SMD=standardized mean difference.

that are based on analyses of group data can be biased. Another one is limited variability in the observed means, which could be overcome by meta-analyses of individual patient data, which capture large interindividual variability.

Conclusions

Our results are important on several levels.

First, clinicians can expect that approximately two times more patients improve when treated with antipsychotics compared with placebo, but only a minority will experience a good response in the short term. We need to document better whether antipsychotics only suppress positive symptoms or whether they also help social reintegration, reflected by improvements in social functioning and quality of life.

Second, network meta-analyses need to consider possible temporal changes. If placebo-controlled trials on one drug developed in the 1970s are combined with those for a drug developed in the 2010s, the older drug might artificially turn out better owing to higher effect sizes in that period. In a previous report we therefore excluded placebo-controlled trials in a sensitivity analysis and examined publication year as a moderator (11).

Third, industry sponsorship has not inflated effect sizes. But there was publication bias, because companies do not always publish inconclusive studies. Increasing placebo response, but not decreasing drug response, contributed to the decreasing effect sizes over time. Finally, sample size and related measures arose several times as significant moderators, and these are modifiable design features for drug development. There could be a vicious circle. Sample sizes have increased continually over the years (see online Figure S4). Companies conduct large trials to assure statistical significance. The inclusion of many patients and sites leads to more recruitment pressure and variability, which, by definition, reduces effect sizes (SMD=mean difference/*standard deviation*). The next sample size estimation will suggest an even larger sample. We recommend somewhat smaller studies, but with better selected patients, to reverse this trend.

AUTHOR AND ARTICLE INFORMATION

From the Department of Psychiatry and Psychotherapy, Technische Universität München, Munich, Germany; the Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford, U.K.; the Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece; the Institute of Social and Preventive Medicine. University of Bern, Bern, Switzerland; and the Psychiatric Institute, University of Illinois at Chicago, and the Maryland Psychiatric Research Center, Baltimore.

Address correspondence to Dr. Leucht (stefan.leucht@tum.de).

Supported by grant FKZ 01KG1115 from the German Federal Ministry for Education and Research (Bundesministerium für Bildung und Forschung, BMBF). Dr. Cipriani is supported by the NIHR Oxford Cognitive Health Clinical Research Facility. Dr. Mavridis and Dr. Salanti received funding from the European Research Council (IMMA 260559 grant). Dr. Salanti is a Marie Skłodowska-Curie Fellow. The funding sources had no role in study design, data collection, data analysis, data interpretation, writing of the report, or the decision to submit the paper for publication.

The authors thank Samantha Roberts for help in the literature search, Magdolna Tardy, M.Sc., for help in study selection, and Marc Krause, Philipp Rothe, M.D., Thomas Arndt, and Natalie Peter for help in data extraction; they were all specifically trained for this purpose. The authors also thank Ole Andreassen, Guy Chouinard, Michael Jann, Herbert Meltzer, Bret Rutherford, Merck, Dainippon Sumitomo, Novartis, Sunovion, Eli Lilly, Johnson & Johnson, and Pfizer for replying to requests about their trials; we mainly obtained answers about methodological questions. Ofer Agid sent us his complete data set. The authors also thank the many authors who sent data for our previous publications on antipsychotic drugs that could be used again here.

In the last 3 years Dr. Stefan Leucht has received honoraria for consulting from Roche, TEVA, Otsuka, Lundbeck, and LB Pharma; for lectures from Otsuka, Lundbeck, Janssen, ICON, Lilly, SanofiAventis, AOP Orphan, Roche, and Servier; and for a publication from Roche. Dr. Claudia Leucht is the spouse of Dr. Stefan Leucht; therefore, his conflicts of interest are also relevant for her. Dr. Huhn has received lecture honoraria from Janssen and Lundbeck (Institute). Dr. Cipriani was expert witness for Accord Healthcare for a patent issue about quetiapine extended release. The other authors report having no financial relationships with commercial interests.

Received Dec. 8, 2016; revisions received March 16, March 18, and March 20, 2017; accepted March 20, 2017; published online May 25, 2017.

REFERENCES

1. Gøtzsche PC, Young AH, Crace J: Does long term use of psychiatric drugs cause more harm than good? BMJ 2015; 350:h2435
2. Moncrieff J: The Bitterest Pills: The Troubling Story of Antipsychotic Drugs. Basingstoke, UK, Palgrave McMillan, 2013
3. Angermeyer MC, Van der Auwera S, Matschinger H, et al: The public debate on psychotropic medication and changes in attitudes 1990–2011. Eur Arch Psychiatry Clin Neurosci 2016; 266:165–172
4. Lacro JP, Dunn LB, Dolder CR, et al: Prevalence of and risk factors for medication nonadherence in patients with schizophrenia: a comprehensive review of recent literature. J Clin Psychiatry 2002; 63:892–909

5. Lexchin J, Bero LA, Djulbegovic B, et al: Pharmaceutical industry sponsorship and research outcome and quality: systematic review. BMJ 2003; 326:1167–1170
6. Heres S, Davis J, Maino K, et al: Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: an exploratory analysis of head-to-head comparison studies of second-generation antipsychotics. Am J Psychiatry 2006; 163: 185–194
7. National Institute of Mental Health Psychopharmacology Service Center Collaborative Study Group: Phenothiazine treatment in acute schizophrenia: effectiveness. Arch Gen Psychiatry 1964; 10:246–261
8. Center for Drug Evaluation and Research: Application Number: 200603: Summary Review. Silver Spring, Md, Center for Drug Evaluation and Research, 2010. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2010/200603Orig1s000SumR.pdf
9. Agid O, Siu CO, Potkin SG, et al: Meta-regression analysis of placebo response in antipsychotic trials, 1970-2010. Am J Psychiatry 2013; 170:1335–1344
10. Rutherford BR, Pott E, Tandler JM, et al: Placebo response in antipsychotic clinical trials: a meta-analysis. JAMA Psychiatry 2014; 71: 1409–1421
11. Leucht S, Cipriani A, Spineli L, et al: Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. Lancet 2013; 382:951–962
12. Liberati A, Altman DG, Tetzlaff J, et al: The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009; 6:e1000100
13. Carpenter WT Jr, Buchanan RW: Schizophrenia. N Engl J Med 1994; 330:681–690
14. Leucht S, Tardy M, Komossa K, et al: Antipsychotic drugs versus placebo for relapse prevention in schizophrenia: a systematic review and meta-analysis. Lancet 2012; 379:2063–2071
15. Honigfeld G: Clozapine: antipsychotic activity in treatment-resistant schizophrenics. Adv Ther 1984; 1:77–97
16. Gardner DM, Murphy AL, O'Donnell H, et al: International consensus study of antipsychotic dosing. Am J Psychiatry 2010; 167: 686–693
17. Higgins JPT, Green S: Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. Chichester, UK, Wiley and Sons, 2011
18. McMahon RP, Kelly DL, Boggs DL, et al: Feasibility of reducing the duration of placebo-controlled trials in schizophrenia research. Schizophr Bull 2008; 34:292–301
19. Wu TX, Li YP, Liu GJ, et al: Investigation of authenticity of 'claimed' randomized controlled trials (RCTs) and quality assessment of RCT reports published in China. Abstracts of the 14th Cochrane Colloquium, Dublin, Ireland, 2006. https://abstracts.cochrane.org/2006-dublin/investigation-authenticity-claimed-randomized-controlled-trials-rcts-and-quality
20. Woodhead M: 80% of China's clinical trial data are fraudulent, investigation finds. BMJ 2016; 355:i5396
21. Adams CE, Coutinho E, Davis JM, et al: The Cochrane Library. Chichester, UK, John Wiley & Sons, 2011
22. Klein DF, Davis JM: Diagnosis and Drug Treatment of Psychiatric Disorders. Baltimore, Williams and Wilkins, 1969
23. Adams CE, Awad G, Rathbone J, et al: Chlorpromazine versus placebo for schizophrenia. Cochrane Database Syst Rev 2007; 2: CD000284
24. Joy CB, Adams CE, Laurie S: Haloperidol versus placebo for schizophrenia. Cochrane Database Syst Rev 2006; 4:CD003082; update 2013; 11:CD003082
25. Shen X, Xia J, Adams CE: Flupenthixol versus placebo for schizophrenia. Cochrane Database Syst Rev 2012; 11:CD009777
26. Matar HE, Almerie MQ: Oral fluphenazine versus placebo for schizophrenia. Cochrane Database Syst Rev 2007; (1):CD006352
27. Hartung B, Wada M, Laux G, et al: Perphenazine for schizophrenia. Cochrane Database Syst Rev 2005; (1):CD003443
28. Omori IM, Wang J: Sulpiride versus placebo for schizophrenia. Cochrane Database Syst Rev 2009; (2):CD007811
29. Fenton M, Rathbone J, Reilly J, et al: Thioridazine for schizophrenia. Cochrane Database Syst Rev 2007; (3):CD001944
30. Leucht S, Arbter D, Engel RR, et al: How effective are second-generation antipsychotic drugs? A meta-analysis of placebo-controlled trials. Mol Psychiatry 2009; 14:429–447
31. Kay SR, Fiszbein A, Opler LA: The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. Schizophr Bull 1987; 13:261–276
32. Overall JE, Gorham DR: The Brief Psychiatric Rating Scale. Psychol Rep 1962; 10:790–812
33. Marshall M, Lockwood A, Bradley C, et al: Unpublished rating scales: a major source of bias in randomised controlled trials of treatments for schizophrenia. Br J Psychiatry 2000; 176:249–252
34. Leucht S: Measurements of response, remission, and recovery in schizophrenia and examples for their clinical application. J Clin Psychiatry 2014; 75(suppl 1):8–14
35. Leucht S, Kane JM, Kissling W, et al: What does the PANSS mean? Schizophr Res 2005; 79:231–238
36. Schennach-Wolff R, Obermeier M, Seemüller F, et al: Does clinical judgment of baseline severity and changes in psychopathology depend on the patient population? Results of a CGI and PANSS linking analysis in a naturalistic study. J Clin Psychopharmacol 2010; 30: 726–731
37. Furukawa TA, Barbui C, Cipriani A, et al: Imputing missing standard deviations in meta-analyses can provide accurate results. J Clin Epidemiol 2006; 59:7–10
38. Salanti G, Higgins JP, White IR: Bayesian synthesis of epidemiological evidence with different combinations of exposure groups: application to a gene-gene-environment interaction. Stat Med 2006; 25:4147–4163
39. Samara MT, Cao H, Helfer B, et al: Chlorpromazine versus every other antipsychotic for schizophrenia: a systematic review and meta-analysis challenging the dogma of equal efficacy of antipsychotic drugs. Eur Neuropsychopharmacol 2014; 24:1046–1055
40. Hartling L, Abou-Setta AM, Dursun S, et al: Antipsychotics in adults with schizophrenia: comparative effectiveness of first-generation versus second-generation medications: a systematic review and meta-analysis. Ann Intern Med 2012; 157:498–511
41. Dold M, Samara MT, Li C, et al: Haloperidol versus first-generation antipsychotics for the treatment of schizophrenia and other psychotic disorders. Cochrane Database Syst Rev 2015; 1:CD009831
42. Higgins JPT, Thompson SG, Deeks JJ, et al: Measuring inconsistency in meta-analyses. BMJ 2003; 327:557–560
43. Furukawa TA, Levine SZ, Tanaka S, et al: Initial severity of schizophrenia and efficacy of antipsychotics: participant-level meta-analysis of 6 placebo-controlled studies. JAMA Psychiatry 2015; 72:14–21
44. Rabinowitz J, Werbeloff N, Caers I, et al: Determinants of antipsychotic response in schizophrenia: implications for practice and future clinical trials. J Clin Psychiatry 2014; 75:e308–e316
45. Mallinckrodt CH, Zhang L, Prucka WR, et al: Signal detection and placebo response in schizophrenia: parallels with depression. Psychopharmacol Bull 2010; 43:53–72
46. George EI, McCuloch RE: Variable selection via Gibbs sampling. J Am Stat Assoc 1993; 88:881–889
47. Mattila T, Wohlfarth T, Koeter M, et al. Geographic variation in efficacy of atypical antipsychotics for the acute treatment of schizophrenia—an individual patient data meta-analysis. Eur Neuropsychopharmacol 2014; 24:1067–1077
48. Leucht S, Rothe P, Davis JM, et al: Equipercentile linking of the BPRS and the PANSS. Eur Neuropsychopharmacol 2013; 23: 956–959
49. Furukawa TA, Levine SZ, Tanaka S, et al: Initial severity of schizophrenia and efficacy of antipsychotics: participant-level meta-analysis of 6 placebo-controlled studies. JAMA Psychiatry 2015; 72:14–21
50. Zohar J, Stahl S, Moller HJ, et al: A review of the current nomenclature for psychotropic agents and an introduction to the

Neuroscience-based Nomenclature. Eur Neuropsychopharmacol 2015; 25:2318–2325
51. Schulz KF, Chalmers I, Hayes RJ, et al: Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995; 273:408–412
52. Wood L, Egger M, Gluud LL, et al: Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 2008; 336: 601–605
53. Porta N, Bonet C, Cobo E: Discordance between reported intention-to-treat and per protocol analyses. J Clin Epidemiol 2007; 60: 663–669
54. Egger M, Davey Smith G, Schneider M, et al: Bias in meta-analysis detected by a simple, graphical test. BMJ 1997; 315:629–634
55. Peters JL, Sutton AJ, Jones DR, et al: Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. J Clin Epidemiol 2008; 61:991–996
56. Mavridis D, Welton NJ, Sutton A, et al: A selection model for accounting for publication bias in a full network meta-analysis. Stat Med 2014; 33:5399–5412
57. 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; 56:455–463
58. Leucht S, Kane JM, Etschel E, et al: Linking the PANSS, BPRS, and CGI: clinical implications. Neuropsychopharmacology 2006; 31: 2318–2325
59. Leucht S, Kane JM, Kissling W, et al: Clinical implications of BPRS scores. Br J Psychiatry 2005; 187:363–371
60. Levine SZ, Rabinowitz J, Engel R, et al: Extrapolation between measures of symptom severity and change: an examination of the PANSS and CGI. Schizophr Res 2008; 98:318–322
61. Cohen J: Statistical power analysis for the behavioral sciences, 2nd ed. Hillsdale, New Jersey, Lawrence Erlbaum Associates, 1988
62. Leucht S, Busch R, Kissling W, et al: Early prediction of antipsychotic non-response among patients with schizophrenia. J Clin Psychiatry 2007; 68:352–360
63. Fusar-Poli P, Papanastasiou E, Stahl D, et al: Treatments of Negative Symptoms in Schizophrenia: Meta-Analysis of 168 Randomized Placebo-Controlled Trials. Schizophr Bull 2015; 41:892–899
64. Marder SR, Alphs L, Anghelescu IG, et al: Issues and perspectives in designing clinical trials for negative symptoms in schizophrenia. Schizophr Res 2013; 150:328–333
65. Komossa K, Depping AM, Gaudchau A, et al: Second-generation antipsychotics for major depressive disorder and dysthymia. Cochrane Database Syst Rev 2010; (12):CD008121
66. Trikalinos TA, Churchill R, Ferri M, et al: Effect sizes in cumulative meta-analyses of mental health randomized trials evolved over time. J Clin Epidemiol 2004; 57:1124–1130
67. Dechartres A, Trinquart L, Boutron I, et al: Influence of trial sample size on treatment effect estimates: meta-epidemiological study. BMJ 2013; 346:f2304
68. Siu CO, Agid O, Remington G: Measuring the effects of treatment with antipsychotics. JAMA Psychiatry 2015; 72:514–515
69. Kemmler G, Hummer M, Widschwendter C, et al: Dropout rates in placebo-controlled and active-control clinical trials of antipsychotic drugs: a meta-analysis. Arch Gen Psychiatry 2005; 62:1305–1312
70. Tandon R, Fleischhacker WW: Comparative efficacy of antipsychotics in the treatment of schizophrenia: a critical assessment. Schizophr Res 2005; 79:145–155
71. Adelson D, Epstein LJ: A study of phenothiazines with male and female chronically ill schizophrenic patients. J Nerv Ment Dis 1962; 134:543–554
72. Casey JF, Bennett IF, Lindley CJ, et al: Drug therapy in schizophrenia: a controlled study of the relative effectiveness of chlorpromazine, promazine, phenobarbital, and placebo. Arch Gen Psychiatry 1960; 2:210–220