60 years of placebo-controlled antipsychotic drug trials in acute schizophrenia: systematic review, Bayesian meta-analysis and meta-regression of efficacy predictors

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

Word count body of the text: 4995 words, 3 Tables, 6 Figures

Corresponding author:

Prof. Stefan Leucht, MD

Department of Psychiatry and Psychotherapy, Technische Universität München

Klinikum rechts der Isar, Ismaningerstr. 22, 81675 Munich, Germany

Tel: +49-89-4140-4249, Fax: +49-89-4140-4888, e-mail: Stefan.Leucht@tum.de

Prof. Andrea Cipriani, MD, PhD, (andrea.cipriani@psych.ox.ac.uk) and Prof. John Geddes, MD, (john.geddes@psych.ox.ac.uk), Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford OX3 7JX, UK

Dr. Claudia Leucht, MD, (Claudia.Leucht@tum.de), Dr. Maximilian Huhn, MD, (Maximilian.Huhn@tum.de), Bartosz Helfer, MSci, (bartosz.helfer@gmail.com), Dr. Myrto Samara (samaramyrto@gmail.com), MD, and Matteo Rabaioli (m.rabaioli@gmx.de) and Susanne Bächer (Susanne.Baecher@gmx.net) have the same address as Stefan Leucht.

Dr. Anna Chaimani, PhD, (annachaimani@gmail.com), Dr. Dimitris Mavridis, PhD, (dimi.mavridis@googlemail.com): Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, University Campus Ioannina 45110, Ioannina, Greece

Prof. Georgia Salanti, PhD (georgia.salanti@ispm.unibe.ch): Institute of Social and Preventive Medicine (ISPM). University of Bern, Finkenhubelweg 11, 3012 Bern, Switzerland

Prof. John M. Davis, MD (jdavis@psych.uic.edu) : Psychiatric Institute, University of Illinois at Chicago, 1601 W. Taylor St., Chicago, IL 60612, and Maryland Psychiatric Research Center, Baltimore, MD, USA

Disclosures and acknowledgement

We thank Samantha Roberts for her help in the literature search, Magdolna Tardy, MSc, for her help in study selection and Marc Krause, Philipp Rothe, MD, Thomas Arndt and Natalie Peter for their help in data extraction. They had all been specifically trained for this purpose.

Ole Andreassen, Guy Chouinard, Michael Jann, Herbert Meltzer, Bret Rutherford, Merck, Dainippon Sumitomo, Novartis, Sunovion, EliLilly, Johnson&Johnson and Pfizer for replying to requests about their trials. We mainly obtained answers about methodological questions. Ofer Agid sent us his complete dataset. We also thank the many authors who sent data for our previous publications on antipsychotic drugs which could be used again here.

In the last three years Stefan Leucht has received honoraria for lectures from EliLilly, Lundbeck (Institute), Pfizer, Janssen, BMS, Johnson and Johnson, Otsuka, Roche, SanofiAventis, ICON, Abbvie, AOP Orphan, Servier; for consulting/advisory boards from Roche, Janssen, Lundbeck, EliLilly, Otsuka, TEVA; for the preparation of educational material and publications from Lundbeck Institute and Roche. EliLilly has provided medication for a clinical trial led by SL as principal investigator. Andrea Cipriani was expert witness for Accord Healthcare for a patent issue about quetiapine extended release. The other authors have no conflicts of interest to declare.

The meta-analysis was supported by the German Federal Ministry for Education and Research (Bundesministerium für Bildung und Forschung, BMBF) Grant: FKZ 01KG1115. Andrea Cipriani is supported by the NIHR Oxford Cognitive Health Clinical Research Facility. Dimitris Mavridis and Georgia Salanti received funding from the European Research Council [IMMA 260559 grant]. 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.

ABSTRACT

Objective: Antipsychotic drug efficacy may have decreased over the decades. We, therefore, present a comprehensive meta-analysis of all placebo-controlled trials in acute schizophrenia, we investigate which trial characteristics have changed over the years and which ones are moderators of drug-placebo efficacy differences.

Method: We searched multiple electronic databases, ClinicalTrials.gov and the FDA website. The outcomes were overall efficacy (primary outcome), responder rates, drop-out rates, positive, negative and depressive symptoms, quality of life, functioning, and major side-effects. Multiple potential moderators of overall efficacy were analyzed by meta-regression.

Results: 167 double-blind randomized controlled trials with 28102 participants were included. The standardized mean difference (SMD) for overall efficacy was 0.47 (95% CrI 0.42,0.51), but accounting for small trial effects/publication bias reduced the SMD to 0.38. 51% in the antipsychotic group versus 30% in placebo had at least a 'minimal' response, and 23% versus 14% had a 'good' response. Positive symptoms improved more than negative symptoms and depression. There were also small-to medium sized improvements in quality of life and functioning (SMDs 0.35 and 0.34). In the analysis of response predictors, 17 of 26 trial characteristics analyzed changed over the decades. But in a multivariable meta-regression, only industry-sponsorship and increasing placebo response were significant moderators of effect sizes. Importantly, drug response remained stable over time.

Conclusions: Approximately two times more patients improved under antipsychotics compared to placebo, but only a minority experienced a good response. Industry sponsorship reduced, rather than increased effect sizes. The decrease of effect sizes over the years was caused by increasing placebo response, not by decreasing drug response. Meta-analyses need to take this confounder into account. In drug development, somewhat smaller sample sizes but better selected patients may overcome a possible vicious circle of increasing sample sizes, more variability and smaller effect sizes.

INTRODUCTION

Antipsychotics are attacked by distinguished physicians such as prestigious researchers of the Cochrane Collaboration (1) and others (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 criticism 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 (487 participants) NIMH-sponsored trial from 1964 which is often used as a reference for antipsychotic drug efficacy showed a substantial difference between antipsychotics and placebo (61% versus 22% were much improved). In contrast, in recent years, there have been a number of failed trials where even standard drugs such as haloperidol did not outperform placebo (7). A systematic review and meta-regression suggested that increasing placebo response could explain this phenomenon. However it only analyzed predictors of placebo response, but it had little on the improvement on drug above placebo, which is crucial for patients (8). 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 (8). 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 controlled trials which are very different from placebo-controlled trials. It is not plausible that drug-response decreases when placebo-response increases making a re-assessment important (see discussion) (9). Finally, a recent network meta-analysis did not provide a meta-regression of efficacy predictors. Moreover, it primarily examined newer second-generation antipsychotics (and only two of 52 old ones; http://www.whocc.no/), thus offered limited information on the first forty years of antipsychotic drugs (10).

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 drugs trials and should help to put the debate about antipsychotic drugs on a rational basis.

METHODS

We followed the PRISMA guidelines (11) (checklist in Table S1 of the data supplement) and initially published a protocol in PROSPERO (CRD42013003342, data supplement Table S2).

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, or delusional disorder, because these do not require generally different treatment (12)). We excluded relapse prevention studies in stable patients receiving maintenance medication (13), in patients with predominant negative symptoms, and in patients with major concomitant somatic or psychiatric illness.

Interventions

We included all antipsychotics licensed in at least one country, except clozapine, a more efficacious drug (10), so that pooling with the other compounds would not have been appropriate *(only one clozapine arm with nine patients had to be excluded on this basis(14) making the impact of this decision negligible)*. We excluded intramuscular formulations, because these are used primarily as sources either for emergency use (short-acting i.m. 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 fixed-dose studies we only included target- to- maximum doses according to the International-Consensus-Study-of-Antipsychotic-Dosing (15). We averaged the results of eligible

fixed doses of single studies with appropriate formulas before entering the study in the meta-analysis (Cochrane Handbook (16)).

Types of studies

Published and unpublished, double-blind, placebo-controlled randomized controlled trials of at least 3 weeks duration (17). Studies with a high risk in sequence generation or allocation concealment were excluded (16). We a priori excluded Chinese studies due to serious quality concerns (18, 19). Risk of bias was independently assessed by at least two of the following reviewers (CL, SL, MH, BH) with the Cochrane Collaboration's risk-of-bias-tool (16).

Search strategy

We searched the Cochrane-Schizophrenia-Group-Controlled-Trials-Register (compiled by regular systematic searches of more than 15 databases, clinical trial registers, the FDA website, hand searches and conference proceedings (20), 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, online Table S3 presents search terms), supplemented by screening previous reviews (8, 10, 21-29).

Outcomes

1. The primary outcome was the mean overall change in symptoms, with the following order: change in Positive and Negative Syndrome Scale (PANSS, (30)) total score; if not available, change in Brief Psychiatric Rating Scale (BPRS, (31)); then endpoint values of these scales; then other published schizophrenia rating scales (32).

2. Responders: We analyzed how many patients achieved a) at least a "minimal response", defined as either at least 20% PANSS/BPRS reduction from baseline or a CGI at least slightly improved (33-35); and b) a "good response" – either at least 50% PANSS/BPRS reduction or CGI at least much improved (33-

35). Results of the single definitions were consistent and were presented separately, as well. We also analyzed c) any study-defined definition.

3. Discontinuation: Drop-outs due to a) any cause, b) inefficacy.

4. Positive, negative and depressive symptoms, quality of life, social functioning (measured by published rating scales (32))

6. Major side-effects: 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 MH, MT, MS and SL independently selected potentially relevant publications from the abstracts found by our search and decided to include studies, and at least two reviewers among CL, MH, BH, MS, MR, SB, MK, PR, TA, NP and SL (see acknowledgement) extracted data in duplicate in Excel sheets. 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. We preferably extracted intention-to-treat data and we preferred mixed-effect-model-of-repeated-measurements (MMRM) models over last-observation-carried-forward (LOCF). 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 (36, 37).

Statistical synthesis of study results

We used a Bayesian hierarchical random-effects model in OpenBUGS 3.2.3 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 (10, 38-40), but results of individual drugs are presented as well. For the primary analysis we merged the different antipsychotic arms within multi-arm trials (16) but properly accounted for the inherent correlation in the drug-specific analyses. We estimated

standardized mean differences for continuous outcomes and risk ratios for dichotomous outcomes, together with their 95% credible intervals (CrI). Numbers-needed-to-treat-to-benefit/harm were estimated 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 considerable heterogeneity (41)).

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 by previous evidence (8, 9, 42-44), 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 (45)). 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. Agid et al.(8)) are marked with an asterix*.

Patient related factors

The patient-related factors were: chronicity (8) measured by the patients' age, duration of illness, duration of the current episode and first episode status (8, 43);percentage men (43); US American populations versus not/mixed countries (46); degree of placebo response (8) and degree of drug (9) response* (both measured by the PANSS change or the BPRS change converted to the PANSS using a validated method

(47)); severity at baseline (PANSS total score (48)), in- versus outpatient (8); and operationalized criteria (e.g. ICD-10 or DSM-III to IV-R) versus unspecific 'clinical diagnoses'.

Drug related factors

We classified the antipsychotics by their mechanisms according to the "Neuroscience-based Nomenclature" (49); antipsychotic doses in chlorpromazine equivalents according to the International-Consensus-Study-of Antipsychotic-Dosing (9, 15); and fixed versus flexible dose studies (44).

Study design related factors

We analyzed the impact of risk of bias (appropriate versus unclear randomization (50) and allocation concealment methods (51),blinding (51), and missing outcome data (16, 52)); study duration (8); duration of wash-out (8); requirement of a scale-derived minimum of symptoms at baseline (48); PANSS versus BPRS as a scale; sample size (53); number of sites (8); percentage of academic sites (8)*; number of medications and arms (8);percentage of participants randomized to placebo (44); and drug company sponsorship of at least one study arm (medication donation alone was not considered company sponsorship (6)).

Sensitivity analyses of the 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.*

Publication bias

We used contour enhanced funnel-plots (54), a selection model (software OPEN BUGS (55)), and the trim-and-fill method (56) to assess whether eventual small trial effects were likely due to publication bias (Table S2 presents details).

RESULTS

Description of included studies

Supplemental Figure S1 presents the PRISMA (11) flow diagram. Overall, 167 studies published from 1955 to 2016 with 28102 participants were included (see Table S4). The mean duration of illness was 13.4 (SD 4.7) years, the mean age 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. In studies on acutely ill 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 (8 studies), paliperidone (8 studies), aripiprazole (9 studies), thioridazine (7 studies), lurasidone (7 studies), asenapine (6 studies) and loxapine (6 studies), for all other drugs less than 5 studies were available. Risk of bias is presented in supplemental Table S6. We only included randomised, double-blind trials, but the reports often did not indicate full details about sequence generation, allocation concealment, although this has improved over the years (see Table 1). Descriptions of 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%) 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 (N) with 22741 participants (n)). Patients treated with antipsychotics were twice as likely to respond as those on placebo when any response criterion was accepted [N=97, n=20690, response ratio 1.93 (1.72,2.19), number-needed-to-treat-to-benefit 6 (5,8), I^2 61%]. 51% of the antipsychotic-treated patients compared to

30% on placebo had at least a 'minimal' response [N=46, n=8918, response ratio 1.75 (1.59,1.97), number-needed-to-treat-to-benefit 5 (4,5)], while 23% versus 14% had a 'good' response [N=38, n=8403, response ratio 1.96 (1.65,2.44), number-needed-to-treat-to-benefit 8 (6,11)]; Figure 1. Similar results were obtained when responder rates based on PANSS/BPRS or CGI were analyzed separately (Table S6, which also presents odd ratios).

Participants on placebo were more likely to discontinue the studies prematurely, both for any reason [38% drug, 56% placebo, N=105, n=22851, risk ratio 1.25 (1.20,1.31), number-needed-to-treat-to-benefit 11 (9,14), I^2 19%] and for inefficacy of treatment [13% drug, 26% placebo, N=94, n=23017, risk ratio 2.09 (1.90,2.32), number-needed-to-treat-to-benefit 7 (6,9), I^2 46%].

The effect size for positive symptoms [N=64, n=18174, SMD 0.45 (0.40,0.50), I^2 56%] was similar to that of overall symptoms, while effects on negative symptoms [N=69, n=18632, SMD 0.35 (0.31,0.40), I^2 42%], and depression [N=33, n=9658, SMD=0.27 (0.20,0.34), I^2 50%] were smaller.

Based on six trials the quality of life of participants in the antipsychotic groups was better than that in the placebo group [N=6, n=1900, SMD 0.35 (0.16,0.51), I^2 43%], and so were improvements in social functioning in ten trials [N=10, n=3077, SMD 0.34 (0.21,0.47), I^2 46%] (Figure 2).

Antipsychotic drugs produced more movement disorders [antiparkinson medication use 19% drug, 10% placebo, N=63, n=14942, risk ratio 1.93 (1.65, 2.29), number-needed-to-treat-to-harm 12 (9,16), I^2 51%], were more sedating [14% sedated versus 6%, N=86, n=18574, risk ratio 2.80 (2.30,3.55), I^2 54%], led to more weight gain [N=59, n=17076, SMD -0.40 (-0.47,-0.33), I^2 73%], to more prolactin increase [N=51, n=15219, SMD -0.43 (-0.55,-0.30), I^2 91%], and to more QTc prolongation [N=29, n=9883, SMD -0.19 (-0.29,-0.08), I^2 80%] than placebo. Individual drugs are presented in Figures 3 and 4 and in suppl. Figure S2.

Change of trial characteristics over time

Table 1 shows that several trial characteristics changed significantly over the years: the number of participants and sites, the use of minimum baseline severity as inclusion criteria, fixed-dose designs, use

of operationalized criteria, use of the PANSS, percentage of men, studies outside the US and placebo-response increased, while the duration of the wash-out periods, use of dopamine D2 antagonists (49), study duration, risk of bias in terms of incomplete outcome data, mean doses and the number of academic sites decreased.

Moderators of the efficacy of antipsychotics – 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 units 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 (9).

Significant s*tudy design related factors* were: larger sample size (total number of participants and sites), number of drugs, PANSS rather than BPRS, a minimum of symptoms as an entry criterion, and industry sponsorship. With the exception of 'number of drugs' all these factors were associated with *smaller* effect sizes (Table 2).

Significant *patient or drug- related factors* were: operationalised rather than clinical diagnostic criteria, higher placebo response rates, lower dose, and D2, 5-HT1A receptor partial agonists vs D2 antagonists (mainly haloperidol), all of which were associated with *smaller* effect sizes (Table 2).

Moderators of the efficacy of antipsychotics – multivariable analysis

As several significant predictors are related by nature, we made the following choices for the multivariable model: a) We chose sample size as representative for the number of sites, number of drugs and number of arms. b) We chose publication year to represent the use of operationalized criteria (such criteria did not exist for early studies) and of PANSS (introduced in 1987) versus BPRS, and of drug mechanisms according to NbN (49)) which also changed over the years. Only pharmaceutical sponsorship

and the degree of placebo response remained significant and 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 points 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 to 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, Table S7)

Publication bias

Contour enhanced funnel plots revealed small trial effects. As studies were missing in the area of non-significant 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 0.36-0.41 in various scenarios of the selection model, corroborated by the trim-and-fill method [adjusted SMD 0.38 (0.33,0.43), Table S2].

Sensitivity analyses

The use of a fixed-effects rather than a random-effects model [N=105, n=22741, SMD=0.44 (0.42,0.47)] and the exclusion of completer analyses [N=95, n=22352, SMD= 0.46 (0.42,0.51)] did not significantly change the primary outcome, nor did the use of odds ratios (Table S6).

DISCUSSION

This first comprehensive meta-analysis of all acute phase, placebo-controlled antipsychotic drug trials since the introduction of chlorpromazine presents multiple new analyses and results which 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 (22-28) had analyzed whatever response criterion was presented in the individual studies, leading to a difficult-to-interpret criteria mix. Several analyses showed that 20% PANSS/BPRS reduction roughly corresponded to minimal improvement on the CGI and 50% PANSS/BPRS reduction to much improvement, and justified analysing them together (34, 57-59) (results of individual scales were similar, Table S6). Antipsychotic drugs clearly increased the number of patients with 'any response' (51% vs 30%), but, importantly, few patients (23% vs 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 (60) and it translates to a 9.6 PANSS points difference. This contrasts with the large (n=463), early NIMH study from 1964 which has been used frequently as a benchmark for antipsychotic drug efficacy (61). Its impressive difference in response rates (61% under drug vs 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 (61). In the current review not a single study was restricted to first-episode patients. Thus, its results are only representative for chronic 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 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

The current gold-standard for effects of antipsychotics on general negative symptoms is a meta-analysis which erroneously found an almost two times higher superiority compared to placebo (SMD 0.58 vs 0.35 here), but mistakenly standard errors rather than standard deviations were often used in the calculation of effect sizes which artificially inflated them (64). In our analysis antipsychotics the effect size for negative symptoms was smaller, but in a similar range than that for positive symptoms. However, as studies with primary negative symptoms were excluded, the effect-size might mainly reflect reductions of secondary negative symptoms (63). 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 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 suggested a small to medium benefit of antipsychotics in quality of life. As in our systematic review on maintenance treatment with antipsychotics compared to placebo (13), only a few (six) recent trials reported on 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 1900 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 re-integration 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 an effect size of approximately 0.74 and a study from 2015 of 0.38, a trend which has not been stopped by the most recent antipsychotics brexpiprazole and cariprazine which had relatively small effect sizes (Figure 3 and 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 such criteria or using the PANSS, nor, that the old D2 antagonists had larger effect sizes than the recent D2/5-HT1a partial agonists. In the early studies standardized criteria and the PANSS were simply not available, and D2 antagonists were primarily examined in the older, smaller, trials. When we analyzed haloperidol separately, its superiority compared to placebo had become smaller over the years, as well, demonstrating that decreasing effect sizes over time cannot be explained solely by more recent drugs being less efficacious (Figure 5b).

Larger sample sizes and the related moderator, number of sites, were associated with smaller effect sizes, which is consistent with the funnel-plot suggesting substantial small-trial effects that are well-known from other medical fields (53, 67). The patients of small trials might be better selected than those of large trials. In contrast, methodology of the often older, small trials was less stringent. For example, independent monitoring is a relative recent requirement. However, our specific tests suggest that studies were missing, at least in part, due to publication bias.

The purpose of minimum baseline severity thresholds is to have drug-responsive populations (48), but, counterintuitively, they 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, against 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 effective antipsychotics are available, patients think twice before consenting to a placebo-controlled trial and those who do consent can be a negative selection, such as (partial) non-responders to previous drugs or so-called "professional patients" who benefit from a free trial of medication answering a newspaper advertisement. These factors may also contribute to high placebo response.

Differences to previous analyses

Agid and colleagues (8) focused on predictors of *placebo response* while our research question was 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 Agid et al. (8) were also significant predictors of drug-placebo differences here, at least in univariable analyses (publication year and the number of sites). Another explanation than the different research question is that our database was two-times larger which made our results more robust (50 randomized controlled trials in Agid et al.(8) versus 105 here).

Rutherford and colleagues (9) 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 multi-level meta-analysis and hierarchical modeling which is another valid method.

This difference in methods makes it unclear why Rutherford et al.'s (9) abstract findings trial duration and baseline severity were not significant in our analysis, but we had 2.5 times more placebo-controlled studies available which can have changed many findings. The major difference in results was that in Rutherford et al.' analysis (9) 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 Rutherford et al. (9) many of the data on drug response (208 arms) were from trials that compared drugs head-to-head, not from 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 re-analyzed the drug-arms of Rutherford et al.'s 39 placebo controlled studies, drug response remained stable, as well (supplemental Table S8). Thus, drug response has only decreased 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 needs to focus on reducing placebo response rather than on increasing drug response. Finally, Agid et al.(8) did not detect the publication bias and Rutherford et al. (9) didn't 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, *of which one trial with only 9 patients was excluded*) (10, 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 to placebo, however an effect size could not be calculated (71, 72). Only 46/38 studies reported on the number of participants with at least "minimal response" and "good response". It is quite 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. that conclusions about

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

Conclusions

Our results are important on several levels.

First, clinicians can expect that approximately two times more patients improve under antipsychotics compared to 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 re-integration 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 of one developed in the 2010s, the older drug might artificially turn out better due 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 (10).

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, *modifiable* design features for drug development. There could be a vicious circle. Sample sizes have increased continually over the years (see Figure S3). 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, better selected patients to reverse this trend.

REFERENCES

1. Gotzsche 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. Houndmills, Basingstoke, New Hampshire, UK: Palgrave McMillan; 2013.
3. Angermeyer MC, Van der Auwera S, Matschinger H, Carta MG, Baumeister SE, Schomerus G. The public debate on psychotropic medication and changes in attitudes 1990-2011. Eur Arch Psychiatry Clin Neurosci. 2016;266(2):165-72.
4. Lacro JP, Dunn LB, Dolder CR, Leckband SG, Jeste DV. Prevalence of and risk factors for medication nonadherence in patients with schizophrenia: a comprehensive review of recent literature. J Clin Psychiatry. 2002;63(10):892-909.
5. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. BMJ. 2003;326(7400):1167-70.
6. Heres S, Davis J, Maino K, Jetzinger E, Kissling W, Leucht S. Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: An exploratory analysis of head-to-head comparison studies of second-generation antipsychotics. AmJPsychiatry. 2006;163(2):185-94.
7. Study 049 S. A 6-week, double-blind, randomized, fixed dose, parallel-group study of the efficacy and safety of three dose levels of SM-13496 (lurasidone) compared to placebo and haloperidol in patients with schizophrenia who are experiencing an acute exacerbation of symptoms. Center for drug evaluation and researchApplication number 200603Medical review(s)http://wwwfgagov. 2010.
8. Agid O, Siu CO, Potkin SG, Kapur S, Watsky E, Vanderburg D, et al. Meta-regression analysis of placebo response in antipsychotic trials, 1970-2010. Am J Psychiatry. 2013;170(11):1335-44.
9. Rutherford BR, Pott E, Tandler JM, Wall MM, Roose SP, Lieberman JA. Placebo response in antipsychotic clinical trials: a meta-analysis. JAMA psychiatry. 2014;71(12):1409-21.
10. Leucht S, Cipriani A, Spineli L, Mavridis D, Orey D, Richter F, et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. Lancet. 2013;382(9896):951-62.
11. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoSMed. 2009;6(7):e1000100.
12. Carpenter WT, Buchanan RW. Schizophrenia. NEngJMed. 1994;330:681-90.
13. Leucht S, Tardy M, Komossa K, Heres S, Kissling W, Salanti G, et al. Antipsychotic drugs versus placebo for relapse prevention in schizophrenia: a systematic review and meta-analysis. Lancet. 2012;379(9831):2063-71.
14. Honigfeld G. Clozapine: antipsychotic activity in treatment-resistant schizophrenics. AdvTher. 1984;1:77-97.
15. Gardner DM, Murphy AL, O'Donnell H, Centorrino F, Baldessarini RJ. International consensus study of antipsychotic dosing. Am J Psychiatry. 2010;167(6):686-93.
16. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. Chichester, UK: Wiley and Sons; 2011.
17. McMahon RP, Kelly DL, Boggs DL, Li L, Hu Q, Davis JM, et al. Feasibility of reducing the duration of placebo-controlled trials in schizophrenia research. Schizophr Bull. 2008;34(2):292-301.
18. Wu TX, Li YP, Liu GJ, Bian Z, Li J, Zhang J, et al. Investigation of authenticity of 'claimed' randomized controlled trials (RCTs) and quality assessment of RCT reports published in China. Presented at the XIV Cochrane Colloquium, Dublin, Ireland, October 23-26 2006. 2006.
19. Woodhead M. 80% of China's clinical trial data are fraudulent, investigation finds. BMJ. 2016;355:i5396.
20. Adams CE, Coutinho E, Davis JM, Duggan L, Essali A, Fenton M, et al. Cochrane Schizophrenia Group. The Cochrane Library. Chichester, UK: John Wiley & Sons Ltd; 2011.
21. Klein DF, Davis JM. Diagnosis and drug treatment of psychiatric disorders. Baltimore: Williams and Wilkins; 1969.

22. Adams CE, Awad G, Rathbone J. Chlorpromazine versus placebo for schizophrenia. Cochrane Database of Systematic Reviews. 2007(2).
23. Joy CB, Adams CE, Laurie S. Haloperidol versus placebo for schizophrenia. Cochrane Database of Systematic Reviews. 2007(2).
24. Shen X, Xia J, Adams CE. Flupenthixol versus placebo for schizophrenia. Cochrane Database Syst Rev. 2012;11:CD009777.
25. Matar HE, Almerie MQ. Oral fluphenazine versus placebo for schizophrenia. Cochrane Database Syst Rev. 2007(1):CD006352.
26. Hartung B, Wada M, Laux G, Leucht S. Perphenazine for schizophrenia. Cochrane Database of Systematic Reviews. 2005(1):CD003443.
27. Omori IM, Wang J. Sulpiride versus placebo for schizophrenia. Cochrane Database Syst Rev. 2009(2):CD007811.
28. Fenton M, Rathbone J, Reilly J, Sultana A. Thioridazine for schizophrenia. Cochrane Database Syst Rev. 2007(3):CD001944.
29. Leucht S, Arbter D, Engel RR, Kissling W, Davis JM. How effective are second-generation antipsychotic drugs? A meta-analysis of placebo-controlled trials. MolPsychiatry. 2009;14(4):429-47.
30. Kay SR, Fiszbein A. The positive and negative symptome scale (PANSS) for schizophrenia. Schizophrenia Bulletin. 1987;13(2):261-75.
31. Overall JE, Gorham DR. The Brief Psychiatric Rating Scale. PsycholRep. 1962;10:790-812.
32. Marshall M, Lockwood A, Bradley C, Adams C, Joy C, Fenton M. Unpublished rating scales: a major source of bias in randomised controlled trials of treatments for schizophrenia. British Journal of Psychiatry. 2000;176:249-52.
33. 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.
34. Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel RR. What does the PANSS mean? SchizophrRes. 2005;79:231-8.
35. Schennach-Wolff R, Obermeier M, Seemuller F, Jager M, Schmauss M, Laux G, 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(6):726-31.
36. Furukawa TA, Barbui C, Cipriani A, Brambilla P, Watanabe N. Imputing missing standard deviations in meta-analyses can provide accurate results. Journal of Clinical Epidemiology. 2006;59(1):7-10.
37. Salanti G, Higgins JP, White IR. Bayesian synthesis of epidemiological evidence with different combinations of exposure groups: application to a gene-gene-environment interaction. StatMed. 2006;25(24):4147-63.
38. Samara MT, Cao H, Helfer B, Davis JM, Leucht S. Chlorpromazine versus every other antipsychotic for schizophrenia: a systematic review and meta-analysis challenging the dogma of equal efficacy of antipsychotic drugs. European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology. 2014;24(7):1046-55.
39. Hartling L, Abou-Setta AM, Dursun S, Mousavi SS, Pasichnyk D, Newton AS. Antipsychotics in Adults With Schizophrenia: Comparative Effectiveness of First-Generation Versus Second-Generation Medications: A Systematic Review and Meta-analysis. Annals of internal medicine. 2012.
40. Dold M, Samara MT, Li C, Tardy M, Leucht S. Haloperidol versus first-generation antipsychotics for the treatment of schizophrenia and other psychotic disorders. Cochrane Database Syst Rev. 2015;1:CD009831.
41. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327:557-60.
42. Furukawa TA, Levine SZ, Tanaka S, Goldberg Y, Samara M, Davis JM, et al. Initial severity of schizophrenia and efficacy of antipsychotics: participant-level meta-analysis of 6 placebo-controlled studies. JAMA psychiatry. 2015;72(1):14-21.

43. Rabinowitz J, Werbeloff N, Caers I, Mandel FS, Stauffer V, Menard F, et al. Determinants of antipsychotic response in schizophrenia: implications for practice and future clinical trials. J Clin Psychiatry. 2014;75(4):e308-16.
44. Mallinckrodt CH, Zhang L, Prucka WR, Millen BA. Signal detection and placebo response in schizophrenia: parallels with depression. PsychopharmacolBull. 2010;43:53-74.
45. George EI, McCuloch RE. Variable selection via Gibbs sampling. Journal of the American Statistical Association. 1993;88(423):881-9.
46. Mattila T, Wohlfarth T, Koeter M, Storosum J, van den Brink W, de Haan L, et al. Geographic variation in efficacy of atypical antipsychotics for the acute treatment of schizophrenia - an individual patient data meta-analysis. European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology. 2014;24(7):1067-77.
47. Leucht S, Rothe P, Davis JM, Engel RR. Equipercentile linking of the BPRS and the PANSS. European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology. 2013;23(8):956-9.
48. Furukawa TA, Levine SZ, Davis JM, Samara M, Leucht S. Initial severity and efficacy of antipsychotics: Pooled analysis of six placebo-controlled studies in acute and predominantly negative schizophrenia. JAMA Psychiatry (revision under review). 2014.
49. Zohar J, Stahl S, Moller HJ, Blier P, Kupfer D, Yamawaki S, et al. A review of the current nomenclature for psychotropic agents and an introduction to the Neuroscience-based Nomenclature. European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology. 2015;25(12):2318-25.
50. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA. 1995;273:408-12.
51. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ. 2008;336(7644):601-5.
52. Porta N, Bonet C, Cobo E. Discordance between reported intention-to-treat and per protocol analyses. J Clin Epidemiol. 2007;60(7):663-9.
53. Egger M, Davey mG, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315:629-34.
54. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. J ClinEpidemiol. 2008;61(10):991-6.
55. Mavridis D, Welton NJ, Sutton A, Salanti G. A selection model for accounting for publication bias in a full network meta-analysis. Stat Med. 2014;33(30):5399-412.
56. 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.
57. Leucht S, Kane JM, Etschel E, Kissling W, Hamann J, Engel RR. Linking the PANSS, BPRS, and CGI: Clinical implications. Neuropsychopharmacology. 2006;31(10):2318-25.
58. Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel RR. Clinical implications of BPRS scores. BrJPsychiatry. 2005;187:363-71.
59. Levine SZ, Rabinowitz J, Engel R, Etschel E, Leucht S. Extrapolation between measures of symptom severity and change: An examination of the PANSS and CGI. SchizophrRes. 2007.
60. Cohen J. Statistical power analysis for the behavioral sciences. 2 ed. Hillsdale, New Jersey: Lawrence Erlbaum Associates; 1988.
61. Cole JO. Phenothiazine treatment in acute schizophrenia. ArchGenPsychiatry. 1964;10:246-61.
62. Leucht S, Busch R, Kissling W, Kane JM. Early prediction of antipsychotic non-response. JClinPsychiatry. 2007;68:352-60.
63. Marder SR, Alphs L, Anghelescu IG, Arango C, Barnes TR, Caers I, et al. Issues and perspectives in designing clinical trials for negative symptoms in schizophrenia. Schizophr Res. 2013;150(2-3):328-33.

64. Fusar-Poli P, Papanastasiou E, Stahl D, Rocchetti M, Carpenter W, Shergill S, et al. Treatments of Negative Symptoms in Schizophrenia: Meta-Analysis of 168 Randomized Placebo-Controlled Trials. Schizophr Bull. 2015;41(4):892-9.
65. Komossa K, Depping AM, Gaudchau A, Kissling W, Leucht S. Second-generation antipsychotics for major depressive disorder and dysthymia. Cochrane Database Syst Rev. 2010(12):CD008121.
66. Trikalinos TA, Churchill R, Ferri M, Leucht S, Tuunainen A, Wahlbeck K, et al. Effect sizes in cumulative meta-analyses of mental health randomized trials evolved over time. JClinEpidemiol. 2004;57(11):1124-30.
67. Dechartres A, Trinquart L, Boutron I, Ravaud P. 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(5):514-5.
69. Kemmler G, Hummer M, Widschwendter C, Fleischhacker W. Dropout rates in placebo-controlled and active-control clinical trials of antipsychotic drugs - A meta-analysis. ArchGenPsychiatry. 2005;62(12):1305-12.
70. Tandon R, Fleischhacker WW. Comparative efficacy of antipsychotics in the treatment of schizohrenia: a critical assessment. SchizophrRes. 2005;79:145-55.
71. Adelson D, Epstein LJ. A study of phenothiazines with male and female chronically ill schizophrenic patients. J Nerv Ment Dis. 1962;134:543-54.
72. Casey JF, F. BI, J. LC, E. HL, H. GM, N. SN. Drug therapy in schizophrenia. A controlled study of the relative effectiveness of chlorpromazine, promazine, phenobarbital, and placebo. Archives of General Psychiatry1960. p. 210-20.

Table 1: Meta-regressions showing which trial characteristics have changed over the years					
Explanatory variable[a]	**Categories**	**Weighted mean publication year**	**Number of studies**	**Meta-regression coefficient**	**95% credible interval**
Study design related 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 (ref)	1988	29	-	
	Yes	2008	73	2.52[c]	1.18,5.38
Duration of wash-out period (days)[a]	-	-	89	-9.20[b]	-11.78,-6.62
Study duration (weeks)[a]	-	-	96	-0.92[b]	-1.33,-0.50
Randomization[a]	Low risk (ref)	2007	48	-	
	Unclear	2006	57	0.80[c]	0.53,1.20
Allocation concealment[a]	Low risk (ref)	2008	33	-	
	Unclear	2006	72	0.76[c]	0.48,1.21
Intention-to-treat analysis/completers[a]	ITT (ref)	2007	95	-	
	Completers	1981	7	0.21[cc]	0.11,0.39
Risk of bias due to missing outcome data[a]	Low risk (ref)	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 (ref)	2007	57	-	
	Unclear	2006	48	0.80[c]	0.52,1.23
	High risk	-	-	-	
Number of arms	Two arms (ref)	2006	10	-	
	More than two arms	2007	95	1.11 [c]	0.74,1.65
Number of medications	Two medications (ref)	2009	33	-	
	More than two medications	2005	72	0.53[c]	0.27,1.02
Industry sponsored drug or not	Non-sponsored drugs (ref)	2006	32	-	
	At least one sponsored drug	2007	65	1.15[c]	0.73,1.81
Percentage patients randomized to PBO	-	-	105	0.00[b]	-0.04,0.03

Scale[a]	BPRS (ref)	1990	33		
	PANSS	2009	68	-0.02[c]	-0.00,0.07
Drug related factors					
Drug mechanism[a]	M1 (ref)	1998	18	-	
	M2 vs M1	1999	47	0.49 [c]	0.26,0.92[c]
	M3 vs M1	2012	12	14.79 [c]	2.74,79.93[c]
	M4 vs M1	2008	17	4.40 [c]	1.17,16.51[c]
	M5 vs M1	2007	7	4.79[c]	0.65,37.17[c]
Fixed/flexible dose[a]	Fixed dose (ref)	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
Patient related factors					
Percentage men[a]	-	-	91	6.81[b]	3.81,9.82
Operationalized criteria or not[a]	Operationalized (ref)	2007	88	-	
	Not operationalized	1977	16	0.07[c]	0.02,0.20
Country[a]	USA (ref)	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 ill (years)	-	-	60	0.66[b]	-0.21,1.53
Baseline severity (PANSS total score)	-	-	85	-0.48[b]	-1.57,0.62

(ref) = the reference category for dichotomous outcomes

[a]Characteristics in these columns resulted in a statistically significant association with publication year.

[b]These coefficients show the average increase or decrease for the respective moderator associated with a 10-year increase in publication year. For example, a 10 years newer study would on average have 79.77 more participants.

[c]These coefficients show the average odds ratio of the respective moderator associated with 10-year increase in publication year. For example, a 10 years newer study would on average have 2.52 times the odds of having a baseline severity entry minimum score.

M1 – M5 are drug mechanisms of action according to the "Neuroscience-based Nomenclature (NbN)" (49): M1 = receptor antagonists (D2) clopenthixol, fluphenazine, haloperidol, perphenzaine, pimozide, pipotiazine, sulpiride, trifluoperazine. M2 = receptor antagonists (D2, 5-HT2) chlorpromazine, iloperidone, loxapine, lurasidone, olanzapine, sertindole, thioridazine, ziprasidone, zotepine. M3 = receptor partial agonists (D2, 5-HT1A) aripiprazole, brexpiprazole, cariprazine. M4= receptor antagonists (D2, 5-HT2, NE alpha2) asenapine, paliperidone, risperidone. M5= receptor antagonist (D2, 5-HT2) and reuptake inhibitor (NET) quetiapine. A few old drugs have not been classified by NbN yet.

Table 2: Univariable meta-regressions

Explanatory variable	Coefficient	95%CrI	Coefficient corresponds to	N, n	SMD at the mean value/reference category of moderator	95%CrI	Mean value/reference category of moderator	Heterogeneity SD	95%CrI	% heterogeneity explained
Study design related factors										
Publication year[a]	-0.08[a]	-0.12,-0.04	10-year increase	105, 22741	0.50	0.45,0.55	2000	0.14	0.10,0.19	12.5%
Number of total participants[a]	-0.04[a]	-0.06,-0.01	100 part. more	105, 22741	0.49	0.44,0.54	225	0.15	0.11,0.20	6.3%
Number of sites[a]	-0.02[a]	-0.04,0.00	10-site increase	96, 20941	0.49	0.44,0.55	28	0.17	0.12,0.22	-
Number of medications[a]	0.08[a]	0.02,0.15	1 drug more	105, 22741	0.40	0.33,0.47	2 drugs	0.15	0.11,0.20	6.3%
Baseline severity entry minimum score[a]	-0.17[a]	-0.29,-0.04	Min. entry score	102, 22291	0.61	0.50,0.72	without entry score	0.15	0.11,0.20	6.3%
Industry sponsored drug or not[a]	-0.15[a]	-0.25,-0.05	Sponsored	97, 22397	0.57	0.48,0.66	non-sponsored	0.14	0.10,0.19	12.5%
Scale (PANNS or BPRS) [a]	0.18[a]	0.07,0.30	BPRS	101, 22589	0.43	0.38,0.48	PANSS	0.15	0.11,0.20	6.3%
Risk of bias due to missing outcome data	0.05	-0.03,0.12	Unclear or high risk	105, 22741	0.45	0.40,0.50	low risk	0.16	0.12,0.21	0%
Percentage of academic sites	0.01	-0.01,0.03	10% increase	59, 9379	0.57	0.51,0.64	58%	0.15	0.08,0.23	6.3%
Number of arms	0.00	-0.04,0.04	1 arm increase	105, 22741	0.47	0.38,0.57	2 arms	0.16	0.12,0.21	0%
Minimum duration of the wash-out phase	0.03	-0.01,0.06	10-day increase	89, 18586	0.50	0.45,0.55	9	0.15	0.11,0.20	6.3%
Percentage randomized to placebo	0.01	-0.05,0.07	10% increase	105, 22741	0.47	0.42,0.52	28.3%	0.16	0.12,0.21	0%
Study duration	0.10	-0.10,0.29	10-week increase	96, 22443	0.46	0.42,0.51	6.5 weeks	0.16	0.11,0.21	0%
Blinding	0.00	-0.09,0.09	Unclear or high risk	105, 22741	0.47	0.41,0.53	low risk	0.16	0.12,0.21	0%
Allocation concealment	0.01	-0.08,0.11	Unclear risk	105, 22741	0.46	0.39,0.53	low risk	0.16	0.12,0.21	0%
Randomization	-0.03	-0.13,0.06	Unclear risk	105, 22741	0.48	0.42,0.55	low risk	0.16	0.12,0.21	0%
Drug related factors										
Drug mechanism M2 vs M1	-0.13	-0.28,0.01	M2	101, 22315[b]	0.60	0.48,0.73	M1	0.16[b]	0.12,0.21	0%
Drug mechanism M3 vs M1[a]	-0.26[a]	-0.43,-0.09	M3	101, 22315[b]	0.60	0.48,0.73	M1	0.16b	0.12,0.21	0%
Drug mechanism M4 vs M1	-0.11	-0.28,0.05	M4	101, 22315[b]	0.60	0.48,0.73	M1	0.16b	0.12,0.21	0%
Drug mechanism M5 vs M1	-0.18	-0.39,0.03	M5	101, 22315[b]	0.60	0.48,0.73	M1	0.16b	0.12,0.21	0%
Fixed/flexible dose	0.04	-0.08,0.17	Flexible dose	105, 22741	0.46	0.41,0.51	fixed dose	0.16b	0.12,0.21	0%

Mean dose[a]	0.03[a]	0.00,0.05	100 CPZ unit increase	91, 19957	0.49	0.45,0.54	580.6	0.15	0.11,0.20	6.3%
Patient related factors										
Operationalized criteria or not[a]	0.22[a]	0.04,0.40	No operationalized criteria	103, 22151	0.45	0.41,0.50	Op.criteria	0.16	0.11,0.20	0%
Placebo response (mean PANSS change score in placebo arm) [a]	-0.15[a]	-0.21,-0.09	10-unit PANSS increase	99, 22520	0.48	0.44,0.52	6.24	0.13	0.08,0.18	18.8%
Drug response (mean change score in drug arm)	0.05	-0.02,0.12	10-unit PANSS increase	100, 22564	0.46	0.42,0.51	17.45	0.16	0.11,0.21	0%
Average age	-0.08	-0.20,0.03	10-year increase	100, 22567	0.47	0.42,0.51	38	0.16	0.11,0.20	0%
Baseline severity	0.10	-0.01,0.20	10-unit PANSS increase	85, 21259	0.45	0.41,0.50	94.6	0.16	0.11,0.21	0%
Duration ill	-0.07	-0.23,0.08	10-year increase	60, 14278	0.47	0.42,0.53	14	0.15	0.09,0.21	6.3%
Percentage of men	-0.01	-0.04,0.02	10% increase	91, 21119	0.46	0.41,0.51	66.3%	0.16	0.12,0.21	0%
Country	0.02	-0.07,0.12	Non-USA or mixed study	105, 22741	0.45	0.38,0.53	USA	0.16	0.12,0.21	0%
First episode#										
Duration of current episode#										
In-outpatients at study start#										

[a]Statistically significant moderators, N = number of studies, n= number of patients.

"Coefficient corresponds to": for example publication year: a 10-year increase in publication year on the average reduces the standardized mean difference by 0.08 units.

Standardized mean difference at the mean value/reference category = standardised mean difference after adjustment for covariate. For example, after adjustment for publication year, a study published in 2000 would on average have an standardized mean difference of 0.50, or non-sponsored studies would have an average standardized mean difference of 0.57. M1 – M5 are drug mechanisms of action according to the "Neuroscience-based Nomenclature" (49): M1 = receptor antagonists (D2) clopenthixol, fluphenazine, haloperidol, perphenzaine, pimozide, pipotiazine, sulpiride, trifluoperazine. M2 = receptor antagonists (D2, 5-HT2) chlorpromazine, iloperidone, loxapine, lurasidone, olanzapine, sertindole, thioridazine, ziprasidone, zotepine. M3 = receptor partial agonists (D2, 5-HT1A) aripiprazole, brexpiprazole, cariprazine. M4= receptor antagonists (D2, 5-HT2, NE alpha2) asenapine, paliperidone, risperidone. M5= receptor antagonist (D2, 5-HT2) and reuptake inhibitor (NET) quetiapine. A few old drugs have not been classified by NbN yet. [b]Numbers for the overall model

#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 parameters could not be analyzed in a meaningful way.

Table 3: Multivariable meta-regression model

Moderator	Coefficient	95% CrI	Coefficient corresponds to	Interpretation	Proba-bility [b]
Placebo response[a]	-0.13	-0.20,-0.06[a]	10-unit increase	A 10 PANSS points higher mean change score in the placebo arm would reduce the standardized mean difference on average by 0.13 units	80.6%
Industry sponsored or not[a]	-0.16	-0.28,-0.04[a]	Industry sponsored	The standardized mean difference for studies including at least one sponsored drug would be on average 0.16 units smaller than non-sponsored studies	82.8%
Publication year	-0.02	-0.09,0.05	10-year increase	A 10 years later published study would have an on average 0.02 units smaller standardized mean difference	25.0%
Sample size	0.01	-0.02,0.04	100 participants increase	A 100 participants larger study would have an on average 0.01 units larger standardized mean difference	3.3%
Mean dose	0.01	-0.03,0.04	100 chlorpromazine units increase	A 100 chlorpromazine units higher mean dose would increase the standardized mean difference on average by 0.01 units	3.3%
Baseline severity minimum score	0.05	-0.13,0.21	Baseline severity minimum score	The standardized mean difference for studies having a minimum baseline severity entry score would be on average 0.05 units larger than studies without a minimum baseline severity entry score.	48.4%

<u>Summary of the model</u>: 78 studies with 19060 participants, heterogeneity standard deviation 0.11 (0.07,0.16), the model explained 31.3% of the heterogeneity

[a]Statistically significant moderators. [b] In a simulation process, this is the probability that a model that includes this moderator would have been selected as the preferred model

Legends of Figures

Figure 1: Number of patients at least minimally improved and at least much improved

PANSS = Positive and Negative Syndrome Scale, BPRS = Brief Psychiatric Rating Scale, CGI = Clinical Global Impression Improvement Scale, CrI = Credible Interval

Figure 2: Quality of life and social functioning

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

Figure 3: Single antipsychotics compared to placebo: positive and negative symptoms (Figure 3 is a panel of figures)

Footnote: It should be noted that 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% credible interval.

Abbreviations: N = number of trials, n = number of participants, SMD = standardized mean difference, SD = standard deviation, CrI = credible interval, ARI = aripiprazole, ASE = asenapine, BLO = blonanserine, BRE = brexpiprazole, CAR = cariprazine, CPZ = chlorpromazine, FLUPE = flupenthixol, FLUPH = fluphenazine, HAL = haloperidol, ILO = iloperidone, LOX = loxapine, LUR = lurasidone, MOL = molindone, OLA = olanzapine, PAL = paliperidone, QUE = quetiapine, RIS = risperidone, SER = sertindole, THIOT = thiothixene, ZIP = ziprasidone, ZOT = zotepine

Figure 4: Single antipsychotics compared to placebo: weight gain, QTc prolongation, prolactin increase, antiparkinson medication (Figure 4 is a panel of figures)

It should be noted that for some drugs few data were available making the results unreliable. For example, the results on weight gain for reserpine are based on only one study with 20 patients. This caused uncertainty about the true effect which is expressed by a large 95% credible interval. 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 to placebo. Abbreviations: N = number of trials, n = number of participants, r = number of participants with an event, RR = relative risk, SMD = standardized mean difference, SD = standard deviation, CrI = credible interval, 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, QUE = quetiapine, RES = reserpine, RIS = risperidone, SER = sertindole, THIOT = thiothixene, THIOR = thioridazine, TRIFLU = trifluperazine, ZIP = ziprasidone, ZOT = zotepine

* These relative risks were obtained after a continuity correction and from a fixed effect model.

Figure 5: Effect sizes over time (Figure 5 is a panel of figures)

Figure 5a: Efficacy of antipsychotic drugs compared to placebo versus publication year

Figure 5b: Efficacy of haloperidol compared to placebo versus publication year

Figure 5c: Placebo response (PANSS total score change from baseline) versus publication year

Figure 5d: Drug response (PANSS total score change from baseline) versus publication year

Footnote: SMD = standardized mean difference, PANSS = Positive and Negative Syndrome Scale, B = regression coefficient

Figure 6: Contour enhanced funnel-plot

Footnote: se = standard error, SMD = standardized mean difference, p = p-value