

# Comparative efficacy and tolerability of 32 oral antipsychotics for the acute treatment of adults with multi-episode schizophrenia: a systematic review and network meta-analysis

Maximilian Huhn<sup>1</sup>, MD, Adriani Nikolakopoulou<sup>2</sup>, PhD, Johannes Schneider-Thoma<sup>1</sup>, MD, Marc Krause<sup>1,3</sup>, MA, Myrto Samara<sup>1</sup>, MD, Natalie Peter<sup>1</sup>, Thomas Arndt<sup>1</sup>, Lio Bäckers<sup>1</sup>, Philipp Rothe<sup>4</sup>, MD, Andrea Cipriani<sup>5</sup>, PhD, John Davis<sup>5,6</sup>, Prof MD, Georgia Salanti<sup>2</sup>, PhD, Stefan Leucht<sup>1\*</sup>, MD

<sup>1</sup> Department of Psychiatry and Psychotherapy, Klinikum rechts der Isar, Technical University of Munich, Germany

<sup>2</sup> Institute of Social and Preventive Medicine (ISPM), University of Bern, Switzerland

<sup>3</sup> Institute for Evidence in Medicine (for Cochrane Germany Foundation), Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Germany

<sup>4</sup> Department of Forensic Psychiatry and Psychotherapy at the University of Ulm, Günzburg District Hospital, Germany

<sup>5</sup> Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford, UK; Oxford Health NHS Foundation Trust, Warneford Hospital, Oxford, UK.

<sup>6</sup> Psychiatric Institute, University of Illinois at Chicago, Chicago, IL, USA

<sup>7</sup> Maryland Psychiatric Research Center, Baltimore, USA

\*Corresponding author

Department of Psychiatry and Psychotherapy  
Klinikum rechts der Isar, Technische Universität München  
Ismaningerstr. 22, 81675 München, Germany  
Phone: +49-89-4140-4249  
E-mail: [Stefan.Leucht@tum.de](mailto:Stefan.Leucht@tum.de)

6 Figures

3 Tables

Webappendix

Wordcount: 3470

# Summary

## Background

Schizophrenia is one of the most common, burdensome, and costly psychiatric disorders worldwide in adults. Antipsychotic drugs are its treatment of choice, but there is controversy about which agent should be used. We, therefore, attempted to compare and rank antipsychotics by quantifying the information from randomized-controlled trials.

## Methods

We performed a network meta-analysis of placebo-controlled and head-to-head randomized controlled trials to compare 32 antipsychotics. We searched EMBASE, MEDLINE, PsycINFO, PubMed, BIOSIS, Cochrane Library, WHO International Clinical Trials Registry Platform (ICTRP) and ClinicalTrials.gov (last search 8<sup>th</sup> January 2019). Two authors independently selected studies and extracted data. Our primary outcome was change in overall symptoms measured with standardised rating scales. Additionally we extracted data for seven efficacy and nine safety outcomes. Differences in the findings of the studies were explored in meta-regressions and sensitivity analyses. Effect sizes measures were standardized mean differences (SMDs), weighted mean differences (WMDs) or risk ratios (RRs) with 95% credible intervals (CrIs). Confidence in the evidence was assessed using CINeMA (Confidence in Network Meta-Analysis). The study protocol was registered at PROSPERO (CRD42014014919).

## Findings

We identified 53,791 citations and included 402 studies with data for 53,463 participants. All antipsychotics reduced overall symptoms more than placebo, with SMDs ranging from -0.89 (95% CrI -1.08 to -0.71) for clozapine to -0.03 (-0.59 to -0.52) for levomepromazine (40,815 participants). SMDs compared to placebo for reduction of positive symptoms (31,179 participants) varied from -0.69 for amisulpride (best drug) to -0.17 for brexpiprazole (worst drug), for negative symptoms (32,015 participants) from -0.62 (clozapine) to -0.1 (flupenthixol), for depressive symptoms (19,683 participants) from -0.9 (sulpiride) to 0.04 (flupenthixol). Risk ratios compared to placebo for all-cause discontinuation (42,672 participants) ranged from 0.71 (thiothixene) to 0.96 (sertindole), for sedation (30,770 participants) from 0.92 (pimozide) to 10.2 (zuclopenthixole), for use of antiparkinson medication (24,911 participants) from 0.46 (clozapine) to 6.14 (pimozide). Weighted mean differences compared to placebo for weight gain (28,317 participants) ranged from -0.16 kg (ziprasidone) to 3.21 kg (zotepine), for prolactin elevation (21,569 participants) from -77.05 ng/ml (clozapine) to 48.51 ng/ml (paliperidone) and for QTc-prolongation (15,467 participants) from -2.21 ms (lurasidone) to 23.9 ms (sertindole). Conclusions for the primary outcome did not materially change after adjusting for possible effect moderators or in sensitivity analyses (e.g. when excluding placebo-controlled studies). The confidence in evidence, depending on the treatment comparison, was often low or very low.

## Interpretation

There are some efficacy differences between antipsychotics, but most of them are gradual rather than discrete. Differences in side effects, many of which are serious, are marked. These findings will aid clinicians in balancing risks

versus benefits of those drugs available in their countries. They should consider the importance of each outcome, the patients' medical problems, and their preferences.

### **Funding**

The project was founded by the German Ministry of Education and Research (FKZ01KG1406).

## **Research in context**

### **Evidence before this study**

Treatment with antipsychotic drugs is the standard for the acute phase of schizophrenia according to most national and international guidelines. Nevertheless, their use is strongly debated due to side effects and possible brain changes. As many antipsychotic drugs are available according to the WHO, it would be important to know how each of the many substances compares to each other, ideally ranked in a hierarchy. Due to their diverging receptor binding profiles, different antipsychotics can vary considerably in their efficacy and side effect profile. Even if there is a large evidence base of randomized clinical trials for the acute treatment of schizophrenia many evidence gaps remain, as many substances have never been compared directly in trials. This is especially true for older antipsychotics. We searched PubMed for network meta-analyses on the acute treatment of schizophrenic patients with antipsychotics. Using the search terms “antipsychotic” AND “schizophrenia” AND (network meta-analysis OR multiple treatment meta-analysis), we found several relevant systematic reviews in our last search October 2018. Some examined special subgroups like children or first episode patients. Others were restricted to special populations like Chinese or Japanese people. Some focused on special outcomes like weight gain or glucose. One network meta-analysis from 2013 did not include the newly approved antipsychotics cariprazine and brexpiprazole and examined only the two older antipsychotics haloperidol and chlorpromazine. Furthermore, not all clinically important efficacy and safety outcomes were studied in this report. Altogether we found no comprehensive and systematic network meta-analysis comparing older and newer antipsychotics for the acute treatment of schizophrenia over several efficacy and safety outcomes.

### **Added value of this study**

To our knowledge the present analysis is the largest network-meta-analysis carried out in the field of schizophrenia, based on 402 studies including 53,463 participants randomized to 32 different older and newer antipsychotics or placebo. The addition of two new and sixteen old antipsychotics is a major extension of a previous report. We investigated ten additional important outcomes, such as specific aspects of efficacy, quality of life and many more side effects, and a number of methodological issues, such as placebo response and sample sizes. The primary outcome was reduction of overall schizophrenic symptoms, but we also examined other domains: reduction in positive and negative symptoms, dropouts, depression, quality of life and functioning. Newer and older

antipsychotics reduced overall symptoms more than placebo and had lower all-cause discontinuation rates than placebo. Side-effect differences between drugs were often large. As there were fewer data for older antipsychotics, comparisons involving them were often more uncertain and their evidence of lower quality. Conclusions for the primary outcome did not materially change after adjusting for possible effect moderators or in sensitivity analyses (e.g. when excluding placebo-controlled studies).

### **Implications of all the available evidence**

These findings provide treatment rankings for older and newer antipsychotics for seventeen different outcomes, which will inform the decision-making process.

# Introduction

Schizophrenia is a common (1% of the world population are afflicted) debilitating, disorder lasting most of the patients life, with a huge burden for patients and a total cost of 155.7 billion USD only in the US.<sup>1,2</sup> Antipsychotics are the mainstay of treatment,<sup>3</sup> but are associated with important side-effects which can cause serious disability including death.<sup>4</sup> Many drugs are available according to the WHO,<sup>5</sup> which vary considerably in their affinity to different synaptic receptors, leading to possibly diverging efficacy and safety profiles. Many guidelines recommend newer antipsychotics as the treatment of choice, but older antipsychotics are less costly and still used worldwide, especially in low income countries.<sup>6</sup> Moreover, neither these old ones nor the recently approved antipsychotics brexpiprazole and cariprazine have been compared in a comprehensive network meta-analysis.<sup>7</sup> However, a clear understanding of the relative risks and benefits are essential for informed decision-making. The project extends our previous work that combined evidence from 212 randomized trials on 15 antipsychotics on seven outcomes to 32 antipsychotics and placebo from 402 randomized trials and seventeen efficacy and safety outcomes, with change in overall symptoms as the primary outcome.<sup>3</sup> The aim of this systematic review is to better inform clinical practice and mental health policies by comparing all licensed second-generation, as well as sixteen first-generation antipsychotics, in the acute treatment of adults with schizophrenia.

# Methods

## Search strategy and selection criteria

Presentation of the results follows the PRISMA guidelines for network meta-analyses (appendix 1).<sup>8</sup> We registered the protocol with PROSPERO (CRD42014014919, appendix 2).

We included randomized controlled trials (RCTs) in adults with acute symptoms of schizophrenia or related disorders (such as schizophreniform, or schizoaffective disorders). We excluded studies in patients with treatment resistance, first episode, predominant negative or depressive symptoms, concomitant medical illness, and relapse prevention studies.

We included all second-generation (“atypical”) antipsychotics (SGAs) available in Europe or the US, placebo and a selection of first-generation (“typical”, “conventional”) antipsychotics (FGA) (benperidol, chlorpromazine, clopenthixol (cis- and trans-isomer) , flupenthixol, fluphenazine, haloperidol, levomepromazine, loxapine,

molindone, penfluridol, perazine, perphenazine, pimozide, sulpiride, thioridazine, thiothixene, trifluoperazine, zuclopenthixol (cis-isomer) guided by a survey among 50 international schizophrenia experts.<sup>9</sup> We excluded intramuscular formulations because they are primarily used for relapse prevention (long-acting) or emergency use (short-acting). We included all flexible-dose studies since these allow the investigators to titrate to optimum dose for the individual patient. In fixed-dose studies, we included target to maximum doses according to the “International Consensus Study on Antipsychotic dose”.<sup>10</sup> If studies used several doses, we averaged the results of the individual arms using weighted means.<sup>11</sup>

We included published and unpublished RCTs comparing one antipsychotic with another or with placebo. Trials in which antipsychotics were used as an augmentation or combination strategy were excluded. For subjective outcomes (e.g. overall change in symptoms), we included only double-blind studies, because a lack of blinding can exaggerate differences between treatments in this area.<sup>12</sup> For objective outcomes, open studies were included (see appendix 18 for full details). We included short term studies with a follow-up period between three and 13 weeks.<sup>13</sup> Studies with a high risk of bias in sequence generation or allocation concealment according to the Cochrane Collaboration’s risk of bias tool were excluded.<sup>11</sup> We a priori excluded studies from mainland China due to serious quality concerns.<sup>14</sup> We searched Medline, Cochrane Central register of Controlled Trials (CENTRAL), Embase, Biosis, Psychinfo, Pubmed, Clinicaltrials.gov, WHO International Trial Registry and the Food and Drug Administration website without language restrictions. The search strategy combined terms for schizophrenia or psychosis and various drug names (appendix 3) until January 8, 2019. Reference lists of the included studies and previous reviews were screened for additional studies.

At least two reviewers (MH, MK, MS and YZ) screened the search results independently, retrieved full text articles and checked inclusion criteria. In case of doubt a third reviewer (SL) was involved. Two reviewers independently extracted data and entered them in electronic forms in Microsoft Access 2010 (MH, JS, MK, NP, LB, TA, YZ, PR, MS, SB, MR, LR and HR). An algorithm checked for conflicting data entries. Differences were discussed and if consensus was not reached a third reviewer (SL) was contacted. Study authors were contacted in case of missing or unclear information. 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.<sup>15</sup> Risk of bias in RCTs for the primary outcome was independently assessed by the same reviewers as above using the Cochrane Collaboration’s risk of bias tool (Appendix 19).<sup>11</sup> The

overall risk of bias was classified into high, medium and low as proposed in a large network meta-analysis for antidepressants.<sup>16</sup>

## Outcomes

The single primary outcome was change in overall symptoms of schizophrenia as measured by rating scales such as the Positive and Negative Syndrome Scale (PANSS), the Brief Psychiatric Rating Scale (BPRS) or any other published scale.<sup>17</sup> Different scales were combined using standardized mean differences (SMD). Secondary outcomes were all-cause discontinuation, discontinuation due to inefficacy, responder rates (study defined), change in positive, negative and depressive symptoms, quality of life and social functioning, measured by means of published rating scales. The following major side-effects were examined: use of antiparkinson drugs as a measure of extrapyramidal side-effects, akathisia, weight gain in kg, >7% weight gain, prolactin levels, sedation/somnolence, QTc prolongation and at least one anticholinergic side-effect (appendix 18).

## Data analysis

We conducted a network meta-analysis combining direct and indirect treatments in a Bayesian hierarchical model using the *rjags* package (appendix 4).<sup>18</sup> Effect sizes were risk ratios (RRs) for dichotomous outcomes and standardized mean differences (SMD) for continuous outcomes, except weight gain, QTc-prolongation and prolactin elevation for which weighted mean differences (WMD) were applied. Data were combined using a random effects model. Treatments were ranked using the surface under the curve cumulative ranking (SUCRA) probabilities. The transitivity assumption was evaluated by comparing the distribution of potential effect modifiers (placebo response, publication year, sample size, baseline severity, mean age and percentage male) across studies grouped by comparison (appendix 5). We assumed a common heterogeneity parameter across the various treatment comparisons, and presented the between-study variance  $\tau^2$  for each outcome. We characterized the amount of heterogeneity as low, moderate or high using the first and third quantiles of their empirical distributions.<sup>19</sup> We statistically evaluated consistency (the agreement of the various sources of evidence) using the design-by treatment test<sup>20</sup> and by separating indirect from direct evidence<sup>21</sup> using the R *netmeta* package (appendix 13).<sup>22</sup> We explored residual heterogeneity and inconsistency by several *a priori* defined meta-regressions (with covariates: placebo response rate, study sample size, study publication year, baseline severity, sponsoring, study duration, mean age, percentage men) and sensitivity analyses (excluding studies at high risk of bias, that did a completer analysis, with imputed standard deviations, placebo controlled studies, with duration more than six weeks, that were published before 1990, that were considered

as failed trials, with unfair dose comparison (studies in which the lowest dose was 50% less than the largest dose in olanzapine equivalents., e.g. a study comparing haloperidol 2mg/day (olanzapine equivalence dose =4mg) with olanzapine 10mg/day was excluded) and excluding placebo arms) (details appendix 4, 16 and 17). We used contour-enhanced funnel plots and the trim-and-fill method for the primary outcome to investigate the presence of small-study effects, whereby small studies give different results from the large studies for all comparisons against placebo and against haloperidol.<sup>23,24</sup>

The certainty of evidence produced by the synthesis for outcome was evaluated using the framework described in Salanti et al.<sup>25</sup> and implemented using the CINeMA (Confidence in Network Meta-Analysis) web application which allows to grade the confidence in the results into high, moderate, low and very low (details appendix 15).<sup>26</sup> For the primary outcome we examined the confidence of evidence of all comparisons, for the remaining outcomes we examined the comparisons of antipsychotics versus placebo.

## **Role of the funding source**

The funder of the study had no role in study design, data collection, data analysis, data interpretation or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

## **Results**

The search identified 53,791 citations and 2,827 full-text articles were retrieved. We included 550 reports from 402 studies with 53,463 participants (PRISMA flowchart figure 7, detailed appendix 6). The full list of included studies can be found in appendix 7. The sample had the following characteristics: mean age 37.40 years, 50.09% males and a mean illness duration of 11.90 years. Further characteristics are summarized in tables 1 (overall) and 2 (individual antipsychotics). We excluded studies with high risk of bias for randomization and allocation, but methods for sequence generation and allocation concealment were often not described in detail and therefore coded as unclear. Risk of bias is presented in appendix 8. The percentage of studies with high/unclear/low risk of bias for the individual items was: randomization (0%/73.1%/26.9%), allocation (0/78.4%/21.6%), rater blinding (13.9%/34.8%/51.3%), blinding of patients and personnel (10.7%/38.6%/50.7%), missing outcomes

(23.9%/35.6%/40.5%), selective reporting (27.4%/20.9%/51.7%) and other bias (5.7%/11.9%/82.3%). 23% of studies had an overall high risk of bias.

A total of 218 (54%) studies with 40,815 (76%) participants presented usable results for change in overall symptoms (Fig. 1). Figure 2a presents the overall change in symptoms for all antipsychotics compared to placebo as a reference. Figure 2a includes a color code, reflecting the strength of recommendations according to CINeMA. It also includes the sample sizes which should be considered together with the credible intervals as measures of the uncertainty of the estimates. All antipsychotics were associated with larger improvement compared to placebo (Fig. 2a), and the differences were statistically significant except for penfluridol, pimozide, perazine, trifluoperazine, fluphenazine and levomepromazine. The SMDs for drugs with significant results ranged between -0.89 (95% CrI -1.08;-0.71) for clozapine to -0.26 (95% CrI -0.4;-0.12) for brexpiprazole.

Table 3 presents the detailed results for all possible antipsychotic comparisons ordered according to their SUCRA ranking. This table needs to be consulted for the question of which drug is more efficacious than another drug. Results of network meta-analysis are presented in the left lower half and results from pairwise meta-analysis in the right upper half, if available. Table 3 shows a pattern that clozapine, amisulpride, zotepine, olanzapine and risperidone reduced overall symptoms significantly more than many other drugs. Most differences between the remaining drugs were small or very uncertain.

Ranking tables for all secondary outcomes are presented in appendix 11. Secondary efficacy outcomes were reported less frequently especially for older drugs including clozapine. 117 studies (29%) with 31,179 participants (58%) presented results for reduction of positive symptoms. All antipsychotics reduced positive symptoms more than placebo. The SMDs for drugs with significant superiorities ranged between -0.69 (95% CrI -0.86;-0.52) for amisulpride to -0.17 (95% CrI -0.31;-0.04) for brexpiprazole (Fig. 2b). Amisulpride, risperidone, olanzapine, paliperidone and haloperidol were significantly more effective than many other drugs (appendix 11).

132 studies (33%) with 32,015 (60%) participants reported usable results for negative symptoms. Most antipsychotics reduced negative symptoms more than placebo. The SMDs for significant drugs ranged between -0.62 (95% CrI -0.89;-0.39) for clozapine to -0.22 (95% CrI -0.33;-0.11) for iloperidone (Fig. 2c). Clozapine, amisulpride and olanzapine and to a lesser extent zotepine and risperidone reduced negative symptoms significantly more than many other drugs. Differences between the remaining drugs were uncertain (appendix 11).

89 studies (22%) with 19,683 participants (37%) reported usable results for depressive symptoms. Most antipsychotics reduced depressive symptoms more than placebo. The SMDs for significant drugs ranged between -0.9 (95% CrI -1.36;-0.44) for sulpiride to -0.16 (95% CrI -0.29;-0.03) for brexpiprazole (Fig. 2d). Sulpiride, clozapine, amisulpride and olanzapine were associated with significantly more reduction of depressive symptoms compared to many other drugs (appendix 11), but credible intervals were wide.

Only 10 studies (3%) with 3,341 participants (6%) reported usable quality of life data. As 50% of the network loops were inconsistent, we performed only a pairwise meta-analysis. Compared to placebo, antipsychotics improved quality of life with SMDs of drugs with significant effects ranging from -0.49 (95% CI:-0.26;-0.72) for aripiprazole to -0.18 (95% CI: -0.02;-0.34) for paliperidone (appendix 12).

16 studies (4%) with 4,370 participants (8%) presented usable results for social functioning. Based on a limited number of studies, thioridazine, olanzapine, paliperidone, quetiapine, lurasidone and brexpiprazole were associated with significant improvement in social functioning compared to placebo with a SMD range from -0.69 (95% CrI: -1.24;-0.14) for thioridazine to -0.25 (95% CrI: -0.38;-0.12) for brexpiprazole (Fig. 2e).

192 studies with 35,115 participants reported study defined response rates, using very different cut-offs (discussion s. Appendix 20.1). All antipsychotics had higher response rates compared to placebo. RRs for significant results ranged from 2.16 (95%CrI: 1.53;3.55) for thioridazine to 1.11 (95%CrI: 1.01;1.19) for brexpiprazole (Appendix 20.1).

226 studies (56%) reported all-cause discontinuation rates for 42,672 participants (80%). All drugs except pimozide had at least trends for lower discontinuation rates compared to placebo. RRs for significant results ranged from 0.52 (95%CrI: 0.11;0.95) for clopenthixol to 0.9 (95%CrI: 0.85;0.95) for haloperidol (Fig. 2f). Examining discontinuation due to inefficacy we found comparable results as for the primary outcome overall change in symptoms (appendix 20.2).

116 studies (29%) with 28,317 (53%) participant usable results for weight gain. Many antipsychotics caused more weight gain than placebo with WMDs in kilogram ranging from 0.54 kg (95%CrI: 0.15;0.95) for haloperidol to 3.21 (95%CrI: 2.1;4.31) for zotepine (Fig. 3a). Zotepine, olanzapine and sertindole produced significantly more weight gain than most other drugs (appendix 11). The hierarchy for patients with at least 7% weight gain was similar, confirming the findings (appendix 20.3)

136 studies (34%) with 24,911 (47%) participants reported use of antiparkinson medication. All drugs except clozapine, perazine and sertindole were associated with more use of antiparkinson medication compared to placebo. Significant RRs ranging from 1.61 (95%CrI: 1.17;2.1) for paliperidone to 6.14 (95%CrI: 4.81;6.55) for pimozide (Fig. 3b). The following drugs were significantly better than haloperidol starting with the best: clozapine, perazine, sertindole, placebo, olanzapine, quetiapine, asenapine, aripiprazole, thioridazine, amisulpride, iloperidone, brexpiprazole, paliperidone, ziprasidone, risperidone, lurasidone, zotepine and chlorpromazine (appendix 11). 116 studies (29%) with 25,783 (48%) reported results for akathisia. The hierarchy was similar to use of antiparkinson medication with RRs ranging from 1.95 (95%CrI: 1.3;2.74) for aripiprazole to 23.81 (95%CrI: 7.41;142.86) for zuclopenthixol (Fig. 3c).

90 studies (20%) with 21,569 participants (40%) reported usable results for prolactin. Compared to placebo clozapine, zotepine and aripiprazole seemed to reduce prolactin with WMD -77.05 ng/ml, -42.23 ng/ml, -7.1 ng/ml (95%CrI: -120.23;-33/-82.42;-1.71/-11.17;-3.09). In contrast olanzapine, asenapine, lurasidone, sertindole, haloperidol, amisulpride, risperidone and paliperidone were associated with significantly elevated prolactin levels (WMD range: 4.47-48.51 ng/ml) . Amisulpride, risperidone and paliperidone caused more prolactin elevation than haloperidol. Overall prolactin data for many antipsychotics (e.g. sulpiride) was not available (Fig. 3d).

51 studies (13%) with 15,467 participants (29%) reported usable data for QTc prolongation. Many antipsychotics caused significantly more QTc-prolongation than placebo with WMDs ranging from 3.43 ms (95%CrI: 0.94;6) for quetiapine to 23.9 ms (95%CrI:20.56;27.33) for sertindole (Fig. 3e).

162 studies (40%) with 30,770 participants (58%) reported results for sedation. All drugs except pimozide were more sedating than placebo with significant RRs ranging from 1.33 (95%CrI: 1;1.68) for paliperidone to 10.2 (95%CrI: 4.72;29.41) for zuclopenthixol and a trend to sedation for most of the remaining antipsychotics (Fig. 3f).

134 studies (33%) with 276,904 participants (50%) reported anticholinergic side effects. Most drugs were associated with a higher risk of anticholinergic side effect compared to placebo. This outcome can be influenced by use of anticholinergic rescue medication, which is often needed for classical antipsychotics. Significant evidence was present for risperidone, haloperidol, olanzapine, clozapine, iloperidone, chlorpromazine, zotepine, thioridazine and quetiapine (RR range: 1.31-4.17) (Fig. 3g)

Heterogeneity was low (to moderate) for most outcomes, moderate to high for use of antiparkinson medication and high for prolactin elevation. Separating indirect from direct evidence (SIDE test), the percentage of inconsistent comparisons was between 2% and 26% for all outcomes, except for quality of life with 50% inconsistent comparisons; so this outcome was examined only in a pairwise meta-analysis (appendix 12). In addition prolactin results were significantly inconsistent according to the design-by-treatment interaction test. As prolactin values vary widely between men and women and assays used in different laboratories, we also applied SMDs, and heterogeneity and inconsistency were substantially lower (appendix 13, 20.4).

The most important differences in terms of study characteristics were that older antipsychotics had less placebo response than newer ones and that the antipsychotics differed in their median baseline severity across studies (appendix 5). These potential threats to the transitivity assumption and other potential effect modifiers were addressed by metaregressions and sensitivity analyses of the primary outcome. We excluded antipsychotics studied in less than 100 participants. The degree of placebo-response, which has increased over the years,<sup>27</sup> had the greatest influence on the heterogeneity. The effect sizes of the individual antipsychotics changed after accounting for response to placebo, but the overall hierarchy did not materially change (fig 4a and appendix 9). This finding was corroborated by removing placebo arms or placebo controlled studies in sensitivity analyses (fig. 4b, 4c and appendix 10). Publication year, mean participants' age, baseline severity, percentage of male patients, sample size and sponsoring also did not materially change relative treatment effects compared to the unadjusted analysis (details and plots appendix 9). Sensitivity analyses removing studies with overall high risk of bias, completer analyses, imputed standard deviations, duration more than six weeks, unfair dose comparisons, failed trials and trials conducted before 1990 did not impact materially on the results (appendix 10).

The quality of the evidence was overall poor. Concerning the primary outcome we judged the confidence in the evidence for 75% of the comparisons with placebo to be low or very low (Figure 2a), and this was the case for 92% of the comparisons of two antipsychotic drugs (Appendix 15.2). Figure 5 shows that many older antipsychotics are among those with poor CINeMA ratings, and that for older drugs often no evidence was available at all for several secondary outcomes.

We conducted two contour-enhanced funnel plots of the primary outcome. Comparing all antipsychotics to haloperidol did not reveal any asymmetry and the SMD did not change using the trim-and-fill method (Fig. 6a). In

contrast comparing all antipsychotics to placebo revealed that smaller trials exaggerate the effectiveness of the active interventions versus placebo (Fig. 6b). SMD changed from 0.45 to 0.38, confirming an earlier analysis.<sup>27</sup>

## Discussion

The present analysis is the largest network-meta-analysis in the field of schizophrenia, based on 402 studies including 53,463 participants randomized to 32 different first- and second-generation antipsychotics or placebo. We extended our previous report by two SGAs and 15 FGAs and by investigating ten additional important outcomes, such as specific aspects of efficacy, quality of life and many more side effects, and a number of methodological issues, such as placebo response and sample sizes.<sup>3</sup>

All antipsychotic drugs reduced overall symptoms more than placebo with mean effect sizes between -0.89 and -0.03 (median -0.42). We emphasize that as can be derived from overlapping credible intervals in the figures comparing the drugs with placebo, most differences between drugs are gradual (not statistically significant) rather than discrete (statistically significant). With few exceptions only clozapine, amisulpride, zotepine, olanzapine and risperidone were significantly more efficacious for the primary outcome than other ones. We emphasize that readers should consult the league tables which provide this information (table 3). Amisulpride was among the most efficacious antipsychotics, but no placebo-controlled study was available making this evidence entirely indirect. Nevertheless amisulpride was significantly superior to placebo in elderly patients (SMD=0.86) and in patients with predominant negative symptoms (SMD=0.47).<sup>28,29</sup>

Mainly newer antipsychotics provided data separately for positive and negative symptoms, but the picture was similar to overall change in symptoms. However, all included studies focused on positive symptoms, as studies with predominant negative symptoms were excluded in this analysis and evaluated separately.<sup>28</sup> It is impossible to clarify in populations with positive symptoms whether differences in negative symptoms relate to primary or just secondary negative symptoms. That many drugs improved depressive symptoms more than placebo may also mainly reflect a reduction of anxiety and distress associated with schizophrenia. Nevertheless, lurasidone and quetiapine are licensed for major depression. So is flupenthixol, but we did not find an antidepressant effect for it, based on sparse data (63 participants).<sup>30</sup> Many antipsychotics did not have data for quality of life, an important outcome for patients as it combines efficacy and safety. If reported, most drugs showed better effects than placebo. Some but not all drugs also

outperformed placebo in terms of social functioning in these short-term studies, an outcome associated with recovery and social reintegration.

As all-cause discontinuation combines efficacy and tolerability, it has been used as a measure of effectiveness in the CATIE trial.<sup>31</sup> If reported separately more patients dropped out due to inefficacy (40%) rather than due to adverse events (20%) in the included trials so that all-cause discontinuation is mainly an efficacy measure.

As antipsychotics are often taken for a long period, side-effects play an important role concerning morbidity, adherence and may affect cognition.<sup>32</sup> Antipsychotics very often scored worse than placebo for side-effect outcomes, with different profiles. In general, older antipsychotics were associated often with more extrapyramidal motoric side-effects and prolactin elevation (with markable exceptions such as amisulpride, paliperidone and risperidone), whereas many newer antipsychotics produced more weight gain and sedation. We consider weight gain to be a good proxy for metabolic side-effects in this already dense review.<sup>33</sup> Specific metabolic side effects such as glucose, insulin, HOMA-IR, total cholesterol, LDL cholesterol, HDL cholesterol, triglycerides will be addressed in future reviews. In contrast to our previous report we present QTc prolongation in original units (msec) which make clinical interpretation easier. Lurasidone and the partial dopamine agonists were the most benign drugs.

Looking at efficacy and safety outcomes many older antipsychotics, limited by few direct comparisons, performed well compared to newer antipsychotics. This finding is important, because in low-and middle-income countries second-generation antipsychotics may not be affordable. However, older studies with negative results could have remained more frequently unpublished in contrast to nowadays where all clinical trials and their results should be registered. In an analysis of all antipsychotics compared to placebo, contour-enhanced funnel plots suggested unpublished studies (Fig. 6b).

Our analysis had limitations. We used strict inclusion criteria to obtain a homogenous sample, but nevertheless, the included studies have been conducted in an over sixty-year period, during which study characteristics changed. Checking for consistency revealed only few inconsistent loops and low-to-moderate heterogeneity in most outcomes (appendix 13), but the overall power to detect inconsistency is low.<sup>16</sup> The major exceptions were quality of life, for which a NMA was not calculated, and prolactin increase. As prolactin results may depend on the laboratory assay used, we calculated SMDs in addition to WMDs (appendix 20.4), which strongly reduced heterogeneity. Still, the fact that clozapine and zotepine significantly reduced prolactin compared to placebo (with wide credible intervals)

may be a statistical artefact, because only two small trials were available. The most important threat to the transitivity assumption of NMA was the increase of placebo response over the years,<sup>27,34</sup> because adjusting for placebo response in a meta-regression strongly reduced heterogeneity ( $\tau$ ) by 60%-63% (Appendix 9.1). In this meta-regression model the ranking was not substantially different from the primary analysis. Removing placebo arms, placebo-controlled studies and failed studies in sensitivity analyses did not materially change the results. Nor did metaregressions of six other moderators and further sensitivity analyses, speaking for the robustness of the findings. The NMA results were overall consistent with those of pairwise meta-analyses (Table 3) and single studies. For example, a study comparing brexpiprazole with placebo and quetiapine found that brexpiprazole was better than placebo, but worse than quetiapine, similar to the hierarchy of our analysis.<sup>35</sup> In a long-term study (sponsored by asenapine's manufacturer) olanzapine was significantly better than asenapine.<sup>36</sup> Thus, we do not believe that placebo response explains all efficacy differences between the compounds. Nevertheless, it is possible that the statistical methods could not fully account for the heterogeneity.<sup>27</sup>

Our decision to exclude studies from mainland China limits the generalisability of the results to this country. A recent literature and telephone interview study revealed that the majority of Chinese trials continue to be of low quality.<sup>14</sup> Chinese reports are usually very short and as communication with the authors is often difficult due to language problems, risk of bias is difficult to assess. We had therefore a priori decided to exclude Chinese studies.

Clinical trials exclude suicidal patients and the severely ill are unlikely to be included in modern trials as giving informed consent is often not possible for them. With a mean duration of illness of twelve years, our sample consisted mainly of chronic patients, which are known to respond worse compared to first-episode patients.<sup>37</sup> These factors reduce generalisability.

For feasibility reasons our risk of bias assessment focussed on the primary outcome, but strictly speaking risk of bias is outcome specific (appendix 19). Moreover, the evidence for many secondary outcomes, e.g. social functioning, was based on much lower sample sizes compared to the primary outcome (42576 vs. 5199).

These limitations clearly limited the strength of the derivable recommendations. This is most clearly, albeit not only, the case for older antipsychotics, as their effect sizes are based mainly on one or two studies with sample sizes lower than 100. Small sample sizes leave room for small-trial effects which may have inflated some results. For example,

the large effect of clozapine concerning reduction of negative symptoms is based on 63 participants, as clozapine is mainly studied in treatment-resistant patients which were excluded from the analysis. The contribution of indirect evidence is large for older drugs, resulting in large credible intervals, higher uncertainty and lower confidence in the evidence evaluated by CINeMA (Fig 5). The generally lower number of data available for old drugs, except for perphenazine which had more evidence of good quality from a large trial,<sup>31</sup> is highlighted in the figures and must thus be taken into account in the interpretation of all findings.

As so many antipsychotic options are available, our results should help to find the most suitable drug for the individual patient balancing side-effect profiles and efficacy of different drugs. We confirm that antipsychotics differ more in their side-effects than in their efficacy. We believe that efficacy differences between compounds exist, but it is a problem that their measurement is based on subjective rating scales. The development of objective efficacy measures would make interpretation easier. Finally, clinicians must remember, that the reported results are averages and that response and side effects may vary considerably in individual patients.

## ***Contributors***

SL obtained funding. SL, GS, AN and MH designed the study. MH and SL set up the database. MH, MS, JS, MK, NP, LB, TA, PR, SL screened the literature search, acquired reports of relevant trials, selected included studies and extracted data. MH and SL contacted trial investigators for additional information. AN and GS performed all statistical analyses. MH, AC, JD and SL analyzed and interpreted the data. MH and SL drafted the report. All authors critically reviewed the report for important intellectual content and approved the final submitted version.

## **Acknowledgement**

We thank all study authors that responded to our data requests. Data were sent by G. Chouinard, K. Hatta, R. Nielsen, M. Dossenbach, O. Lemming, and MK de Boer. We also thank the numerous researchers who sent information for our previous reviews on which the present one was built. The authors thank Samantha Roberts for help in the literature search, Yikang Zhu for help with study screening and Yikang Zhu, Matteo Rabaioli-Fischer, Susanne Bächer, Leonie Reichelt and Hannah Röder for help with data extraction.

## Declaration of interests

MH has received speaker's honoraria from Janssen and Lundbeck. SL has received honoraria for consulting from LB Pharma, Lundbeck, Otsuka, TEVA, Geodon Richter, Recordati, LTS Lohmann, and Boehringer Ingelheim; and for lectures from Janssen, Lilly, Lundbeck, Otsuka, SanofiAventis, and Servier. AC is supported by the National Institute for Health Research (NIHR) Oxford Cognitive Health Clinical Research Facility. All other authors declare no competing interests.

## References

- 1 Kassebaum NJ, Arora M, Barber RM, et al. Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet* 2016; **388**: 1603–58. doi:10.1016/S0140-6736(16)31460-X.
- 2 Research BBC. Antipsychotic Drugs: Technologies and Global Markets. Technologies and Global Markets. <https://www.bccresearch.com/market-research/pharmaceuticals/antipsychotic-drugs-markets-phm063a.html> (accessed Sep 06, 2018).
- 3 Leucht S, Cipriani A, Spineli L, et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia. A multiple-treatments meta-analysis. *Lancet (London, England)* 2013; **382**: 951–62. doi:10.1016/S0140-6736(13)60733-3.
- 4 Oehl M, Hummer M, Fleischhacker WW. Compliance with antipsychotic treatment. *Acta psychiatrica Scandinavica. Supplementum* 2000; 83–86.
- 5 WHOCC. WHOCC - ATC/DDD Index - Antipsychotics. [https://www.whocc.no/atc\\_ddd\\_index/?code=N05A](https://www.whocc.no/atc_ddd_index/?code=N05A) (accessed Jul 19, 2018).
- 6 Chisholm D, Gureje O, Saldivia S, et al. Schizophrenia treatment in the developing world: an interregional and multinational cost-effectiveness analysis. *Bulletin of the World Health Organization* 2008; **86**: 542–51. doi:10.2471/BLT.07.045377.
- 7 Higgins JPT, Welton NJ. Network meta-analysis. A norm for comparative effectiveness? *Lancet (London, England)* 2015; **386**: 628–30. doi:10.1016/S0140-6736(15)61478-7.
- 8 Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions. Checklist and explanations. *Annals of internal medicine* 2015; **162**: 777–84. doi:10.7326/M14-2385.
- 9 Leucht S, Huhn M, Rothe P, Schneider J, Zhu Y. Which are the most important first-generation antipsychotic drugs? Survey of international schizophrenia experts. *npj Schizophrenia* 2016; **2**: 25. doi:10.1038/npjSchz2016.8.
- 10 Gardner DM, Murphy AL, O'Donnell H, Centorrino F, Baldessarini RJ. International consensus study of antipsychotic dosing. *The American journal of psychiatry* 2010; **167**: 686–93. doi:10.1176/appi.ajp.2009.09060802.
- 11 Higgins JPT, ed. Cochrane handbook for systematic reviews of interventions. Chichester: Wiley-Blackwell, 2011.
- 12 Leucht S, Tardy M, Komossa K, et al. Antipsychotic drugs versus placebo for relapse prevention in schizophrenia: a systematic review and meta-analysis. *Lancet (London, England)* 2012; **379**: 2063–71. doi:10.1016/S0140-6736(12)60239-6.
- 13 McMahon RP, Kelly DL, Boggs DL, et al. Feasibility of reducing the duration of placebo-controlled trials in schizophrenia research. *Schizophrenia Bulletin* 2008; **34**: 292–301. doi:10.1093/schbul/sbm152.

- 14 Tong Z, Li F, Ogawa Y, Watanabe N, Furukawa TA. Quality of randomized controlled trials of new generation antidepressants and antipsychotics identified in the China National Knowledge Infrastructure (CNKI): a literature and telephone interview study. *BMC Med Res Methodol* 2018; **18**: 96. doi:10.1186/s12874-018-0554-2.
- 15 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**: 7–10. doi:10.1016/j.jclinepi.2005.06.006.
- 16 Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *The Lancet* 2018; **391**: 1357–66. doi:10.1016/S0140-6736(17)32802-7.
- 17 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.
- 18 Plummer M, Stukalov A, Denwood M. Rjags: Bayesian Graphical Models using MCMC (version 4-6) 2016. <https://CRAN.R-project.org/package=rjags>.
- 19 Rhodes KM, Turner RM, Higgins JPT. Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. *Journal of clinical epidemiology* 2015; **68**: 52–60. doi:10.1016/j.jclinepi.2014.08.012.
- 20 Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis. Concepts and models for multi-arm studies. *Res.Synth.Meth.* 2012; **3**: 98–110.
- 21 Dias S, Welton NJ, Marinho VCC, Salanti G, Higgins JPT, Ades AE. Estimation and adjustment of bias in randomized evidence by using mixed treatment comparison meta-analysis. *J.Roy.Statist.Soc.A* 2010; **173**: 613–29.
- 22 Gerta Rücker, Guido Schwarzer, Ulrike Krahn, Jochem König, Orestis Efthimiou. netmeta: Network Meta-Analysis using Frequentist Methods 2018.
- 23 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. *Journal of clinical epidemiology* 2008; **61**: 991–96.
- 24 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**: 455–63.
- 25 Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JPT. Evaluating the Quality of Evidence from a Network Meta-Analysis. *PloS one* 2014; **9**. doi:10.1371/journal.pone.0099682.
- 26 Cinema: Confidence in Network Meta-Analysis. Bern: Institute of Social and Preventive Medicine, 2017.
- 27 Leucht S, Leucht C, Huhn M, et al. Sixty Years of Placebo-Controlled Antipsychotic Drug Trials in Acute Schizophrenia. Systematic Review, Bayesian Meta-Analysis, and Meta-Regression of Efficacy Predictors. *The American journal of psychiatry* 2017: appiajp201716121358. doi:10.1176/appi.ajp.2017.16121358.
- 28 Krause M, Zhu Y, Huhn M, et al. Antipsychotic drugs for patients with schizophrenia and predominant or prominent negative symptoms: a systematic review and meta-analysis. *European archives of psychiatry and clinical neuroscience* 2018. doi:10.1007/s00406-018-0869-3.
- 29 Howard R, Cort E, Bradley R, et al. Antipsychotic treatment of very late-onset schizophrenia-like psychosis (ATLAS): a randomised, controlled, double-blind trial. *The Lancet Psychiatry* 2018; **5**: 553–63. doi:10.1016/S2215-0366(18)30141-X.
- 30 BNF 76. London: BMJ Group; PHARMACEUTICAL PRESS, 2018.
- 31 Lieberman JA, Stroup TS, McEvoy JP, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *The New England journal of medicine* 2005; **353**: 1209–23. doi:10.1056/NEJMoa051688.
- 32 McIntyre RS. Understanding needs, interactions, treatment, and expectations among individuals affected by bipolar disorder or schizophrenia: the UNITE global survey. *The Journal of clinical psychiatry* 2009; **70 Suppl 3**: 5–11. doi:10.4088/JCP.7075su1c.02.

- 33 Rummel-Kluge C, Komossa K, Schwarz S, et al. Head-to-head comparisons of metabolic side effects of second generation antipsychotics in the treatment of schizophrenia: a systematic review and meta-analysis. *Schizophrenia research* 2010; **123**: 225–33. doi:10.1016/j.schres.2010.07.012.
- 34 Agid O, Siu CO, Potkin SG, et al. Meta-regression analysis of placebo response in antipsychotic trials, 1970-2010. *The American journal of psychiatry* 2013; **170**: 1335–44. doi:10.1176/appi.ajp.2013.12030315.
- 35 Marder SR, Hakala MJ, Josiassen MK, et al. Brexpiprazole in patients with schizophrenia: overview of short- and long-term phase 3 controlled studies. *Acta neuropsychiatrica* 2017; **29**: 278–90. doi:10.1017/neu.2016.57.
- 36 Schoemaker J, Stet L, Vrijland P, Naber D, Panagides J, Emsley R. Long-term efficacy and safety of asenapine or olanzapine in patients with schizophrenia or schizoaffective disorder: an extension study. *Pharmacopsychiatry* 2012; **45**: 196–203. doi:10.1055/s-0031-1301310.
- 37 Zhu Y, Li C, Huhn M, et al. How well do patients with a first episode of schizophrenia respond to antipsychotics: A systematic review and meta-analysis. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology* 2017; **27**: 835–44. doi:10.1016/j.euroneuro.2017.06.011.