The effect of differences in trial design on estimates of efficacy of olanzapine in randomized studies

Psychiatry Research 336 (2024) 115895

Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

The effect of differences in trial design on estimates of efficacy of olanzapine in randomized studies

F. Rotella [a,*], A. Falone [a], E. Cassioli [a], E. Mannucci [b], V. Ricca [a], F. Del Monaco [a]

[a] Department of Health Sciences, University of Florence, Psychiatry Unit, AOU Careggi, Largo Brambilla 3, 50134, Florence, Italy
[b] Department of Health Sciences, University of Florence, Diabetology, AOU Careggi, Largo Brambilla 3, 50134, Florence, Italy

ARTICLE INFO

Keywords:
Schizophrenia
Antipsychotics
Pharmacotherapy
Meta-analysis

ABSTRACT

Background: Differences in trial design may affect estimates of efficacy of psychotropic drugs. The purpose of this meta-analysis is to evaluate whether the use of Olanzapine (OLZ) as either investigational or control drug affects the observed efficacy of OLZ.
Methods: We performed a search for Randomized-Controlled Trials (RCTs) in which the efficacy of OLZ is assessed in patients with schizophrenia or schizoaffective disorder. We assessed overall efficacy of OLZ and performed subgroup analyses of studies with OLZ as intervention or comparator. Mixed-effect meta-regression analyses were performed.
Results: Of the 25 RCTs included, OLZ was considered as investigational drug or active control in 13 and 12 studies, respectively. The reduction of PANSS score was greater in trials in which OLZ was used as investigational drug. Multivariate meta-regression models showed that a higher PANSS score at baseline and trial duration were the main predictors of greater PANSS score reduction.
Conclusions: Trials with OLZ used as investigational drug differ from those of trials with OLZ as comparator for baseline PANSS scores and study duration; these differences may produce differences in estimates of efficacy. As a consequence, the severity of illness at enrollment and trial duration should be carefully considered to ensure the reliability of indirect comparisons among antipsychotics.

1. Introduction

Olanzapine (OLZ), a Second-Generation Antipsychotic (SGA) widely used in the treatment of Schizophrenia (Duggan et al., 2005), has been reported to have a higher efficacy and tolerability than First-Generation Antipsychotics (FGA), at least in the short and medium term (Solmi et al., 2017). Available evidence on the clinical effects of antipsychotics mostly derives from Randomized Controlled Trials (RCTs) specifically designed for submission to regulatory authorities (Huhn et al., 2019).

Recently, a third wave of antipsychotic drugs has been released (namely aripiprazole, brexpiprazole, cariprazine, asenapine, lurasidone). Their efficacy is comparable to "older" SGAs, but with a higher tolerability; for this reason, they are sometimes referred to as Third Generation Antipsychotics (TGAs). The authors are aware of the non-univocal nomenclature used for this new wave of antipsychotics, but for simplicity decided to use the term TGA throughout the paper.

SGAs, including OLZ, which had been used as investigational drugs in RCTs designed for their development, have also been used as comparator drugs in (usually more recent) RCTs designed for the development of TGAs (Greger et al., 2021; Jindal et al., 2013). Conversely, few RCTs directly compared TGAs with FGAs (Girgis et al., 2011; Kane et al., 2010); our knowledge of differences between TGAs and FGAs is mostly derived from inferences based on comparisons of FGAs with SGAs, and SGAs with TGAs. Such indirect evidence can be affected by peculiarities in the design of those latter comparisons: differences between trials using SGAs either as investigational drugs compared with FGAs or as comparators for TGAs could distort estimates of indirect comparisons between TGAs and FGAs.

SGAs and TGAs have both been claimed to be at least as effective as, or more effective than, FGAs, and to have a higher tolerability; TGAs display a higher tolerability than SGA, especially in the long term (McDonagh et al., 2020; Leucht et al., 2013). However, the generalizability of the results derived from RCTs to real-world practice has been questioned (Katona et al., 2021; Younis et al., 2020), because of the enrollment of small and non-representative samples and of the use of fixed doses (Katona et al., 2021). In addition, the number of available

* Corresponding author.
E-mail address: francesco.rotella@unifi.it (F. Rotella).

https://doi.org/10.1016/j.psychres.2024.115895
Received 3 June 2023; Received in revised form 3 April 2024; Accepted 5 April 2024
Available online 12 April 2024

RCTs with an adequately long follow-up is insufficient to draw consistent conclusions on long-term safety and tolerability (Rotella et al., 2020).

In order to obtain a complete overview of all available evidence, the results of RCTs can be combined in a meta-analysis. Heterogeneity across studies is one of the main issues of this approach, as RCTs with the same endpoint can display wide methodological differences (e.g., superiority or non-inferiority design, diversity in inclusion and exclusion criteria, different length of follow-up, the same drug used as investigational or comparator treatment, etc.).

The aim of this study is to assess the possible effect on estimates of OLZ efficacy based on the trial design, comparing studies in which OLZ was used either as an investigational drug or comparator, and adjusting for other potential confounders.

2. Methods

The reporting of this systematic review and meta-analysis follows PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) guidelines (Page et al., 2021). The protocol of this study was registered in PROSPERO (registration number: CRD42022276894). Database search, selection process and data extraction were performed independently by two of the authors (A.F. and F.D.M.), and conflicts were resolved by a third investigator (E.C.).

2.1. Literature search and selection process

We searched Medline, Embase and clinicaltrials.gov for superiority, non-inferiority, or equivalence designed RCT in which efficacy of OLZ is assessed, published up to Sept 15, 2021, using the following search string: "(olanzapine) AND (schizophrenia OR schizoaffective OR psychotic OR psychosis OR delusional)", with the filters "Randomized Controlled Trial", "English", and "Humans". We used EndNote™ 20 to merge database searches and delete duplicates. Eligibility criteria were screened on the basis of title and abstracts (Supplementary Material, Table 1); selected items were examined in full text, and excluded in case of: 1) Effect size not available or computable, and not provided by contacted authors; 2) Absence of a PANSS assessment of OLZ efficacy; 3) Impossibility to state if OLZ was the test drug or the active comparator, due to the absence of funding or a clear declaration by authors regarding the study design.

2.2. Data extraction

The following data for each study were collected: publication year, OLZ group sample size, OLZ group mean age, OLZ group female %, OLZ group PANSS at baseline, the variable "tested drug"/"active comparator" referred to OLZ, superiority/non-inferiority/equivalence design, flexible/fixed OLZ doses, mean OLZ dose, end-point (weeks), name of the drug compared to OLZ, its mean dosage, observed mean change from baseline to end-point (end of randomized treatment) in PANSS scores as effect size measure and standard deviations. To determine if OLZ was the tested drug or the active control, statements in which the information was precisely provided by authors were searched or, alternatively, data about the presence of funding. In case of the absence of these data, we considered OLZ as the tested drug if a FGA was the other drug studied, or the active control if the other drug was a third-generation antipsychotic. OLZ compared with a SGA, in the absence of funding or more precise information by the authors of the study, was a condition of exclusion due to the impossibility of establishing whether OLZ is a tested or a comparator drug. If in one study there were independent arms with different doses of OLZ, we considered the arm with the maximum dosage. If missing, SDs values were computed using conversion formulas from p-value, confidence interval, or T values; if this was not possible, authors were contacted to obtain missing data.

2.3. Quality assessment

The risk of bias was assessed using the Cochrane Collaboration's tool (Higgins et al., 2019) assigning an evaluation of "low risk", "high risk" or "uncertain risk" of bias concerning the following items: random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other bias. The risk of publication bias was evaluated through the interpretation of the funnel plot symmetry and with Egger's test (Egger et al., 1997).

2.4. Statistical analysis

The extracted data were presented in an aggregated form using weighted means and standard deviations, or frequencies/percentages where appropriate. Comparisons were made between groups (OLZ as intervention vs OLZ as comparator) on these descriptive measures, employing the chi-square test for frequencies, the Mann-Whitney U test for trial duration, and the meta-regression analysis described below for age, OLZ dosage, and baseline PANSS score.

A meta-analysis on the efficacy of OLZ was performed using random-effects models (Borenstein et al., 2010), using reduction of PANSS (Kay et al., 1987) total score as the main outcome (negative values correspond to a reduction from baseline). For each subgroup (OLZ as intervention, OLZ as comparator), overall effects were computed together with 95 % confidence intervals. Between-study heterogeneity was tested using Cochran's Q-test; heterogeneity variance ($\tau 2$) and the I^2 statistic were also computed. A test for subgroup differences was performed to verify if the overall effect of OLZ was different when used as a comparator or intervention.

Mixed-effects meta-regression analyses were used to test the additional role of putative moderators in determining heterogeneity. Study publication year, trial duration, and PANSS baseline scores were tested as moderators (both individually and in a multivariate comprehensive model), keeping OLZ group as a predictor in all models. Furthermore, to investigate any differences in terms of baseline PANSS scores between the two OLZ subgroups (OLZ as intervention, OLZ as comparator), baseline PANSS scores were also meta-analyzed, with OLZ group and year of the study as predictors. A further post-hoc analysis was performed comparing the effects of OLZ on PANSS (versus baseline) in trials using OLZ either as an investigational drug or comparator, after excluding non-sponsored trials. All statistical analyses were performed using R Statistical Software version 4.2.1 (R Core Team, 2022) and the "metafor" package (Viechtbauer, 2010).

3. Results

The study flow summary is reported in Fig. 1. Of the 1548 items identified after the exclusion of duplicates, 1496 were excluded on the basis of title and abstract, and 27 after the analysis of full text. Data were therefore extracted from 25 studies.

The characteristics of the studies and demographic information are summarized in Table 1. The global population treated with OLZ consisted of 4546 patients, with mean PANSS scores at baseline ranging from 80.5 to 110.5. Net of missing data, the weighted mean age was 37.78 years, with 36 % of females. Mean dosage of OLZ was 13.42 mg (SD 1.89) and mean length of trials was 17.12 weeks. In 13 studies OLZ was considered tested drug and in 12 studies as active control. Studies were powered for superiority of test drug in 6 cases, whereas 5 trials were designed as non-inferiority studies; the superiority or non-inferiority design was not reported in publications in 14 studies. Twenty-three studies claimed they had received funding from pharmaceutical companies. The study by Jindal et al. (2013) declared the absence of funding and conflicts of interest, therefore OLZ was considered the comparator due to aripiprazole being the tested drug. In the study by Ishigooka et al. (2001) it was clearly stated that OLZ was the

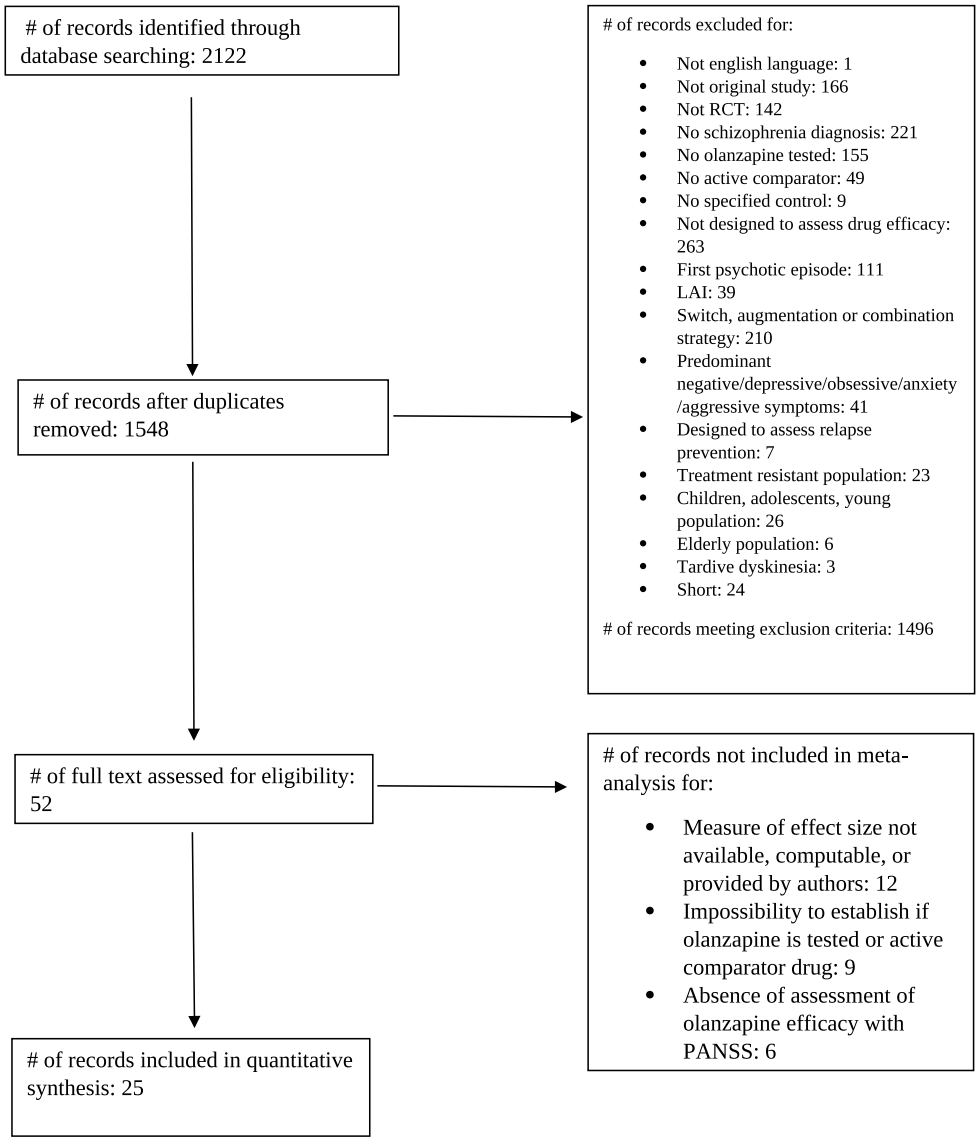

Fig. 1. Study flow chart.

tested drug.

The characteristics of studies in which OLZ was used either as investigational drug or comparator are summarized in Table 2. Trials in which OLZ was used as investigational drug showed a significantly ($p = 0.037$) longer duration of treatment and baseline PANSS scores ($p = 0.005$); no other characteristic of trials showed significant differences between groups of trials (all $p > 0.05$).

3.1. Meta-analysis

Treatment with OLZ was associated with a reduction of PANSS score with respect to baseline. A relevant heterogeneity was detected, with $I^2 > 90$ %. The mixed-effects Egger's Test was not statistically significant ($z = -1.26$, $p = 0.209$), indicating no relationship between the observed effect sizes and standard errors, and suggesting symmetry in the funnel plot, which is illustrated in Supplementary Fig. 1. PANSS reduction with OLZ versus baseline was significantly greater in trials using the drug as investigational treatment than in those using it as comparator (Fig. 2); heterogeneity was high in both subgroups of trials.

Only two studies resulted to be non-sponsored, one with OLZ as comparator and one with OLZ as investigational drug; PANSS reduction from baseline with OLZ was greater in trials using OLZ as investigational drugs even after the exclusion of those two non-sponsored trials (difference estimate $= -10.03$, $Q_M = 13.57$, df $= 1$, $p < 0.001$).

In order to explore the mechanisms underlying this heterogeneity and to test whether the efficacy of OLZ was significantly different based on whether it was used as an intervention or as a comparator, meta-regression analyses were performed. The first analysis added "study year of publication" as an additional moderator (Fig. 3A and B). The results show that a greater PANSS reduction was observed in more recent trials (Fig. 3A), with a high between-study heterogeneity (Fig. 3B). The second meta-regression analysis used "mean baseline PANSS scores" as additional moderator (Fig. 3A and C). The results show that a greater PANSS reduction was observed in trials that enrolled subjects with higher mean baseline PANSS scores (Fig. 3C), with high between study heterogeneity (Fig. 3D). A third meta-regression highlighted a negative, although non-significant, coefficient for trial length as a predictor of PANSS change over time (Fig. 3E). The coefficient for the "OLZ as intervention" subgroup was significantly different than zero when "study year of publication" was used as moderator but was not different than zero when "mean baseline PANSS scores" or duration of trial were used as moderators (Fig. 3B, D, F). Therefore, the reduction of PANSS scores with OLZ was significantly greater in trials using OLZ as intervention, when compared to those using the drug as comparator,

Table 1
Characteristics of included studies.

Study	Year	% female OLZ group	OLZ group sample size	OLZ group mean age	OLZ test/ control	OLZ dosage	Other Drug	Other Drug sample size	Study design (superiority/non inferiority)	Sponsor	OLZ group PANSS baseline scores	End point (weeks)	National setting
Kane et al. (2009)	2009	31.0	281	38.3	test	16.7	Aripiprazole	285	superiority	Lilly	95.7 ± 15.9	28	North and South America, Australia
Shen et al. (2014)	2014	35.2	71	40.1	control	15.0	Vabicaserin	70	not specified	Pfizer	94.5 ± 11.7	6	United States
Mortimer et al. (2004)	2004	35.6	188	37.4	control	12.5 ± 7.5	Amisulpride	186	non-inferiority	Sanofi	93.2 ± 16.0	24	EU, UK, Switzerland, Morocco, Tunisia
Schmidt et al. (2012)	2012	47.0	93	38.6	control	15.0	JNJ-37,822,681	69	not specified	Janssen	91.0 ± 11.1	6	Bulgaria, Estonia, Korea, Lithuania, Malaysia, Romania, Russia, South Africa, Taiwan, Ukraine
Conley and Mahmoud (2001)	2001	27.0	186	38.9	control	12.5 ± 7.5	Risperidone	175	not specified	Janssen	81.2 ± 13.5	8	United States
Jindal et al. (2013)	2013	/	27	50.0	control	11.0 ± 2.1	Aripiprazole	26	not specified	–	99.6 ± 11.9	6	India
Landbloom et al. (2017)	2017	39.1	45	40.8	control	15.0	Asenapine	111	superiority	Forest Research Institute	92.7 ± 10.5	6	USA, Bulgaria, Croatia, Romania, Russia, Ukraine
Kongsakon et al. (2006)	2006	49.0	139	32.7	test	10.2 ± 4.6	Haloperidol	123	superiority	Lilly	104.2 ± 28.1	24	Philippines, Pakistan, Malaysia, Thailand, Singapore
Tollefson et al. (2001)	2001	32.2	90	/	test	20.0 ± 5.0	Clozapine	87	non-inferiority	Lilly	108.2 ± 15.7	18	Western Countries, South Africa
Tran et al. (1997)	1997	/	172	/	test	15.0 ± 5.0	Risperidone	165	superiority	Lilly	96.3 ± 17.0	28	Western Countries, South Africa
Dossenbach et al. (2004)	2004	53.3	27	/	test	11.7 ± 3.0	Fluphenazine	27	not specified	Lilly	110.5 ± 16.4	22	Croatia
Marder et al. (2007)	2007	/	110	/	control	10.0	Paliperidone ER	111	not specified	Johnson & Johnson	94.9 ± 12.4	6	United States
Schoemaker et al. (2010)	2012	41.0	297	36.2	control	15.0 ± 5.0	Asenapine	869	not specified	Schering-Plough Corporation and Pfizer	92.1 ± 13.4	52	Western Countries, Russia, South Africa
Meltzer et al. (2011)	2011	22.0	121	38.3	control	15.0	Lurasidone	118	not specified	Lurasidone Study Group	96.3 ± 12.2	6	United States, Colombia, Lithuania, India, Philippines
Dossenbach et al. (2007)	2007	25.3	80	30.9	test	12.5 ± 7.5	Chlorpromazine	40	not specified	Lilly	104.7 ± 18.8	6	Turkey, UAE, Morocco
Ishigooka et al. (2001)	2001	/	90	/	test	10.0 ± 5.0	Haloperidol	78	non-inferiority	–	88.3 ± 21.3	8	Japan
Revicki et al. (1999)	1999	34.2	600	38.0	test	12.5 ± 7.5	Haloperidol	228	not specified	Lilly	87.7 ± 18.9	52	Western Countries
Tollefson et al. (1997)	1997	38.7	1312	/	test	12.5 ± 7.5	Haloperidol	636	not specified	Lilly	90.1 ± 19.2	6	North America, Europe
Beasley et al. (1997)	1997	36.0	89	37.0	test	15.0 ± 2.5	Haloperidol	79	not specified	Lilly	105.6 ± 18.9	6	Western Countries
Breier et al. (2005)	2005	35.0	268	40.1	test	15.0 ± 5.0	Ziprasidone	261	not specified	Lilly	99.5 ± 18.5	28	Western Countries
Gureje et al. (2003)	2003	37.5	32	35.6	test	15.0 ± 5.0	Risperidone	33	superiority	Lilly	94.7 ± 18.4	30	Australia, New Zealand
Naber et al. (2005)	2005	40.0	57	32.9	test	15.0 ± 10.0	Clozapine	56	non-inferiority	Lilly	101.3 ± 22.7	26	Germany
Simpson et al. (2005)	2005	/	114	/	control	12.6	Ziprasidone	55	not specified	Pfizer	89.0 ± 16.9	6	
Lublin et al. (2009)	2009	/	22	/	control	15.0 ± 5.0	Ziprasidone	126	non-inferiority	Pfizer	80.5*	12	Denmark, Finland, Switzerland, Iceland
Grootens et al. (2011)	2011	14.0	35	23.1	control	15.0 ± 5.0	Ziprasidone	34	superiority	Pfizer	80.6 ± 13.4	8	The Netherlands, Belgium

* The standard deviation of baseline PANSS scores was not obtainable.

Table 2
Descriptive summary of trial characteristics.

	Overall OLZ (25)	OLZ test (13)	OLZ control (12)
Year	Median: 2005 IQR: 6	Median: 2005 IQR: 7.5	Median: 2010.5 IQR: 4.5
Female (%)	36.13 °	36.72 °,†	34.43 °,†
Population (n)	4546	3237	1309
Age (pM, pSD)	37.45 (3.01) *	37.28 (2.55) *,†	37.71 (3.56) *,†
Dosage (pM, pSD)	13.42 (1.89)	13.37 (1.98) †	13.57 (1.64) †
Study design	Superiority: 6 Non-inferiority: 5 Not specified: 14	Superiority: 4 Non-inferiority: 3 Not specified: 6	Superiority: 2 Non-inferiority: 2 Not specified: 8
PANSS score baseline (pM, pSD)	92.72 (5.94)	93.50 (6.08) ‡	90.78 (5.08) ‡
Endpoint -weeks (M)	17.12	21.69 ‡	12.17 ‡

° data obtained from 19 studies.
* data obtained from 18 studies.
† comparison not statistically significant ($p \geq 0.05$).
‡ comparison statistically significant ($p < 0.05$).

after adjusting for year of publication (Fig. 3B), but not after adjusting for baseline PANSS scores or trial length (Fig. 3D and F).

A further meta-regression analysis was performed to test the relationship between "mean baseline PANSS scores", "study year of publication" and the use of OLZ as intervention or comparator. This analysis was performed on 24 out of 25 included studies since the standard deviation of baseline PANSS scores was not obtainable (Lublin et al., 2009). Although baseline PANSS scores appeared to be higher in more recent studies (Fig. 3G), studies that used OLZ as comparator showed lower baseline PANSS scores and the association between "mean baseline PANSS scores" and "study year of publication" was no longer statistically significant when adjusting for the use of OLZ as either intervention or comparator (Fig. 3H).

Finally, when "study year of publication", "mean baseline PANSS scores", and "trial length" were all entered as moderators, together with intervention/comparator group, only baseline PANSS scores and trial duration emerged as statistically significant predictors of higher PANSS scores reduction (Fig. 4).

3.2. Risk of bias assessment

Supplementary Fig. 2 provided an overview of the risk of bias assessment. Overall, the quality of evidence was good due to the presence, apart from two cases (Lublin et al., 2009; Dossenbach et al., 2007) of double-blind randomized trials. Inadequate or missing description of random sequence generation and allocation concealment made assessment of risk of selection bias impossible in most of the studies. About half of the studies were affected by high drop-out rates, but in all the studies there was a good handling of incomplete outcome data thanks to the presence of an Intention-To-Treat (ITT) or Last-Observation-Carried-Forward (LOCF) analysis.

4. Discussion

This is the first study to investigate within the context of RCTs on antipsychotic drugs the potential role of their use as investigational or comparator as a moderator of efficacy. OLZ showed a higher efficacy in RCTs in which it has been studied as investigational drug than in those in which it has been used as comparator. This result could be considered

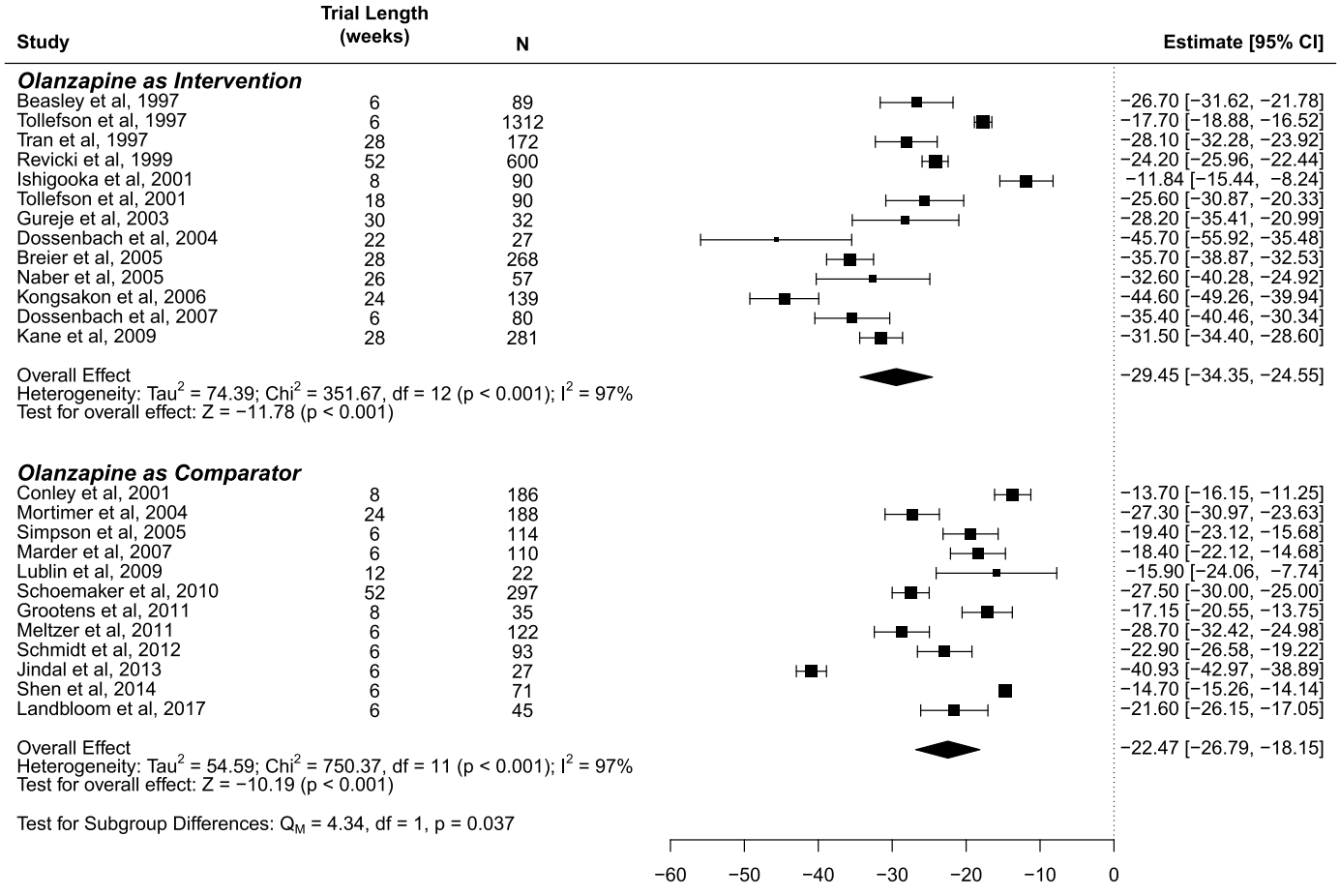

Fig. 2. Forest plots for mean PANSS reduction in trials that used Olanzapine as investigational drug (upper part) and as comparator (lower part).

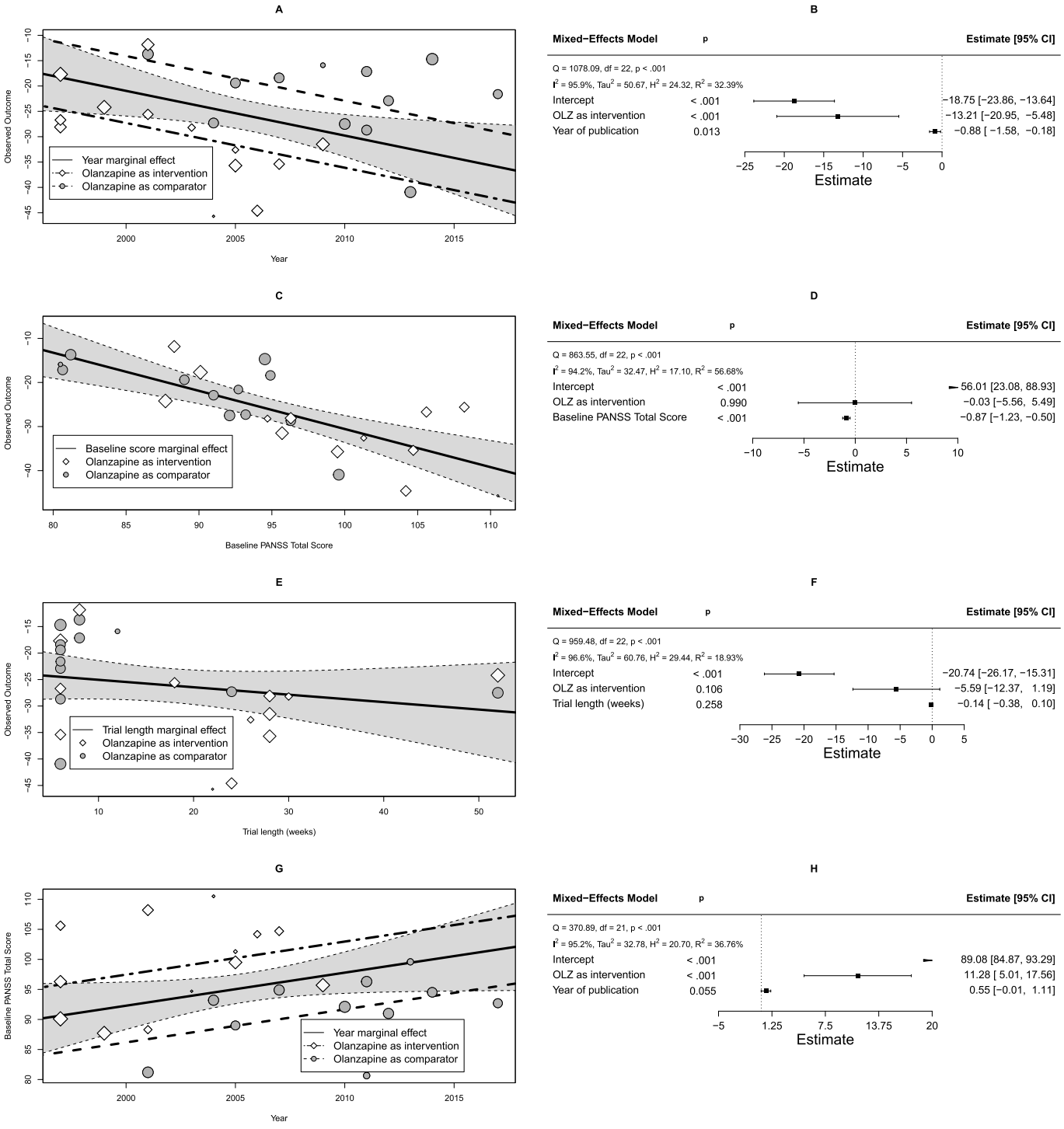

Fig. 3. Meta-regression models for mean PANSS reduction (panels A-F) and baseline PANSS Total Score (panels G-H), together with bubble plots illustrating moderators marginal effects and differences between the use of Olanzapine as investigational or comparator drug.

surprising, since the same therapy should be expected to have the same efficacy irrespective of the objective of the trial. Notably, most trials included in the analysis were blinded; therefore, the investigators' awareness of the treatment arm cannot have influenced the estimates of efficacy. On the other hand, differences in trial design or execution can affect study results. All included trials enrolled patients affected by schizophrenia or schizoaffective disorder; although definitions of these disorders have been revised over the time (American Psychiatric Association, 1987, 1994, 2013) it is unlikely that changes intervened in diagnostic criteria were sufficient to produce major differences in case mix.

Meta-regression analyses were performed in order to consider the possible effect of other confounding factors. A reasonable candidate is publication year: changes in the characteristics of patients with schizophrenia or schizoaffective disorder, or in the surrounding environment that occurred over the years, could theoretically affect the efficacy of therapeutic interventions. In addition, quality standards for RCTs have been progressively arisen over the years, affecting several aspects of trial design and execution. Most RCTs with OLZ as investigational drug were performed several years before those in which OLZ

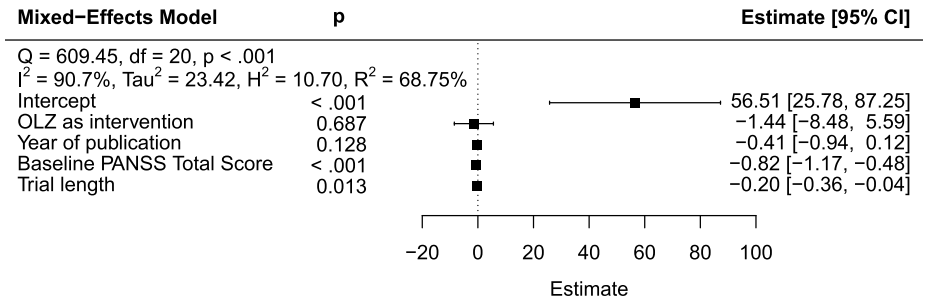

Fig. 4. Full meta-regression model for mean PANSS reduction.

was used as comparator for TGAs, possibly producing differences in estimated efficacy. Although the lag between trial completion and publication is variable, the year of publication can be considered a proxy for timing of trial design and execution. Notably, in meta-regression analyses, a greater efficacy of OLZ can be observed in more recent trials. However, in multivariate analysis both OLZ as investigational drug/comparator and year of publication are independently associated with reported efficacy of OLZ. This means that the greater efficacy of OLZ when used as investigational drug cannot be explained by the year of publication.

Another evident confounder is the severity of symptoms at baseline. Patients with higher PANSS scores can be expected to experience greater reductions in PANSS when undergoing treatment. In fact, in meta-regression analyses OLZ shows a greater efficacy in trials enrolling patients with higher mean PANSS scores. In a multivariate model, the difference between trials with OLZ as investigational or comparator drug is no longer statistically significant after adjusting for baseline PANSS. This suggests that the main determinant of this difference is the severity of symptoms at enrollment. Interestingly, mean PANSS scores are higher in RCTs with OLZ as investigational drugs than as comparator, but they do not appear to decrease over the years. Notably, the design of a trial can be affected by the principal aim of the study; we could expect that trials aimed at demonstrating a greater efficacy of the investigational drug (as the majority of those with olanzapine as investigational drug) select patients with higher symptom scores, in order to amplify differences across treatment arms. Conversely, trials aimed at showing non-inferiority on efficacy (as the majority of those with olanzapine as comparator) are more likely to enroll patients with milder symptoms, thus blunting possible differences in efficacy.

A further possible confounder is the duration of trials, with longer studies revealing a greater effect of treatment. Trials with OLZ as comparator were usually shorter than those with OLZ as investigational drug. It can be speculated that investigators designing non-inferiority trials could be more prone to plan a shorter duration of treatment, blunting possible differences between treatment arms.

In a further multivariate model including all variables reported above, the efficacy of OLZ in RCTs increases with baseline PANSS scores and duration of trial, without being affected by publication year or by the use of OLZ as investigational or comparator drug, confirming the central role of severity of symptoms at enrollment and trial duration in determining apparent efficacy.

Eight out of 25 trials identified maximum PANSS scores for patient inclusion. Since most of these trials (6 out of 8) use OLZ as comparator drug, this could have produced a difference in case mix enrolling patients with a lower severity. Since the reduction of symptoms from baseline is wider in case of greater severity, the enrollment of patients with milder symptoms reduced the probability of observing differences in efficacy between treatments when olanzapine was used as comparator.

Since trials comparing TGAs with OLZ (as comparator) enrolled patients with a lower severity of symptoms, the sensitivity of those studies in detecting differences in efficacy between treatment arms is somewhat smaller than that of trials comparing OLZ (as investigational drug) with FGAs. This means that the greater efficacy of TGAs with respect to FGAs cannot be assumed on the basis of their equivalence with SGAs. A recent network meta-analysis by Huhn et al. (2019) clearly showed that the evidence for comparisons between TGA and FGA is almost indirect.

The execution of a network meta-analysis, specifically designed for calculating indirect comparisons, including all trials with FGAs, SGAs, and TGAs, can be highly informative when there is a large number of different comparators, all included in a single network model. On the other hand, its reliability is weaker when the number of nodes in the network is smaller. In fact, one of the assumptions of a network meta-analysis is that the effects of a treatment in included trials are independent of comparators. In addition, the results of a network meta-analysis can be distorted by differences in the characteristics of included trials. We have shown here that the severity of symptoms in enrolled patients is greater in trials comparing OLZ with FGAs than in those comparing TGAs with OLZ; consequently, the reliability of a network meta-analysis would be highly questionable.

It should be recognized that other possible confounding factors, different from those considered in the present work, may have a role in determining observed results. Even if the year of publication of a trial can be considered a proxy of the general quality of the trial, as more recent trials are usually based on high-quality international standards, a correction for the risk of bias was not performed. Another possible confounding factor may be represented by the presence or absence of a sponsor. In fact, even small differences in inclusion/exclusion criteria, determined by the willingness to provide an advantage for the investigational drug, could produce a significant distortion of the results. However, in RCTs performed for submission to regulatory authorities, which are the large majority of those included in the present meta-analysis, inclusion and exclusion criteria have an impact on indications and contraindications, and therefore they cannot be designed for obtaining a more favorable result for the investigational drug. A further limitation is represented by the possibility of publication/reporting bias: although the analyses performed do not suggest such bias, the possibility that some trials with unfavorable results for the investigational drug were not reported cannot be completely ruled out.

5. Conclusions

Randomized trials that studied olanzapine as comparator enrolled patients with lower mean baseline PANSS scores and had a shorter duration compared to randomized trials that used olanzapine as investigational drug. These differences explain the apparent lower efficacy of olanzapine in these studies. As a consequence, the reliability of indirect comparisons between first- and third-generation antipsychotics is questionable. In order to have a clearer picture of their efficacy, specific trials with TGAs of appropriate duration, enrolling patients with severe schizophrenia or schizoaffective disorder, and/or subgroup analyses on patients with higher PANSS scores at enrollment, should be performed.

CRediT authorship contribution statement

F. Rotella: Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization. **A. Falone:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Data curation. **E. Cassioli:** Writing – review & editing, Writing – original draft, Supervision, Software, Methodology, Data curation, Conceptualization. **E. Mannucci:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **V. Ricca:** Supervision, Project administration, Methodology, Investigation, Conceptualization. **F. Del Monaco:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation.

Declaration of competing interest

Valdo Ricca received fees from agencies for speaking in symposia sponsored by Angelini, Janssen, Lundbeck, Otsuka, Viatris. Psychiatric Unit received fees from ArcaPharma and LundBeck. Edoardo Mannucci has received consultancy fees from Merck and Novartis, speaking fees from Abbott, Astra Zeneca, Bristol Myers Squibb, Boehringer-Ingelheim, Eli-Lilly, Merck, Novo Nordisk, Sanofi, and Novartis, and research grants from Merck, Novartis, and Takeda. Diabetology received fees from Astra Zeneca, Bayer, Boehringer-Ingelheim, Eli-Lilly, Janssen, Merck, Novartis and Novo Nordisk. Francesco Rotella, Emanuele Cassioli, Andrea Falone and Francesco Del Monaco declare no conflict of interests.

Acknowledgements

Francesco Rotella thanks Carlo Faravelli for his enlightening teachings that have stimulated the initial critical thoughts leading to the conceptualization and design of the present work.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.psychres.2024.115895.

References

American Psychiatric Association, 1987. Diagnostic and Statistical Manual of Mental Disorders—(DSM-III-R), third ed. American Psychiatric Association, Washington D. C. (revised).

American Psychiatric Association, 1994. Diagnostic and Statistical Manual of Mental Disorders—(DSM-IV), fourth ed. American Psychiatric Association, Washington D.C.

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders (fifth ed.). Arlington, VA: Author.

Beasley Jr., C.M., Hamilton, S.H., Crawford, A.M., Dellva, M.A., Tollefson, G.D., Tran, P. V., Blin, O., Beuzen, J.N., 1997. Olanzapine versus haloperidol: acute phase results of the international double-blind olanzapine trial. Eur. Neuropsychopharmacol. 7 (2), 125–137. https://doi.org/10.1016/s0924-977x(96)00392-6.

Borenstein, M., Hedges, L.V., Higgins, J.P., Rothstein, H.R., 2010. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res. Synth. Methods 1 (2), 97–111. https://doi.org/10.1002/jrsm.12.

Breier, A., Berg, P.H., Thakore, J.H., Naber, D., Gattaz, W.F., Cavazzoni, P., Walker, D.J., Roychowdhury, S.M., Kane, J.M., 2005. Olanzapine versus ziprasidone: results of a 28-week double-blind study in patients with schizophrenia. Am. J. Psychiatry 162 (10), 1879–1887. https://doi.org/10.1176/appi.ajp.162.10.1879.

Conley, R.R., Mahmoud, R., 2001. A randomized double-blind study of risperidone and olanzapine in the treatment of schizophrenia or schizoaffective disorder. Am. J. Psychiatry 158 (5), 765–774. https://doi.org/10.1176/appi.ajp.158.5.765.

Dossenbach, M.R., Folnegovic-Smalc, V., Hotujac, L., Uglesic, B., Tollefson, G.D., Grundy, S.L., Friedel, P., Jakovljevic, M.M., Olanzapine HGCH Study Group, 2004. Double-blind, randomized comparison of olanzapine versus fluphenazine in the long-term treatment of schizophrenia. Prog. Neuropsychopharmacol. Biol. Psychiatry 28 (2), 311–318. https://doi.org/10.1016/j.pnpbp.2003.10.010.

Dossenbach, M., Treuer, T., Kryzhanovskaya, L., Saylan, M., Dominguez, S., Huang, X., HGCQ, HGDQ, HGDT, and HGDV Study Team, 2007. Olanzapine versus chlorpromazine in the treatment of schizophrenia: a pooled analysis of four 6-week, randomized, open-label studies in the Middle East and North Africa. J. Clin. Psychopharmacol. 27 (4), 329–337. https://doi.org/10.1097/JCP.0b013e3180ca83b1.

Duggan, L., Fenton, M., Rathbone, J., Dardennes, R., El-Dosoky, A., Indran, S., 2005. Olanzapine for schizophrenia. Cochrane Database Syst. Rev. (2), CD001359 https://doi.org/10.1002/14651858.CD001359.pub2.

Egger, M., Davey Smith, G., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. BMJ 315 (7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629.

Girgis, R.R., Merrill, D.B., Vorel, S.R., Kim, Edward, Portland, Kimberly, You, M., Pikalov, A., Whitehead, R., Lieberman, J.A., 2011. Aripiprazole versus haloperidol treatment in early-stage schizophrenia. J. Psychiatr. Res. 45 (6), 756–762. https://doi.org/10.1016/j.jpsychires.2010.09.003.

Greger, J., Aladeen, T., Lewandowski, E., Wojcik, R., Westphal, E., Rainka, M., Capote, H., 2021. Comparison of the metabolic characteristics of newer second generation antipsychotics: brexpiprazole, lurasidone, asenapine, cariprazine, and iloperidone with olanzapine as a comparator. J. Clin. Psychopharmacol. 41 (1), 5–12. https://doi.org/10.1097/JCP.0000000000001318.

Grootens, K.P., van Veelen, N.M., Peuskens, J., Sabbe, B.G., Thys, E., Buitelaar, J.K., Verkes, R.J., Kahn, R.S., 2011. Ziprasidone vs. olanzapine in recent-onset schizophrenia and schizoaffective disorder: results of an 8-week double-blind randomized controlled trial. Schizophr. Bull. 37 (2), 352–361. https://doi.org/10.1093/schbul/sbp037.

Gureje, O., Miles, W., Keks, N., Grainger, D., Lambert, T., McGrath, J., Tran, P., Catts, S., Fraser, A., Hustig, H., Andersen, S., Crawford, A.M., 2003. Olanzapine vs risperidone in the management of schizophrenia: a randomized double-blind trial in Australia and New Zealand. Schizophr. Res. 61 (2–3), 303–314. https://doi.org/10.1016/s0920-9964(02)00226-8.

Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J., Welch, V.A. (Eds.), 2019. Cochrane Handbook for Systematic Reviews of Interventions, second ed. John Wiley & Sons, Chichester (UK).

Huhn, M., Nikolakopoulou, A., Schneider-Thoma, J., Krause, M., Samara, M., Peter, N., Arndt, T., Bäckers, L., Rothe, P., Cipriani, A., Davis, J., Salanti, G., 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. https://doi.org/10.1016/S0140-6736(19)31135-3.

Ishigooka, J., Inada, T., Miura, S., 2001. Olanzapine versus haloperidol in the treatment of patients with chronic schizophrenia: results of the Japan multicenter, double-blind olanzapine trial. Psychiatry Clin. Neurosci. 55 (4), 403–414. https://doi.org/10.1046/j.1440-1819.2001.00882.x.

Jindal, K.C., Singh, G.P., Munjal, V., 2013. Aripiprazole versus olanzapine in the treatment of schizophrenia: a clinical study from India. Int J Psychiatry Clin Pract 17 (1), 21–29. https://doi.org/10.3109/13651501.2011.653376.

Kane, J.M., Cohen, M., Zhao, J., Alphs, L., Panagides, J., 2010. Efficacy and safety of asenapine in a placebo- and haloperidol-controlled trial in patients with acute exacerbation of schizophrenia. J. Clin. Psychopharmacol. 30 (2), 106–115. https://doi.org/10.1097/JCP.0b013e3181d35d6b.

Kane, J.M., Osuntokun, O., Kryzhanovskaya, L.A., Xu, W., Stauffer, V.L., Watson, S.B., Breier, A., 2009. A 28-week, randomized, double-blind study of olanzapine versus aripiprazole in the treatment of schizophrenia. J. Clin. Psychiatry 70 (4), 572–581. https://doi.org/10.4088/jcp.08m04421.

Katona, L., Bitter, I., Czobor, P., 2021. A meta-analysis of effectiveness of real-world studies of antipsychotics in schizophrenia: are the results consistent with the findings of randomized controlled trials? Transl. Psychiatry 11 (1), 510. https://doi.org/10.1038/s41398-021-01636-9.

Kay, S.R., Fiszbein, A., Opler, L.A., 1987. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr. Bull. 13 (2), 261–276. https://doi.org/10.1093/schbul/13.2.261.

Kongsakon, R., Trinidad-Oñate, P., Chaudhry, H.R., Raza, S.B., Leynes, C.R., Khan, I.U., Ismail, H.C., Chan, B., Ignacio, J.C., Rodriguez, S.C., Lowry, A.J., Brnabic, A.J., Buenaventura, R., 2006. Asian outpatients with schizophrenia: a double-blind randomized comparison of quality of life and clinical outcomes for patients treated with olanzapine or haloperidol. J. Med. Assoc. Thai. 89 (8), 1157–1170.

Landbloom, R., Mackle, M., Wu, X., Kelly, L., Snow-Adami, L., McIntyre, R.S., Mathews, M., Hundt, C., 2017. Asenapine for the treatment of adults with an acute exacerbation of schizophrenia: results from a randomized, double-blind, fixed-dose, placebo-controlled trial with olanzapine as an active control. CNS Spectr. 22 (4), 333–341. https://doi.org/10.1017/S1092852916000377.

Leucht, S., Cipriani, A., Spineli, L., Mavridis, D., Orey, D., Richter, F., Samara, M., Barbui, C., Engel, R.R., Geddes, J.R., Kissling, W., Stapf, M.P., Lässig, B., Salanti, G., Davis, J.M., 2013. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. Lancet 382 (9896), 951–962. https://doi.org/10.1016/S0140-6736(13)60733-3.

Lublin, H., Haug, H.J., Koponen, H., Sigmundsson, T., Kolb, S.A., 2009. Ziprasidone versus olanzapine, risperidone or quetiapine in patients with chronic schizophrenia: a 12-week open-label, multicentre clinical trial. World J. Biol. Psychiatry 10 (4 Pt 3), 710–718. https://doi.org/10.1080/15622970802269589.

Marder, S.R., Kramer, M., Ford, L., Eerdekens, E., Lim, P., Eerdekens, M., Lowy, A., 2007. Efficacy and safety of paliperidone extended-release tablets: results of a 6-week, randomized, placebo-controlled study. Biol. Psychiatry 62 (12), 1363–1370. https://doi.org/10.1016/j.biopsych.2007.01.017.

McDonagh, M.S., Dana, T., Selph, S., Devine, E.B., Cantor, A., Bougatsos, C., Blazina, I., Grusing, S., Fu, R., Haupt, D.W., 2020. Updating the comparative evidence on second-generation antipsychotic use with schizophrenia. Psychiatric. Res. Clin. Pract. 2 (2), 76–87. https://doi.org/10.1176/appi.prcp.20200004.

Meltzer, H.Y., Cucchiaro, J., Silva, R., Ogasa, M., Phillips, D., Xu, J., Kalali, A.H., Schweizer, E., Pikalov, A., Loebel, A., 2011. Lurasidone in the treatment of schizophrenia: a randomized, double-blind, placebo- and olanzapine-controlled

study. Am J Psychiatry 168 (9), 957–967. https://doi.org/10.1176/appi.ajp.2011.10060907.

Mortimer, A., Martin, S., Lôo, H., Peuskens, J., SOLIANOL Sudy Group, 2004. A double-blind, randomized comparative trial of amisulpride versus olanzapine for 6 months in the treatment of schizophrenia. Int. Clin. Psychopharmacol. 19 (2), 63–69. https://doi.org/10.1097/00004850-200403000-00002.

Naber, D., Riedel, M., Klimke, A., Vorbach, E.U., Lambert, M., Kühn, K.U., Bender, S., Bandelow, B., Lemmer, W., Moritz, S., Dittmann, R.W., 2005. Randomized double blind comparison of olanzapine vs. clozapine on subjective well-being and clinical outcome in patients with schizophrenia. Acta Psychiatr. Scand. 111 (2), 106–115. https://doi.org/10.1111/j.1600-0447.2004.00486.x.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., Moher, D., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372, n71. https://doi.org/10.1136/bmj.n71.

R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

Revicki, D.A., Genduso, L.A., Hamilton, S.H., Ganoczy, D., Beasley Jr., C.M., 1999. Olanzapine versus haloperidol in the treatment of schizophrenia and other psychotic disorders: quality of life and clinical outcomes of a randomized clinical trial. Qual. Life Res. 8 (5), 417–426. https://doi.org/10.1023/a:1008958925848.

Rotella, F., Cassioli, E., Calderani, E., Lazzeretti, L., Ragghianti, B., Ricca, V., Mannucci, E., 2020. Long-term metabolic and cardiovascular effects of antipsychotic drugs. A meta-analysis of randomized controlled trials. Eur. Neuropsychopharmacol. 32, 56–65. https://doi.org/10.1016/j.euroneuro.2019.12.118.

Schmidt, M.E., Kent, J.M., Daly, E., Janssens, L., Van Osselaer, N., Hüsken, G., Anghelescu, I.G., Van Nueten, L., 2012. A double-blind, randomized, placebo-controlled study with JNJ-37822681, a novel, highly selective, fast dissociating D_2 receptor antagonist in the treatment of acute exacerbation of schizophrenia. Eur. Neuropsychopharmacol. 22 (10), 721–733. https://doi.org/10.1016/j.euroneuro.2012.02.007.

Schoemaker, J., Naber, D., Vrijland, P., Panagides, J., Emsley, R., 2010. Long-term assessment of Asenapine vs. Olanzapine in patients with schizophrenia or schizoaffective disorder. Pharmacopsychiatry 43 (4), 138–146. https://doi.org/10.1055/s-0030-1248313.

Shen, J.H., Zhao, Y., Rosenzweig-Lipson, S., Popp, D., Williams, J.B., Giller, E., Detke, M. J., Kane, J.M., 2014. A 6-week randomized, double-blind, placebo-controlled, comparator referenced trial of vabicaserin in acute schizophrenia. J. Psychiatr. Res. 53, 14–22. https://doi.org/10.1016/j.jpsychires.2014.02.012.

Simpson, G.M., Weiden, P., Pigott, T., Murray, S., Siu, C.O., Romano, S.J., 2005. Six-month, blinded, multicenter continuation study of ziprasidone versus olanzapine in schizophrenia. Am. J. Psychiatry 162 (8), 1535–1538. https://doi.org/10.1176/appi.ajp.162.8.1535.

Solmi, M., Murru, A., Pacchiarotti, I., Undurraga, J., Veronese, N., Fornaro, M., Stubbs, B., Monaco, F., Vieta, E., Seeman, M.V., Correll, C.U., Carvalho, A.F., 2017. Safety, tolerability, and risks associated with first- and second-generation antipsychotics: a state-of-the-art clinical review. Ther. Clin. Risk Manag. 13, 757–777. https://doi.org/10.2147/TCRM.S117321.

Tollefson, G.D., Beasley Jr., C.M., Tran, P.V., Street, J.S., Krueger, J.A., Tamura, R.N., Graffeo, K.A., Thieme, M.E., 1997. Olanzapine versus haloperidol in the treatment of schizophrenia and schizoaffective and schizophreniform disorders: results of an international collaborative trial. Am. J. Psychiatry 154 (4), 457–465. https://doi.org/10.1176/ajp.154.4.457.

Tollefson, G.D., Birkett, M.A., Kiesler, G.M., Wood, A.J., Lilly Resistant Schizophrenia Study Group, 2001. Double-blind comparison of olanzapine versus clozapine in schizophrenic patients clinically eligible for treatment with clozapine. Biol. Psychiatry 49 (1), 52–63. https://doi.org/10.1016/s0006-3223(00)01026-x.

Tran, P.V., Hamilton, S.H., Kuntz, A.J., Potvin, J.H., Andersen, S.W., Beasley Jr., C., Tollefson, G.D., 1997. Double-blind comparison of olanzapine versus risperidone in the treatment of schizophrenia and other psychotic disorders. J. Clin. Psychopharmacol. 17 (5), 407–418. https://doi.org/10.1097/00004714-199710000-00010.

Viechtbauer, W., 2010. Conducting meta-analyses in R with the metafor package. J. Stat. Softw. 36 (3), 1–48.

Younis, I.R., Gopalakrishnan, M., Mathis, M., Mehta, M., Uppoor, R., Zhu, H., Farchione, T., 2020. Association of end point definition and randomized clinical trial duration in clinical trials of schizophrenia medications. JAMA Psychiatry 77 (10), 1064–1071. https://doi.org/10.1001/jamapsychiatry.2020.1596.