Psychological Medicine

cambridge.org/psm

Antipsychotic drugs *v.* barbiturates or benzodiazepines used as active placebos for schizophrenia: a systematic review and meta-analysis

Spyridon Siafis[1], Giacomo Deste[2], Anna Ceraso[2], Christian Mussoni[2], Antonio Vita[2], Senad Hasanagic[3], Johannes Schneider-Thoma[1], Georgios Papazisis[4], John M Davis[5,6] and Stefan Leucht[1]

[1]Department of Psychiatry and Psychotherapy, School of Medicine, Technische Universität München, München, Germany; [2]Department of Psychiatry, Spedali Civili Hospital, Brescia, Italy; [3]Medical Center of the University of Sarajevo, Sarajevo, Bosnia and Herzegovina; [4]Department of Clinical Pharmacology, School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece; [5]Department of Psychiatry, University of Illinois at Chicago, Chicago, USA and [6]Illinois and Maryland Psychiatric Research Center, Baltimore, Maryland, USA

Original Article

Cite this article: Siafis S *et al* (2020). Antipsychotic drugs *v.* barbiturates or benzodiazepines used as active placebos for schizophrenia: a systematic review and meta-analysis. *Psychological Medicine* **50**, 2622–2633. https://doi.org/10.1017/S003329171900285X

Received: 20 March 2019
Revised: 9 September 2019
Accepted: 18 September 2019
First published online: 18 October 2019

Key words:
First-generation; phenobarbital; phenothiazines; randomized; response

Author for correspondence:
Spyridon Siafis,
E-mail: spyridon.siafis@tum.de

Abstract

Background. Comparisons of antipsychotics with placebo can be biased by unblinding due to side effects. Therefore, this meta-analysis compared the efficacy of antipsychotics for acute schizophrenia in trials using barbiturates or benzodiazepines as active placebos.

Methods. Randomized controlled trials (RCTs) in acute schizophrenia with at least 3 weeks duration and comparing any antipsychotic with barbiturates or benzodiazepines were eligible. ClinicalTrials.gov, CENTRAL, EMBASE, MEDLINE, PsycINFO, PubMed, WHO-ICTRP as well as previous reviews were searched up to 9 January 2018. Two separate meta-analyses, one for barbiturates and one for benzodiazepines, were conducted using random-effects models. The primary outcome was response to treatment, and mean values of schizophrenia rating scales and dropouts were analyzed as secondary outcomes. This study is registered with PROSPERO (CRD42018086263).

Results. Seven barbiturate-RCTs (number of participants $n = 1736$), and two benzodiazepine-RCTs ($n = 76$) were included in the analysis. The studies were published between 1960 and 1968 and involved mainly chronically ill patients. More patients on antipsychotics in comparison to barbiturates achieved a 'good' response (36.2% *v.* 16.8%; RR 2.15; 95% CI 1.36–3.41; $I^2 = 48.9$) and 'any' response (57.4% *v.* 27.8%; RR 2.07; 95% CI 1.35–3.18; $I^2 = 68.2$). In a single small trial ($n = 60$), there was no difference between antipsychotics and benzodiazepines on 'any' response (74.7% *v.* 65%; RR 1.15; 95% CI 0.82–1.62).

Conclusions. Antipsychotic drugs were more efficacious than barbiturates, based on a large sample size. Response ratios were similar to those observed in placebo-controlled trials. The results on benzodiazepines were inconclusive due to the small number of studies and participants.

Introduction

A comprehensive meta-analysis of placebo-controlled randomized trials (RCTs) since the introduction of chlorpromazine in 1953 found that about twice as many patients with acute exacerbation of schizophrenia respond to antipsychotics in comparison to inert placebo (Leucht *et al.*, 2017). Double-blind RCTs are the gold standard for the assessment of treatments, but a major concern is the risk of unblinding due to antipsychotic side effects (Shader *et al.*, 1964; Leucht *et al.*, 2008, 2013). Such unblinding could lead to overestimates of the efficacy (Jensen *et al.*, 2017). In addition, one might argue that the mechanism of action of antipsychotics may be unspecific sedation rather than direct effects on positive symptoms (Moncrieff and Cohen, 2005), a message that if exaggerated could have a potential consequence of non-adherence leading to unnecessary psychotic relapses, and their negative social consequences.

The use of 'active placebos,' i.e. drugs that induce side effects but are not efficacious on positive symptoms, could provide more convincing results. A greatly discussed systematic review of active placebo RCTs questioned the efficacy of antidepressants based on inert placebo RCTs (Moncrieff *et al.*, 2004; Jensen *et al.*, 2017). Tricyclic antidepressants were found to be only slightly more efficacious for depression in comparison to atropine, a drug with anticholinergic side effects similar to those of tricyclic antidepressants (Moncrieff *et al.*, 2004). In schizophrenia trials conducted in the 1960s, barbiturates and later benzodiazepines were

CAMBRIDGE UNIVERSITY PRESS

used as active placebos because they can mimic antipsychotic-induced sedation (Casey *et al.*, 1960*b*; Holden *et al.*, 1968), and they both enhance $GABA_A$ receptor function. However, the comparison of antipsychotics with barbiturates has not been evaluated systematically before, and reviews on benzodiazepines were focused on short-term sedation or adjunctive treatment (Dold *et al.*, 2012). As a result, this review aims to compare the efficacy of antipsychotics with barbiturates or benzodiazepines as active placebos for acute schizophrenia.

Material and methods

We followed the PRISMA statement (checklist in eAppendix-1) (Liberati *et al.*, 2009) and the *a priori* written protocol was registered on PROSPERO (CRD42018086263).

Search strategy and selection criteria

RCTs comparing antipsychotics with barbiturates or benzodiazepines in patients with exacerbation of schizophrenia or schizophrenia-like psychosis were eligible. We used only the first phase of cross-over studies to avoid carry-over effects (Elbourne *et al.*, 2002). As in our previous review on antipsychotics *v.* inactive placebo (Leucht *et al.*, 2017), the minimum duration of follow-up was 3 weeks, a study duration that has been shown to be sufficient to show the full effects of antipsychotics (McMahon *et al.*, 2008). Any antipsychotic, barbiturate or benzodiazepine was included at any dose range and administered via any form of application, except for short-term intramuscular injections, which are used for sedation purposes. We excluded promazine and mepazine post-hoc, when we found that old reviews had already clearly shown that they are less efficacious than other antipsychotics (Davis *et al.*, 1989), but they were included in a sensitivity analysis of the primary outcome.

We searched ClinicalTrials.gov, the Cochrane Central Register of Controlled Trials, EMBASE, MEDLINE, PsycINFO, PubMed, and World Health Organization International Trial Registry on 9 January 2018 (search strategies in eAppendix-3). Two separate searches were conducted for barbiturates (no restriction in terms of publication date) and benzodiazepines [search for literature published after 2010 and older records were identified from the reference list of our previous Cochrane review (Dold *et al.*, 2012), for which extensive searches had been conducted]. Additional reviews (Klein and Davis, 1969; Wolkowitz and Pickar, 1991) and reference lists of included studies were inspected. At least two independent reviewers or contributors screened all title/abstracts from the search (SS, GA), full texts against the predefined eligibility criteria (SS, GP), extracted data in electronic forms, and assessed the quality of included studies using the Cochrane Collaboration's risk of bias tool (Higgins and Green, 2011; The Cochrane Collaboration, 2014) (SS, GD, AC, SH, CM, and JS). Any disagreements in all stages were resolved by consultation with a third reviewer (SL). The strength of the evidence of the primary outcome was assessed using the GRADE approach (Schünemann *et al.*, 2013).

Outcome variables

As in our previous meta-analysis of antipsychotics *v.* placebo, two response criteria were investigated, 'good' (primary outcome) and 'any' response (Leucht *et al.*, 2017). 'Good' response was defined as either 'at least much improvement' in the Clinical Global Impression scale (CGI) (Guy *et al.*, 1976) or 'at least 50% reduction from baseline of overall symptoms' as measured by published rating scales (Leucht *et al.*, 2005, 2007, 2012; Levine *et al.*, 2008), e.g. the older Lorr's Multidimensional Scale of Rating Psychiatric Patients (MSRPP) (Lorr *et al.*, 1953) and the Psychotic Reaction Profile (PRP) (Lorr *et al.*, 1960). 'Any' response was defined as at least minimal improvement in the CGI or at least 20% reduction of overall symptoms (Leucht *et al.*, 2005, 2007, 2012; Levine *et al.*, 2008). If these cut-offs were not available, then authors' definitions of response were also accepted. It has been shown that pooling studies with different definitions of response are acceptable as long as the effect size is presented as relative risks or odds ratios (Furukawa *et al.*, 2011). When responder rates were not reported, they were imputed from mean overall symptoms applying a validated method (Samara *et al.*, 2013).

Secondary outcomes were overall, positive, and negative symptoms as measured by published rating scales (Marshall *et al.*, 2000). Change from baseline to endpoint was preferred to follow-up scores of the these scales. Intention-to-treat data were used whenever available. Other secondary outcomes were premature discontinuation (dropouts) due to any cause, inefficacy, and adverse events. Missing standard deviations were estimated from reported statistics, or, if not possible, from other arms of the same study or from the average of other studies (Higgins and Green, 2011).

Statistical analysis

The rational to present the results on barbiturates and benzodiazepines in one paper is that they were both used as active placebos and they act via $GABA_A$ receptors. Nevertheless in the primary analysis, separate meta-analyses for barbiturates and benzodiazepines were calculated throughout using random-effects models (DerSimonian and Laird, 1986). The effect size for dichotomous outcomes was the relative risk or response ratio (RR) and its 95% confidence intervals (95% CI). Event rates and number-needed-to-treat for an additional beneficial/harmful outcome (NNTB/NNTH) were estimated according to the Cochrane Handbook (Higgins and Green, 2011), using the relative risk and the occurrence of an outcome in control groups as assumed control risk (see eAppendix-5). If the original authors presented only the results of the per protocol population, we conservatively assumed that participants who were lost to follow-up would not have responded. The effect size for continuous outcomes was the standardized mean difference expressed as Hedges' *g* and its 95% CI.

Assessment of heterogeneity was determined by visual inspection of forest plots and by applying a χ^2 test for homogeneity and the I^2 statistic (considerable heterogeneity when >50%) (Higgins and Green, 2011). Evaluation of the magnitude of heterogeneity was also supplemented using the empirical distribution of τ^2 of SMDs (Rhodes *et al.*, 2015) (see eAppendix-5). Based on available data, we conducted the following predefined subgroup or meta-regression analyses of the primary outcome: follow-up duration (⩽3 months *v.* longer-term), chlorpromazine equivalents (Gardner *et al.*, 2010), and baseline severity (MSRPP total score). Predefined sensitivity analyses of the primary outcome were the use of a fixed-effects model and exclusion of studies that presented per protocol data. Post-hoc sensitivity analyses were also conducted by including promazine/mepazine, excluding studies with imputed responder rates as well as sensitivity analyses of overall symptoms using different estimates of standard

https://doi.org/10.1017/S003329171900285X Published online by Cambridge University Press

deviations (see eAppendix-4). Following a reviewer request, we post-hoc pooled the results of the single benzodiazepine study with those of the barbiturates studies to obtain one estimate for the comparison of antipsychotics with GABAergic drugs. As data on the primary cut-off of 'good' response were not available, this analysis was only possible for the secondary cut-off 'any' response. Post-hoc analyses of the primary outcome were also conducted by comparing barbiturates with inert placebo and antipsychotics with mepazine. We investigated small study effects and publication bias with funnel plots if at least 10 studies were available (Egger *et al.*, 1997; Higgins and Green, 2011). The analyses were conducted with meta v4.9-2/5 (Schwarzer, 2007) and metafor v2.0-0 (Viechtbauer, 2010) packages in R statistical language v3.5 (R Core Team, 2018). The α was set at 0.05, except for heterogeneity at 0.1.

Results

Description of the included studies

Nine studies were included in the analysis, seven comparing barbiturates and two comparing benzodiazepines with antipsychotics for schizophrenia. The PRISMA flow diagram of the search is presented in Fig. 1 (Moher *et al.*, 2009). Study characteristics are displayed in Table 1 and in eAppendix-4. The included studies were published between 1960 and 1968 (median 1961), and were conducted in the USA, four in Veteran Administration Hospitals (VAH) (Casey *et al.*, 1960*a*, 1960*b*; Hollister *et al.*, 1960; Vestre *et al.*, 1962). No study concerning barbiturates was industry-sponsored, whereas both studies concerning benzodiazepines were sponsored by pharmaceutical industries manufacturing benzodiazepines. The median sample size was 80 (ranging from 24 to 805 participants) and the median follow-up duration was 12 weeks (ranging from 4 to 16 weeks). Two studies had a slightly longer follow-up duration (16 weeks) from 3 months (Hollister *et al.*, 1960; Clark *et al.*, 1961), and one <6 weeks (4 weeks) (Merlis *et al.*, 1962). Longer- and shorter-term results are presented together in the manuscript and separately in graphs in eAppendix-5.

The total number of participants (n) was 2099, 1328 on antipsychotics, 436 on barbiturates, 48 on benzodiazepines, and 287 on inert placebo or treatment combination. Almost all the studies included chronically ill patients, but not all were pretreated with antipsychotics (see Table 1 and eAppendix-4). No study examined exclusively treatment-resistant, first-episode patients, patients with predominant negative symptoms, or children and adolescents. The mean/median age of the participants ranged from 33 to 43 years. Five studies included only male patients (four of them from VAH), one included only females (Clark *et al.*, 1961), and in the other three, the ratios between genders were 1:1 (Merlis *et al.*, 1962; Gallant *et al.*, 1965) and 1:2 (Kurland *et al.*, 1961*a*). Ten antipsychotics were investigated: chlorpromazine (number of studies $N = 6$), fluphenazine ($N = 1$), trifluopromazine ($N = 3$), trifluoperazine ($N = 2$), mepazine ($N = 2$), promazine ($N = 2$), thioridazine ($N = 1$), perphenazine ($N = 2$), prochlorperazine ($N = 2$), and trifluperidol ($N = 1$). Excluding mepazine and promazine, the median of chlopromazine equivalents (Gardner *et al.*, 2010) was 500 mg/day (ranging from 150 to about 1250 mg/day). All seven studies on barbiturates used phenobarbital. The two studies on benzodiazepines used chlordiazepoxide ($N = 2$) and diazepam ($N = 1$) (Merlis *et al.*, 1962; Holden *et al.*, 1968).

Assessment of risk of bias

The risk of bias of the included studies is displayed in Fig. 2. As all studies reported randomization without adequately describing the method of random sequence generation as well as allocation concealment, they were given a rating of 'unclear' for these items. All studies were double-blind with a low risk for performance and detection bias in six of them, and unclear detection bias in three (Casey *et al.*, 1960*a*, 1961; Kurland *et al.*, 1961*a*; Vestre *et al.*, 1962; Gallant *et al.*, 1965; Holden *et al.*, 1968). Seven studies were judged to have an unclear risk of bias due to 'incomplete outcome data,' and two studies were assessed with a high risk of bias (Casey *et al.*, 1960*b*; Kurland *et al.*, 1961*a*, 1962). As it is typical for older trials, the studies, in general, were not well reported, such that four of them were rated to be of high risk of bias for selective reporting (Hollister *et al.*, 1960; Merlis *et al.*, 1962; Gallant *et al.*, 1965; Holden *et al.*, 1968). For example, standard deviations sometimes had to be estimated, and the exact number of participants was not always clear. Conservative extraction decisions were made throughout (details are reported in the characteristics of included studies in eAppendix-4). Finally, three studies were judged to be of low-risk of 'other bias,' and six studies of 'unclear risk.'

Antipsychotics v. barbiturates

Response to treatment

Antipsychotics were more efficacious than barbiturates in terms of the primary outcome 'good' response [36.2% of the patients on antipsychotics *v.* 16.8% on barbiturates; $N = 6$; $n = 1302$; RR 2.15; 95% CI (1.36–3.41); $I^2 = 48.9$; NNTB 5, 95% CI (2–17); *low quality of evidence according to GRADE*] (Fig. 3a). The results were not changed substantially in the sensitivity analyses using a fixed-effects model [RR 2.03 (1.57–2.62)] and including promazine and mepazine [$N = 6$; $n = 1676$; RR 1.98 (1.07–3.68), $I^2 = 71.1$; NNTB 6(2–84)]. Excluding studies with imputed response rates, few data remained, however, the results were still significant [$N = 3$; $n = 193$; RR 2.5 (1.07–5.84), $I^2 = 13.5$; NNTB 7(2–150)]. No significant difference between antipsychotics and phenobarbital was found in the small dataset after exclusion of studies with per protocol data for the primary outcome [$N = 2$; $n = 153$; RR 3.46 (0.44–27.05); $I^2 = 56.2$; NNTB 4 (0, NNTH: 18)]. Sensitivity and post-hoc analyses are presented in eAppendix-5.

In terms of the secondary cut-off, 'any' response, antipsychotics were more efficacious than phenobarbital and the response ratio was similar to that of 'good' response [57.4% *v.* 27.8%; $N = 7$; $n = 1362$; RR 2.07 (1.35–3.18); $I^2 = 68.2$; NNTB 3(2–10)] (Fig. 3b).

Overall, positive, and negative symptoms

Antipsychotics decreased overall symptoms more than phenobarbital on the total morbidity scores of MSRPP or PRP [$N = 4$; $n = 928$; SMD (95% CI) −0.56 (−0.96 to −0.16); $I^2 = 83.9\%$, $\tau^2 = 0.13$, high magnitude of heterogeneity] (Fig. 3c). Standard deviations were estimated using the significance threshold of 0.05 reported in three studies (Casey *et al.*, 1960*a*; Kurland *et al.*, 1961*a*, 1961*b*; Vestre *et al.*, 1962), which underestimated the difference, because the original p values were probably smaller. Therefore, we conducted post-hoc sensitivity analyses by using the smallest standard deviation within studies as well as the standard deviation of MSRPP derived by F-values in Casey *et al.* (1960*b*). The SMDs for overall

https://doi.org/10.1017/S003329171900285X Published online by Cambridge University Press

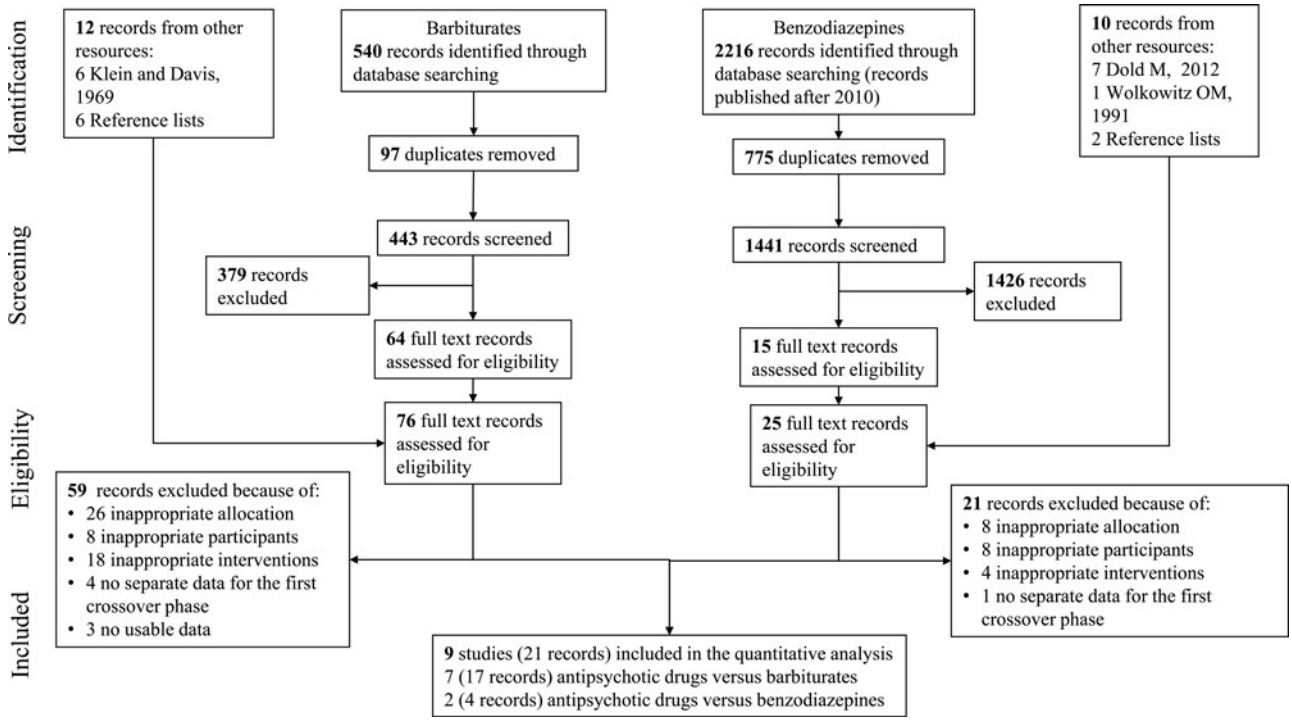


Fig. 1. PRISMA flow diagram of the study selection.

symptoms were larger [−0.73 (−0.95 to −0.50) and −0.82 (−1.01 to −0.62)] as well as less heterogeneity was observed ($I^2 = 49\%$, $\tau^2 = 0.02$ and $I^2 = 32\%$, $\tau^2 = 0.01$) in both scenarios, respectively (eAppendix-5).

Antipsychotics reduced positive symptoms more than phenobarbital, in terms of the subscales 'thinking disorder' of PRP [$N = 1$; $n = 82$; SMD −0.55 (−1.03 to −0.07)], 'perceptual disorganization' ($N = 1$; $n = 114$; SMD −0.58 (−1.05 to −0.12)], 'conceptual disorganization' [$N = 1$; $n = 114$; SMD −0.55 (−1.01 to −0.08)], and 'paranoid belligerence' of MSRPP [$N = 1$; $n = 114$; SMD −0.54 (−1.01 to −0.08)]. No data were available for negative symptoms.

Premature discontinuation

There was no difference between antipsychotics and phenobarbital in dropouts due to any cause [21.3% *v.* 24.5%; $N = 5$; $n = 1242$; RR 0.87 (0.64–1.17); $I^2 = 51$; NNTB 31(11, NNTH 24)] (Fig. 4a). Fewer patients on antipsychotics than phenobarbital discontinued due to inefficacy [3.3% *v.* 8.3%, $N = 5$; $n = 1242$; RR 0.39 (0.16–0.95); $I^2 = 62$; NNTB 20(14–241)] (Fig. 4b), whereas more patients on antipsychotic drugs than phenobarbital discontinued due to side effects [3% *v.* 1%; $N = 5$; $n = 1242$, RR 2.98 (1.12–7.96); $I^2 = 0$; NNTH 51(14, 833)] (Fig. 4c).

Subgroup analysis and meta-regressions of the primary outcome

Some heterogeneity was found on the primary outcome for the comparison between antipsychotics and barbiturates ($I^2 = 48.9\%$; $\chi^2_5 = 9.79$, $p = 0.08$). The *a priori* defined subgroup and meta-regression analyses found no significant differences between studies with shorter and longer follow-up duration ($N = 6$; $\chi^2_1 = 0.19$, $p = 0.66$), while greater response ratios were associated with higher mean doses of chlorpromazine equivalents ($N = 6$; slope = 0.0021; $z = 2.87$, $p = 0.004$) and lower mean baseline severity, as measured by total score of MSRPP ($N = 3$; slope = −0.1209; $z = -2.49$, $p = 0.013$) (eAppendix-5).

Small study effects and publication bias

The analysis of funnel plots was not meaningful because fewer than 10 studies were included in all comparisons.

Antipsychotics v. benzodiazepines

Data for this comparison were very limited. Of the two small studies available, only one provided useable data for response to treatment (Merlis *et al.*, 1962). Data on 'good' response, overall, positive, and negative symptoms were not available. In terms of 'any' response, there was no difference between antipsychotics and benzodiazepines [74.7% on antipsychotics *v.* 65% on benzodiazepines; $N = 1$; $n = 60$; RR 1.15 (0.82–1.62); NNTB 10 (2; NNTH 9); *low quality of evidence*] (Fig. 3b). A post-hoc analysis by pooling this benzodiazepine-controlled study with the barbiturate-controlled studies did not change the results materially [RR 1.81 (1.26–2.60)]. There was also no difference between antipsychotics and benzodiazepines in terms of dropouts due to any cause [$N = 1$; $n = 16$; RR 3(0.14–63.74)], inefficacy [$N = 1$; $n = 16$; RR 3(0.14–63.74)], and side effects ($N = 1$; $n = 16$; no dropouts due to side effects occurred) (Fig. 4). Due to the paucity of available data, subgroup, sensitivity, and meta-regression analyses were not meaningful for benzodiazepines.

Strength of the evidence for the primary outcome according to GRADE

We rated the strength of the evidence for both comparisons, antipsychotics *v.* barbiturates and *v.* benzodiazepines as low, mainly due to concerns about the risk of bias as well as imprecision of the results for the latter (eAppendix-6).

https://doi.org/10.1017/S003329171900285X Published online by Cambridge University Press

Table 1. Study characteristics of barbiturate- and benzodiazepine-controlled randomized trials (RCT) trials (RCT)

	Study	Design	Duration (weeks)	Diagnosis	Characteristic of included participants	Antipsychotic drugs (mg/day, *n*)	Active placebo (mg/day, *n*)	Other interventions (*n*)
Barbiturate-RCT	(Casey *et al.*, 1960*a*)	37 centers (VAH) Crossover	12	Schizophrenia (criteria n.i.)	*n* = 805 Sex (M/F): 805/0 Age: mean age 37y; up to 51y Pretreatment: 65% received 'tranquilizers'	CPZ: 400, *n* = 170* PROM†: 400, *n* = 171*	PHEN: 200, *n* = 173*	PLAC: *n* = 178*
	(Casey *et al.*, 1960*b*)	35 centers (VAH) Crossover	12	Schizophrenia (criteria n.i.)	*n* = 640 Sex (M/F): 640/0 Age: mean age 34y (18–54) Pretreatment: 66% received 'tranquilizers'	CPZ: 200–1200, *n* = 77* PERPH 16–96, *n* = 77* PROCH: 25–150, *n* = 83* TPRO: 50–300, *n* = 69* MEP†: 50–300, *n* = n.i.	PHEN: 32–192, *n* = n.i.	
	(Clark *et al.*, 1961)	Single center Parallel	16	Schizophrenia (DSM-I)	*n* = 60 Sex (M/F): 0/60 Age: mean age 43y (26–52) Pretreatment: 75% received 'chemotherapy'	CPZ: 200–800, *n* = 20	PHEN: 120–480, *n* = 20	PLAC: *n* = 20
	(Gallant *et al.*, 1965)	Single center Parallel	10	Schizophrenia (criteria n.i.)	*n* = 60 Sex (M/F): 30/30 Age: mean age 40.5y (21–59 y) Pretreatment: n.i.	TRIFLU: 4–6, *n* = 20 TPRE: 32–48, *n* = 20	PHEN: 120–180, *n* = 20	
	(Hollister *et al.*, 1960)	Single center (VAH) Parallel	16	Schizophrenia (criteria n.i.)	*n* = 60 Sex (M/F): 60/0 Age: median age 36y Pretreatment: 100% chlorpromazine for at least six months	CPZ: 100–900, *n* = 20 TPRE: 5–45, *n* = 20	PHEN: 32–288, *n* = 20	
	(Kurland *et al.*, 1961*a*, 1961*b*, 1962)	Single center Parallel	6	Newly admitted inpatients candidates for treatment with tranquilizers (predominately schizophrenia; criteria n.i.)	*n* = 277 Sex (M/F): ½ Age: mean age 39y (18–61) Pretreatment: n.i.	CPZ: 300–1200, *n* = 33 PROCH: 30–125, *n* = 32 TPRO: 75–300, *n* = 36 MEP†: 75–450, *n* = 34 PRO†: 300–1600, *n* = 32	PHEN: 97.5–360, *n* = 37	PLAC: *n* = 37
	(Vestre *et al.*, 1962)	Single center (VAH) Parallel	12	Schizophrenia (criteria n.i.)	*n* = 93 Sex (M/F): 93/0 Age: mean age 37 years (25–56 years) Pretreatment: 100% 'ataractic' medication	FLU: 2.5–25, *n* = 31 TPRO: 25–250, *n* = 31	PHEN: 32–320, *n* = 31	

Benzodiazepine-RCT	(Holden *et al.*, 1968)	Single center Crossover	8	Schizophrenia (criteria n.i.)	*n* = 24 Sex (M/F): 24/0 Age: mean age 33y (19 –44) Pretreatment: all of the patients were previously treated, n.i. previous medications	THIOR: 300–500, *n* = 8	CHD: 60–100, *n* = 8	THIOR + CHD: *n* = 8
	(Merlis *et al.*, 1962)	Single center Parallel	4	Psychosis (87.5% schizophrenia, criteria n.i)	*n* = 80 Sex (M/F): 40/40 Age: 14–62y Pretreatment: all of the patients received psychopharmacological agents (most frequently phenothiazines)	CPZ: 150, *n* = 20	CHD: 75, *n* = 20 DZ: 30, *n* = 20	PLAC: *n* = 20

CHD, chlordiazepoxide; CPZ, chlorpromazine; DZ, diazepam; FLU, fluphenazine; MEP, mepazine; PERPH, perphenazine; PHEN, phenobarbital; PLAC, inert placebo; PROCH, prochlorperazine; PROM, promazine; THIOR, thioridazine; TPRE, trifluoperazine; TPRO, triflupromazine; TRIFLU, trifluperidol; VAH, Veteran Administration Hospitals; *n*, number of participants; DSM, Diagnostic and Statistical Manual of Mental Disorders; M/F, males/females; y, years; *, completers' data when the randomized n was not available; n.i., not indicated; †, mepazine and promazine were excluded from the primary analysis.
All studies were randomized double-blind, involving inpatients. In the case of crossover studies, the duration of the first crossover phase is presented. Mean/median and range of age is presented.

Discussion

Antipsychotics were more efficacious than phenobarbital based on substantial evidence. In contrast, no difference compared to benzodiazepines was found based on one small study. Almost twice as many patients on antipsychotics in comparison to phenobarbital achieved a 'good' response or 'any' response. The number of included studies was small, but the results of the primary outcome were based on a considerable number of participants ($n = 1302$), which is higher than the threshold of 1000 participants suggested by Trikalinos *et al.* for the robustness of meta-analysis in psychiatry (Trikalinos *et al.*, 2004). Excluding studies analyzed 'per protocol' was the only sensitivity analysis with no significant difference between antipsychotics and phenobarbital, but it was based only on two studies with 153 participants. For these studies, we employed a conservative approach by assuming that dropouts had not responded; therefore, our primary analysis could be reliable from that point of view. Antipsychotics also decreased overall and positive symptoms with a medium effect size, and fewer patients than in the phenobarbital groups discontinued due to inefficacy. Whereas more patients on antipsychotics than on phenobarbital discontinued due to adverse effects, there were no differences in dropouts due to any cause.

About 72% of the total number of participants in the primary outcome stemmed from two industry-independent studies conducted in VAHs (Casey *et al.*, 1960*a*, 1960*b*). It is impressive that large, multicenter, double-blind RCTs in hospitalized patients were conducted in the 1960s without the support of modern technology, such as online communication (no facsimile, let alone e-mail), statistical software, as well as without industry funding. These early RCTs were milestones in clinical psychopharmacology (Shen, 1999; Carpenter and Davis, 2012). However, as these studies are from an era when trial methodology had not been established clearly yet, they did not meet all current methodological standards. Antipsychotic drugs were clearly more efficacious than phenobarbital, but due to these short comings, we downgraded the evidence, because there is uncertainty about the exact magnitude of the superiority of antipsychotics according to the GRADE approach (see eAppendix-6).

Our results comparing antipsychotics with phenobarbital were similar to those observed in our previous meta-analysis of RCTs with inactive placebo (Leucht *et al.*, 2017), where the response ratio was 1.96(1.65–2.44) for 'good' response (23% participants on antipsychotics had a 'good' response *v.* 14% on inert placebo) and 1.93(1.72–2.19) for 'any' response. The populations in the reviews were also similar, i.e. chronically ill patients, although it is likely that compared to modern trials, at that time, there were more antipsychotic-naïve patients. For example, in the two VAH studies (Casey *et al.*, 1960*a*, 1960*b*), only 65% of the participants had used 'tranquilizers' before the studies. Nevertheless, chronic patients might not respond as well to antipsychotics as first-episode patients (Zhu *et al.*, 2017*a*, 2017*b*). Indeed, in an early large NIMH study, almost half of the participants had their first episode or were antipsychotic-naive, and higher response rates were observed (61% patients on antipsychotics were much improved *v.* 22% on inert placebo) (National Institute of Mental Health Psychopharmacology Service Center Collaborative Study Group, 1964). The meta-analysis with inactive placebo-RCTs included many other antipsychotics than phenothiazines, in particular, haloperidol and second-generation antipsychotics (Leucht *et al.*, 2017). Despite these differences,

https://doi.org/10.1017/S003329171900285X Published online by Cambridge University Press

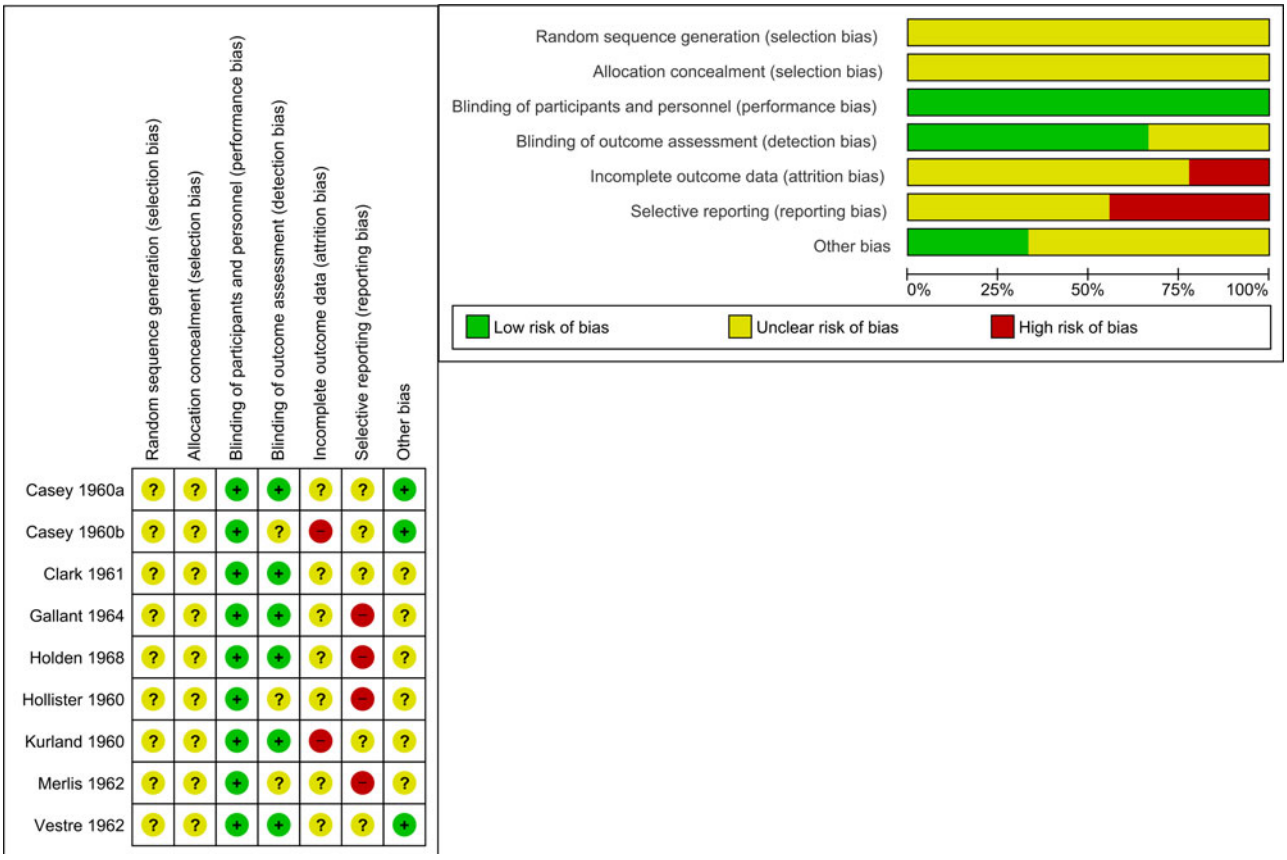


Fig. 2. Summary and graph of risk of bias of the included studies. The risk of bias was evaluated using the Cochrane's Risk of Bias Tool.

phenobarbital and inert placebo appear to be similarly ineffective compared to antipsychotics. Indeed, no difference between phenobarbital and inert placebo was found post-hoc in the primary outcome [$N = 3$; $n = 517$; RR 0.94 (0.68–1.30)] (eAppendix-5).

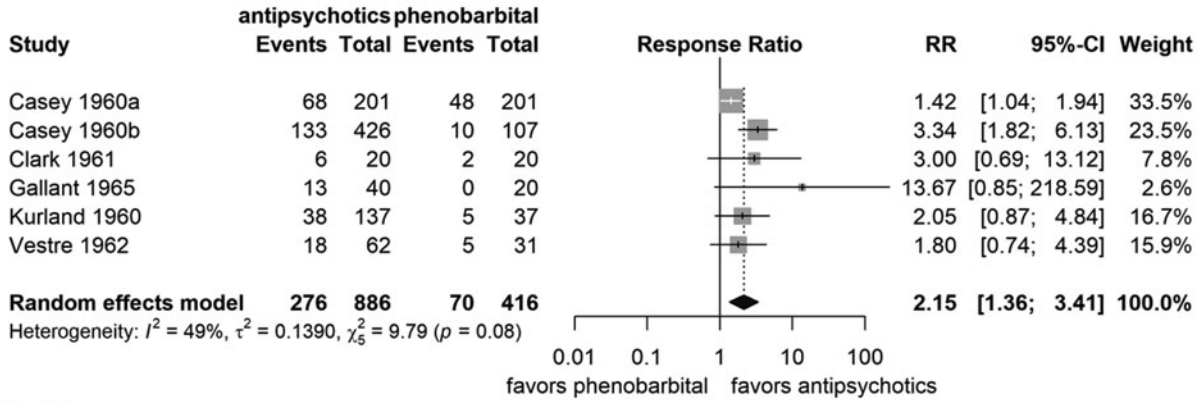
On the other hand, only two small trials were eligible for the comparison between antipsychotics and benzodiazepines (Merlis *et al.*, 1962; Holden *et al.*, 1968). There was no significant difference in response rates between antipsychotics and benzodiazepines based on one small study supported by the manufacturer of the benzodiazepines (Merlis *et al.*, 1962). The response rates were high in all arms (65% in inert placebo and benzodiazepines, 75% in chlorpromazine), although the daily dose of chlorpromazine was about 150 mg/day, which may be insufficient (Davis and Chen, 2004). Benzodiazepines are effective for sedating agitated patients with schizophrenia (Dold *et al.*, 2012), and a large, pragmatic RCT found that midazolam was effective more rapidly than combined haloperidol–promethazine in this regard (TREC Collaborative Group, 2003). However, as so few randomized data are available for longer-term outcomes, the results are inconclusive. We believe that further research is warranted, since benzodiazepines are used frequently in clinical practice, although they are liable for addiction, and observational studies suggest that benzodiazepine augmentation could be associated with increased mortality (Tiihonen *et al.*, 2012).

This review has some limitations. First, barbiturates and benzodiazepines induce sedation, and blinding could be jeopardized by antipsychotic-specific side effects, such as extrapyramidal symptoms (EPS). We did not analyze specific side effects, but four studies reported more EPS in the antipsychotic arms than under phenobarbital (Casey *et al.*, 1960*b*; Clark *et al.*, 1961; Kurland *et al.*, 1961*a*, 1961*b*; Gallant *et al.*, 1965; Hollister, 1972), while one study reported an overall low incidence of EPS (six EPS events in 805 patients, one in phenobarbital) (Casey *et al.*, 1960*a*) and another one, no significant difference in the use of anticholinergic drugs between antipsychotics and phenobarbital (Vestre *et al.*, 1962). We found that more patients on antipsychotics than phenobarbital dropped out due to side effects, but dropout rates were low (3% *v.* 1%, respectively), in comparison to 20% observed in more recent antipsychotic trials (Huhn *et al.*, 2019). Dropout rates in antipsychotic trials have increased since that early era of psychopharmacology (Wahlbeck *et al.*, 2001), which could reflect improvements in the quality of reporting, as well as changes in trial methodology, patient recruitment and ethical aspects (Wahlbeck *et al.*, 2001; Brunoni *et al.*, 2010).

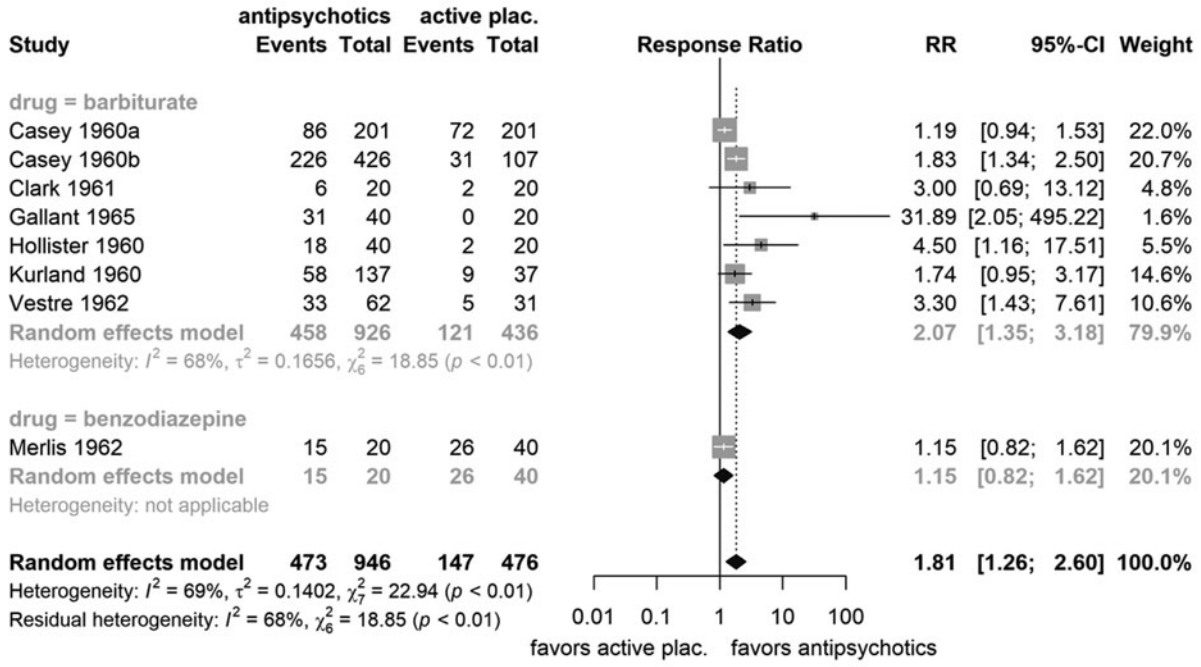
From a theoretical point of view, drugs which before study start were supposed to have antipsychotic properties would better control for investigator bias resulting from unblinding. Mepazine can serve as an example. It was introduced as a chlorpromazine-like phenothiazine with potential superiority to other phenothiazines (Casey *et al.*, 1960*b*). It might inhibit slightly dopaminergic receptors (Heiss *et al.*, 1976), somewhat similar to promethazine, but the dopamine blockade mechanism of antipsychotics had not been established at that time. In a post-hoc analysis, the other phenothiazines were more efficacious than mepazine [$n = 2$; $N = 704$; RR 1.78 (1.21–2.62)] (eAppendix-5), although it was investigated as a potential antipsychotic, reducing investigator bias (Casey *et al.*, 1960*b*; Kurland *et al.*, 1961*a*). In a similar vein, the phenothiazine

https://doi.org/10.1017/S003329171900285X Published online by Cambridge University Press

(*a*) 'Good' Response



(*b*) 'Any' Response



(*c*) Overall symptoms

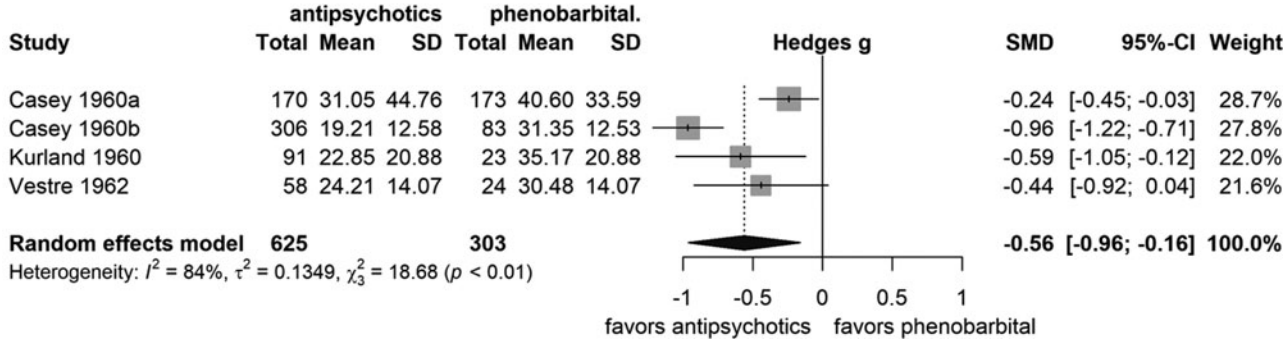


Fig. 3. Forest plots of (*a*) 'good' response, (*b*) 'any' response, (*c*) overall symptoms. Random-effects models were used as well as the Mantel–Haenszel method was used for 'good' and 'any' response. The weight of each study is reflected by the size of the square and the 95% confidence intervals by the associated error bars. The pooled effect (point estimate and 95% CI) is demonstrated with a black diamond. Regarding 'good' and 'any' response, antipsychotic drugs are superior to active placebo when the response ratio (RR) is greater than one. Regarding overall symptoms, antipsychotic drugs are superior to active placebo when the standardized mean difference (Hedges's g) is lower than zero. Heterogeneity between studies is quantified by the I^2 and χ^2 statistics. For overall symptoms, two scales had usable data MSRPP (Casey *et al*., 1960*a*, 1960*b* and Kurland *et al*., 1961*a*) and PRP (Vestre *et al*., 1962). Events: number of participants with 'good' or 'any' response to treatment, Total: total number of participants in the group, mean: the mean endpoint value of the total score of overall symptoms, SD: the standard deviation of the endpoint values of the total score.

https://doi.org/10.1017/S003329171900285X Published online by Cambridge University Press

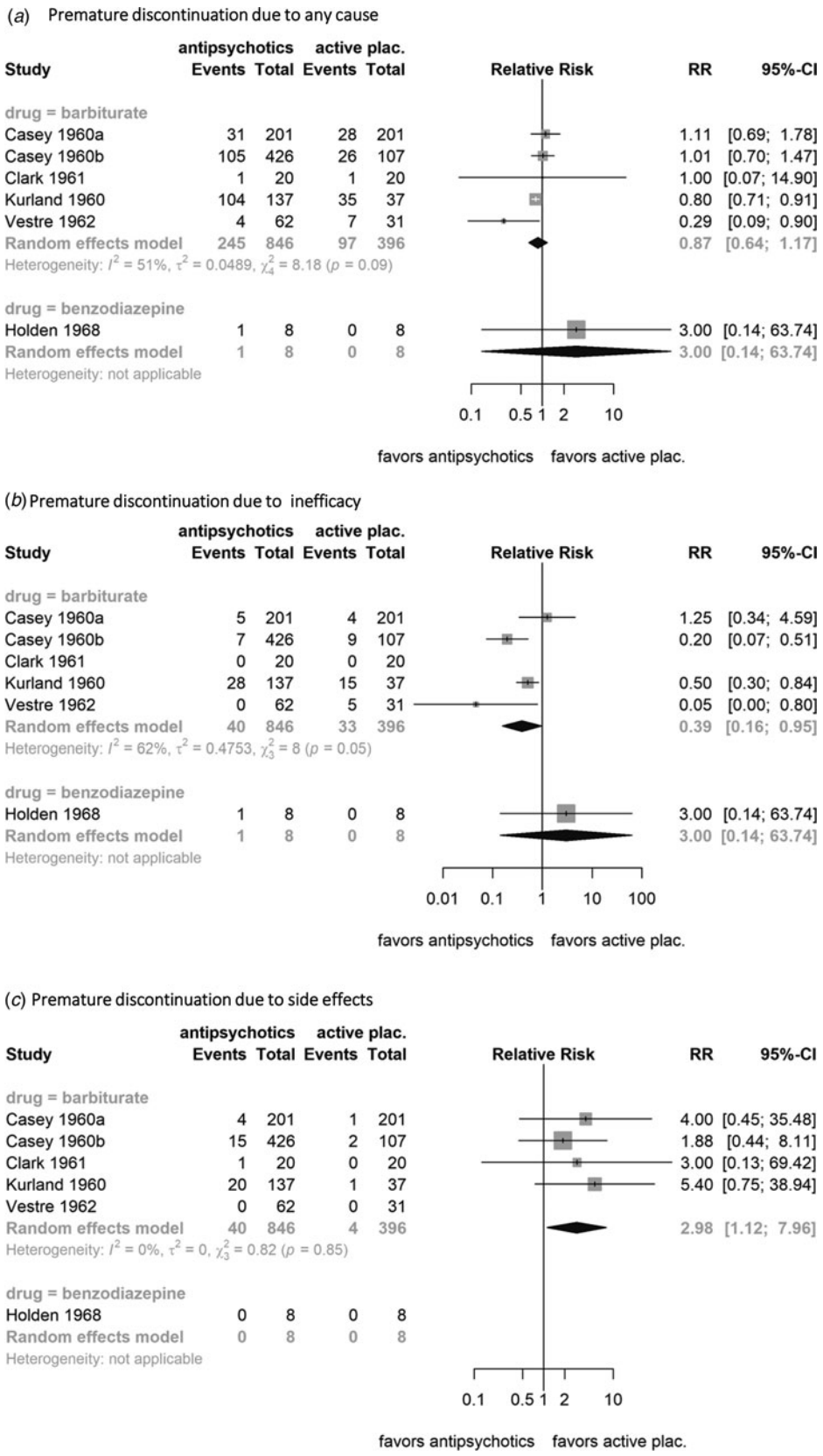


Fig. 4. Forest plots of premature discontinuation due to (*a*) any cause, (*b*) infefficacy, and (*c*) side effects. Random-effects models and the Mantel–Haenszel method was used. The weight of each study is reflected by the size of the square and the 95% confidence intervals by the associated error bars. The pooled effect (point estimate and 95% CI) is demonstrated with a black diamond. Antipsychotic drugs are superior to phenobarbital or benzodiazepines when the relative risk (RR) is lower than one. Heterogeneity between studies is quantified by the I^2 and χ^2 statistics. Events: number of participants who dropout, Total: total number of participants in the group.

perazine was superior to the tricyclic antidepressant trimipramine in a double-blind RCT, which attempted to prove trimipramine's efficacy, inspired by its receptor-binding profile being similar to clozapine (Bender *et al.*, 2003; Leucht *et al.*, 2014). Finally, phenothiazines were also superior to chlorprophenpyridamine/scopolamine (an active placebo mimicking sedative and anticholinergic properties) in a double-blind RCT with 288 participants (Adelson and Epstein, 1962).

https://doi.org/10.1017/S003329171900285X Published online by Cambridge University Press

Second, antipsychotics were not analyzed separately but were pooled together, since small differences in terms of efficacy could be expected (Leucht *et al.*, 2013; Samara *et al.*, 2014). Third, the included studies were published before the CONSORT statements for RCTs (Begg *et al.*, 1996); therefore, their reporting did not follow current standards and their quality was downgraded. Fourth, for this reason in some cases, estimation of values was mandatory, but conservative decisions were made throughout, which should have underestimated differences (eAppendix-4). For example, responder rates were imputed in three studies (Casey *et al.*, 1960*a*, 1960*b*; Kurland *et al.*, 1961*a*, 1961*b*) from mean values using a validated methodology (Samara *et al.*, 2013), which tends to narrow the difference between experimental and control groups (Samara *et al.*, 2013). Standard deviations were also poorly reported and conservative estimations were used in three studies (Casey *et al.*, 1960*a*; Kurland *et al.*, 1961*a*, 1961*b*; Vestre *et al.*, 1962), which underestimated the difference (see post-hoc sensitivity analyses of overall symptoms, eAppendix-5). These estimations, in part, also could explain the considerable heterogeneity in some secondary analyses and outcomes for which no clear other reasons were found.

Last, the results of the predefined meta-regressions on baseline severity and antipsychotic dose should be interpreted with most caution, because they were based on a maximum of six studies, and they were prone to ecological fallacy, potential outliers, and chance findings (eAppendix-5).

We conclude that antipsychotics were more efficacious than phenobarbital with medium effect sizes, based on old studies which are large but do not meet all current methodological standards so that the strength of the evidence about the exact magnitude of the superiority was low according to GRADE (Schünemann *et al.*, 2013). The data on benzodiazepines as 'active placebos' were inconclusive so that trials are still warranted, although possibly of less clinical importance due to their risk of addiction and excess mortality (Tiihonen *et al.*, 2012).

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S003329171900285X.

Acknowledgements. We would like to thank Farhad Shokraneh for conducting the electronic database search and assisting with the retrieval of full texts, Ghazaleh Aali for conducting the double screening of title and abstracts, and Makoto Wada for assisting with the translation of a Japanese full text. We would like to thank Dr Dimitris Mavridis for his statistical consultation regarding the empirical distributions of τ^2. SH was supported during his research internship at the Department of Psychiatry and Psychotherapy, School of Medicine, Technische Universität München, Germany by the European College of Neuropsychopharmacology (ECNP).

Author contributions. SL was the supervisor of the study. SL, SS, GP, and JD conceived and designed the study. SS, GM, AC, CM, SH, GP, JD, JS, and SL selected articles, extracted, and interpreted data. SS did the statistical analysis. SS and SL wrote the draft and the final version of the report. SL and AV obtained funding for the staff. SL, GP, and AV provided administrative, technical, or material support. All authors critically reviewed the report for important intellectual content and approved the final submitted version.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Conflict of interest. In the last 3 years, Stefan Leucht has received honoraria as a consultant/advisor and/or for lectures from LB Pharma, Otsuka, Lundbeck, Boehringer Ingelheim, LTS Lohmann, Janssen, Johnson&Johnson, TEVA, MSD, Sandoz, SanofiAventis, Angelini, Recordati, Sunovion, Geodon Richter. In the last 3 years, Antonio Vita received support directly or indirectly for clinical studies or trials, conferences, consultancies, congress presentation, and advisory boards from Angelini, Boheringer Ingelheim, Eli Lilly, Fidia, Forum Pharmaceutical, Innovapharma, Janssen-Cilag, Lundbeck, Otsuka, Recordati, Takeda. No other disclosures were reported.

References

Adelson D and Epstein L (1962) A study of phenothiazines with male and female chronically ill schizophrenic patients. *Journal of Nervous and Mental Disease* **134**, 543–554.

Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz K, Simel D and Stroup D (1996) Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *Journal of the American Medical Association* **276**, 637–639.

Bender S, Olbrich H, Fischer W, Hornstein C, Schoene W, Falkai P, Haarmann C, Berger M and Gastpar M (2003) Antipsychotic efficacy of the antidepressant trimipramine: a randomized, double-blind comparison with the phenothiazine perazine. *Pharmacopsychiatry* **36**, 61–69.

Brunoni AR, Tadini L and Fregni F (2010) Changes in clinical trials methodology over time: a systematic review of six decades of research in psychopharmacology. *PLoS ONE* **5**, e9479.

Carpenter W and Davis J (2012) Another view of the history of antipsychotic drug discovery and development. *Molecular Psychiatry* **17**, 1168–1173.

Casey J, Bennett I, Lindley C, Hollister L, Gordon M and Springer N (1960*a*) Drug therapy in schizophrenia. A controlled study of the relative effectiveness of chlorpromazine, promazine, phenobarbital, and placebo. *AMA Archives of General Psychiatry* **2**, 210–220.

Casey J, Lasky J, Klett C and Hollister L (1960*b*) Treatment of schizophrenic reactions with phenothiazine derivatives. *American Journal of Psychiatry* **117**, 97–105.

Clark M, Ray T, Paredes A, Costiloe J, Chappell J, Hagans J and Wold S (1961) Chlorpromazine in chronic schizophrenic women: I. Experimental design and effects at maximum point of treatment. *Psychopharmacologia* **2**, 107–136.

Davis J and Chen N (2004) Dose response and dose equivalence of antipsychotics. *Journal of Clinical Psychopharmacology* **24**, 192–208.

Davis J, Barter J and Kane J (1989) Antipsychotic drugs. In Baltimore MWW (ed.), *Comprehensive Textbook of Psychiatry*. Baltimore: Williams & Wilkins Co. pp. 1591–1626.

DerSimonian R and Laird N (1986) Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**, 177–188.

Dold M, Li C, Tardy M, Khorsand V, Gillies D and Leucht S (2012) Benzodiazepines for schizophrenia. *The Cochrane Database of Systematic Reviews* **11**, Cd006391.

Egger M, Smith G, Schneider M and Minder C (1997) Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* **315**, 629.

Elbourne D, Altman D, Higgins J, Curtin F, Worthington H and Vail A (2002) Meta-analyses involving cross-over trials: methodological issues. *International Journal of Epidemiology* **31**, 140–149.

Furukawa T, Akechi T, Wagenpfeil S and Leucht S (2011) Relative indices of treatment effect may be constant across different definitions of response in schizophrenia trials. *Schizophrenia Research* **126**, 212–219.

Gallant D, Bishop M, Nesselhof W and Sprehe D (1965) Further observations on trifluperidol: a butyrophenone derivative. *Psychopharmacologia* **7**, 37–43.

Gardner D, Murphy A, O'Donnell H, Centorrino F and Baldessarini R (2010) International consensus study of antipsychotic dosing. *American Journal of Psychiatry* **167**, 686–693.

Guy W, National Institute of Mental Health, Psychopharmacology Research Branch & Program ECDE (1976) *ECDEU assessment manual for psychopharmacology*. U.S. Dept. of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, National Institute of Mental Health, Psychopharmacology Research Branch, Division of Extramural Research Programs: Rockville, Md.

Heiss WD, Hoyer J and Thalhammer G (1976) Antipsychotic drugs and dopamine-mediated responses in Aplysia neurons. *Journal of Neural Transmission* **39**, 187–208.

Higgins JPT and Green S (2011) *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: Wiley and Sons, UK. (Version 5.1.0 [updated March 2011]).

Holden J, Itil T, Keskiner A and Fixk M (1968) Thioridazine and chlordiazepoxide, alone and combined, in the treatment of chronic schizophrenia. *Comprehensive Psychiatry* **9**, 633–643.

Hollister L (1972) Clinical use of psychotherapeutic drugs. I. Antipsychotic and antimanic drugs. *Drugs* **4**, 321–360.

Hollister L, Erickson G and Motzentecker F (1960) Trifluoperazine in chronic psychiatric patients. *Journal of Clinical & Experimental Psychopathology & Quarterly Review of Psychiatry and Neurology* **21**, 15–23.

Huhn M, Nikolakopoulou A, Schneider-Thoma J, Krause M, Samara M, Peter N, Arndt T, Backers L, Rothe P, Cipriani A, Davis J, Salanti G and Leucht S (2019) Comparative efficacy and tolerability of 32 oral antipsychotics for the acute treatment of adults with multi-episode schizophrenia: a systematic review and network meta-analysis. *Lancet* **394**(10202): 939–951.

Jensen JS, Bielefeldt AO and Hrobjartsson A (2017) Active placebo control groups of pharmacological interventions were rarely used but merited serious consideration: a methodological overview. *Journal of Clinical Epidemiology* **87**, 35–46.

Klein DF and Davis JM (1969) *Diagnosis and Drug Treatment of Psychiatric Disorders*. Baltimore: Williams & Wilkins.

Kurland A, Hanlon T, Tatom M, Ota K and Simopoulos A (1961*a*) The comparative effectiveness of six phenothiazine compounds, phenobarbital and inert placebo in the treatment of acutely ill patients: global measures of severity of illness. *Journal of Nervous and Mental Disease* **133**, 1–18.

Kurland A, Hanlon T, Tatom M and Simopoulos A (1961*b*) Comparative studies of the phenothiazine tranquilizers: methodological and logistical considerations. *Journal of Nervous and Mental Disease* **132**, 61–74.

Kurland A, Michaux M, Hanlon T, Ota K and Simopoulos A (1962) The comparative effectiveness of six phenothiazine compounds, phenobarbital and inert placebo in the treatment of acutely ill patients: personality dimensions. *Journal of Nervous and Mental Disease* **134**, 48–60.

Leucht S, Kane JM, Kissling W, Hamann J, Etschel E and Engel R (2005) Clinical implications of Brief Psychiatric Rating Scale scores. *British Journal of Psychiatry* **187**, 366–371.

Leucht S, Davis J, Engel R, Kane J and Wagenpfeil S (2007) Defining 'response' in antipsychotic drug trials: recommendations for the use of scale-derived cutoffs. *Neuropsychopharmacology* **32**, 1903–1910.

Leucht S, Heres S, Hamann J and Kane J (2008) Methodological issues in current antipsychotic drug trials. *Schizophrenia Bulletin* **34**, 275–285.

Leucht S, Engel R, Davis J, Kissling W, Meyer Zur Capellen K, Schmauss M and Messer T (2012) Equipercentile linking of the Brief Psychiatric Rating Scale and the Clinical Global Impression Scale in a catchment area. *European Neuropsychopharmacology* **22**, 501–505.

Leucht S, Cipriani A, Spineli L, Mavridis D, Orey D, Richter F, Samara M, Barbui C, Engel RR, Geddes JR, Kissling W, Stapf MP, Lassig B, Salanti G and Davis J (2013) Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *Lancet* **382**, 951–962.

Leucht S, Helfer B and Hartung B (2014) Perazine for schizophrenia. *Cochrane Database of Systematic Reviews*, (1), CD002832.

Leucht S, Leucht C, Huhn M, Chaimani A, Mavridis D, Helfer B, Samara M, Rabaioli M, Bacher S, Cipriani A, Geddes JR, Salanti G and Davis J (2017) Sixty years of placebo-controlled antipsychotic drug trials in acute schizophrenia: systematic review, Bayesian meta-analysis, and meta-regression of efficacy predictors. *American Journal of Psychiatry* **174**, 927–942.

Levine SZ, Rabinowitz J, Engel R, Etschel E and Leucht S (2008) Extrapolation between measures of symptom severity and change: an examination of the PANSS and CGI. *Schizophrenia Research* **98**, 318–322.

Liberati A, Altman D, Tetzlaff J, Mulrow C, Gotzsche P, Ioannidis J, Clarke M, Devereaux P, Kleijnen J and Moher D (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *British Medical Journal* **339**, b2700.

Lorr M, Jenkins R and Holsopple J (1953) Multidimensional scale for rating psychiatric patients. *Veterans Administration Technical Bulletin* Volume VI, Series 10–507, 1–44.

Lorr O'Connor J and Stafford J (1960) The psychotic reaction profile. *Journal of Clinical Psychology* **16**, 241–245.

Marshall M, Lockwood A, Bradley C, Adams C, Joy C and Fenton M (2000) Unpublished rating scales: a major source of bias in randomised controlled trials of treatments for schizophrenia. *British Journal of Psychiatry* **176**, 249–252.

McMahon R, Kelly D, Boggs D, Li L, Hu Q, Davis J and Carpenter W (2008) Feasibility of reducing the duration of placebo-controlled trials in schizophrenia research. *Schizophrenia Bulletin* **34**, 292–301.

Merlis S, Turner WJ and Krumholz W (1962) A double-blind comparison of diazepam, chlordiazepoxide and chlorpromazine in psychotic patients. *Journal of Neuropsychiatry and Clinical Neurosciences* **3**(suppl. 1), S133–S138.

Moher D, Liberati A, Tetzlaff J and Altman D (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine* **6**, e1000097.

Moncrieff J and Cohen D (2005) Rethinking models of psychotropic drug action. *Psychotherapy & Psychosomatics* **74**, 145–153.

Moncrieff J, Wessely S and Hardy R (2004) Active placebos versus antidepressants for depression. *Cochrane Database of Systematic Reviews* (1), Cd003012.

National Institute of Mental Health Psychopharmacology Service Center Collaborative Study Group (1964) Phenothiazine treatment in acute schizophrenia: effectiveness. *Archives of General Psychiatry* **10**, 246–261.

R Core Team (2018) R: a language and environment for statistical computing.

Rhodes KM, Turner RM and Higgins JPT (2015) Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. *Journal of Clinical Epidemiology* **68**, 52–60.

Samara M, Spineli L, Furukawa T, Engel R, Davis J, Salanti G and Leucht S (2013) Imputation of response rates from means and standard deviations in schizophrenia. *Schizophrenia Research* **151**, 209–214.

Samara M, Cao H, Helfer B, Davis J and Leucht S (2014) Chlorpromazine versus every other antipsychotic for schizophrenia: a systematic review and meta-analysis challenging the dogma of equal efficacy of antipsychotic drugs. *European Neuropsychopharmacology* **24**, 1046–1055.

Schünemann H, Brożek J, Guyatt G and Oxman A (2013) *GRADE Handbook for Grading Quality of Evidence and Strength of Recommendations*. The GRADE Working Group. https://community.cochrane.org/style-manual/references/reference-types/cochrane-publications#handbook.

Schwarzer G (2007) Meta: an R package for meta-analysis. *R News* **7**, 40–45.

Shader R, Cohler J, Elashoff R and Grinspoon L (1964) Phenobarbital and atropine in combination, an active control substance for phenothiazine research. *Journal of Psychiatric Research* **2**, 169–183.

Shen W (1999) A history of antipsychotic drug development. *Comprehensive Psychiatry* **40**, 407–414.

The Cochrane Collaboration (2014) Review Manager (RevMan). (ed. C. T. N. C. Centre).

Tiihonen J, Suokas JT, Suvisaari JM, Haukka J and Korhonen P (2012) Polypharmacy with antipsychotics, antidepressants, or benzodiazepines and mortality in schizophrenia. *Archives of General Psychiatry* **69**, 476–483.

TREC Collaborative Group (2003) Rapid tranquillisation for agitated patients in emergency psychiatric rooms: a randomised trial of midazolam versus haloperidol plus promethazine. *British Medical Journal* **327**, 708–713.

Trikalinos T, Churchill R, Ferri M, Leucht S, Tuunainen A, Wahlbeck K and Ioannidis J (2004) Effect sizes in cumulative meta-analyses of mental health randomized trials evolved over time. *Journal of Clinical Epidemiology* **57**, 1124–1130.

Vestre N, Hall W and Schiele B (1962) A comparison of fluphenazine, trifluopromazine, and phenobarbital in the treatment of chronic schizophrenic patients: a double-blind controlled study. *Journal of Clinical & Experimental Psychopathology & Quarterly Review of Psychiatry and Neurology* **23**, 149–159.

Viechtbauer W (2010) Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software* **36**, 1–48.

Wahlbeck K, Tuunainen A, Ahokas A and Leucht S (2001) Dropout rates in randomised antipsychotic drug trials. *Psychopharmacology* **155**, 230–233.

https://doi.org/10.1017/S003329171900285X Published online by Cambridge University Press

Wolkowitz O and Pickar D (1991) Benzodiazepines in the treatment of schizophrenia: a review and reappraisal. *American Journal of Psychiatry* **148**, 714–726.

Zhu Y, Krause M, Huhn M, Rothe P, Schneider-Thoma J, Chaimani A, Li C, Davis J and Leucht S (2017*a*) Antipsychotic drugs for the acute treatment of patients with a first episode of schizophrenia: a systematic review with pairwise and network meta-analyses. *The Lancet. Psychiatry* **4**, 694–705.

Zhu Y, Li C, Huhn M, Rothe P, Krause M, Bighelli I, Schneider-Thoma J and Leucht S (2017*b*) How well do patients with a first episode of schizophrenia respond to antipsychotics: a systematic review and meta-analysis. *European Neuropsychopharmacology* **27**, 835–844.

https://doi.org/10.1017/S003329171900285X Published online by Cambridge University Press