Blinding successfulness of antipsychotic trials
: A systematic review and meta-analysis

1. Introduction

Schizophrenia is a serious disease that causes psychosis and cognitive impairment, and the lifetime prevalence is about 1% of the population across different cultures and regions [1]. Antipsychotic therapy is essential for its treatment. There are currently more than 30 antipsychotic drugs on the market and more than 400 randomized controlled trials (RCTs) have been conducted on these drugs. Historically, the first-generation antipsychotics (FGAs) were initially used as the mainstay. However, they had high risks of side effects such as extrapyramidal symptoms. Since the 1990s, various second-generation antipsychotics (SGAs) have been developed and become the mainstream because they cause fewer side effects than FGAs[1, 2]. For example, the risk ratios (RRs) for needing antiparkinsonian drugs compared to placebo were 3.12 (95%CI: 2.74 to 3.50) for haloperidol and 1.02 (95% CI: 0.79 to 1.30) for olanzapine [3]. On the other hand, SGAs tend carry higher risks for weight gain than FGAs: the mean differences (MD) of haloperidol and olanzapine compared to placebo were 0.54 kg (95% CI: 0.15 to 0.95) and 2.78 kg (95% CI: 2.44 to 3.13), respectively. Thus, the profiles of side effects of these drugs are different.

Blinding of RCTs is very important for accurate assessment of drug efficacy. Open or single-blind RCTs of antipsychotics for schizophrenia clearly overestimate the drug efficacy in comparison with double-blind trials. Even among so-called double blind studies, if the blinding is not adhered to, there is a risk of overestimation [4, 5]. For drugs with relatively strong and/or idiosyncratic side effects such as antipsychotics, the blinding may be easily broken.

Fergusson et al. investigated the top journals in psychiatry from 1998 to 2001 and reported that blinding assessments were conducted in 7 of 97 studies. However, no studies on antipsychotic drugs were included among these seven [6]. Hróbjartsson et al. also investigated the RCTs published in the year 2001, and blinding assessment was conducted in 12 psychiatric studies. However, there were only two studies of antipsychotics, one for haloperidol for smoking [7] and the other for olanzapine for social

drinkers [8]. Baethge et al. [9] investigated whether blinding of RCTs of psychiatric drugs was maintained. In this study, they searched for studies from 2000 to 2010 that assessed whether blinding was done properly. However, only three of the 569 antipsychotic studies reported blinding assessment, and studies on schizophrenia tended to assess their blindness less frequently than non-schizophrenia studies [9].

Thus, of numerous double-blind RCTs of antipsychotics conducted to date in the field of schizophrenia, it is unclear how many studies have assessed their blindness. The objective of this study is therefore to clarify the following points: (1) the proportion of RCTs in which blinding was assessed, (2) the degree of their blinding successfulness, and (3) whether the adequacy of blinding affects the effect size of antipsychotics for schizophrenia.

2. Materials and Methods

2.1 Selection of the studies

2.1.1 Study design

We will include double or more-blinded RCTs where participants, study personnel and/or outcome assessors were blinded. We will exclude the study that did not aim to examine the effects of antipsychotics for schizophrenia.

2.1.2 Diagnosis and participants

Participants will be diagnosed as acute phase schizophrenia or related disorders (schizophreniform or schizoaffective disorders) according to the following diagnositic criteria: Feighner criteria[10], Research Diagnostic Criteria[11], DSM-III[12], DSM-III-R, DSM-IV[13], DSM-5[14], and ICD-10[15]. We will exclude the studies focused on the patients with treatment resistance. We will also exclude relapse prevention studies.

2.1.3 Antipsychotic drugs

We will include the following SGAs and FGAs [3].

- FGAs: benperidol, chlorpromazine, clopenthixol, flupentixol, fluphenazine, haloperidol, levomepromazine, loxapine, molindone, penfluridol, perazine, perphenazine, pimozide, sulpiride, thioridazine, thiotixene, trifluoperazine, and zuclopenthixol
- SGAs: all SGAs available in Europa or the United States

Trials of combination therapy with antipsychotics and other psychotropic drugs will be included if other psychotropic drugs are not the interventions of interest of the trial and are prescribed equally between the arms.

2.1.4 Comparison

We will include only placebo-controlled studies because comparisons between “active drug vs. placebo” and “active drug vs. active drug” are different in terms of blinding failures. It is clearer in placebo-controlled trials than in active drug-controlled trials how blind failures affect the outcomes.

Since FGA and SGA have different side effect profiles, we will further analyze these two subcategories.

2.1.4.1 FGA vs. placebo

2.1.4.2 SGA vs. placebo

2.1.5 Definition of assessment of blinding

Full texts of publications meeting the eligibility criteria including supplementary materials or secondary papers will be screened for further information of blinding assessment. It includes any data on guess of group allocation by participants and/or evaluators/physicians.

2.1.6 Literature search

We have already built a comprehensive database of antipsychotic drug trials for schizophrenia to conduct their network meta-analyses [3]. We had searched the Cochrane Schizophrenia Group Controlled Trials Register, MEDLINE, EMBASE, PsycINFO, Cochrane Library, PubMed, Biosis, ClinicalTrials.gov and WHO ICTRP from 1946 to January 8th, 2019 (see the supplementary appendix of Huhn 2019 [3]). We will use this database for our literature search.

2.2 Data extraction

Study characteristics, information regarding blinding assessments, and blinding details will be extracted for studies with blindness assessment in the following pre-specified format:

(1) Study characteristics: journal reference, author names and affiliations, source of

funding, publication year, impact factor of journal (in publication year)

(2) Study design and result: method of diagnosis, sample size, number of arms, parallel or cross-over design, concealment of allocation or not, use of intention-to-treat analysis or not, main results (favor active treatment or not, effect measures and the corresponding 95% confidence interval (CI) or p-value)

(3) Exposure: specific antipsychotics and the generation (FGA or SGA)

(4) Outcome: participant-reported, observer-based, or mixed; outcome measurement tools/scales

(5) Blinding details: the key trial persons blinded; how was blinding achieved

(6) Blinding assessment: when and how was blinding was assessed (during trial or end of trial), the response categories, use of a forced-choice or allowing a 'don't know' option; statistical method to test the effectiveness of blinding; the results of blindness assessment, investigators' conclusions about the blinding success

Two review authors will independently review the full text and supplementary materials and extract the information about blinding assessment. If there is any discrepancy in the interpretation of the information, a third review author will be consulted. If there is insufficient information, we will contact the author of the paper.

2.2.1 Primary outcomes

2.2.1.1 The proportion of RCTs in which blinding was assessed

We will show the proportion of studies evaluated blinding assessment among the included studies.

2.2.1.2 Blinding successfulness

The degree of blinding successfulness will be presented with Cohen's Kappa[16]. We will calculate Kappa statistics between guesses and true allocations from each study and synthesize them. We will interpret Kappa values according to the following Lin's definition (paper under review): a Kappa value of -0.20 to 0.20 as successful blinding, 0.21 to 0.40 as slightly broken, 0.41 to 0.60 as moderately broken, 0.61 to 1 as severely broken. We will show the results of "blinding successfullness among patients" and "blinding successfullness among assesssors" separately. If a study reported the blinding

assessment results at multiple timepoints, we will choose the nearest point to the end of treatment.

2.2.2 Secondary outcomes

2.2.2.1 The relationship between the adequacy of blinding and the effect size

We will use meta-regression and Pearson's r to examine the relationship between blinding success and effect sizes.

2.3 Subgroup analysis

We will show the results separately by antipsychotic types (FGAs vs SGAs).

2.4 Sensitivity analysis

We will exclude the studies which did not report the exact number of patients or assessors tested with blinding assessment.

2.5 Statistical Analysis

Descriptive data of all collected studies will be presented as percentages or means with standard deviations (SDs).

(i) The proportion of RCTs in which blinding was assessed (see 2.2.1.1)

The proportion of blinding examination is derived by dividing the number of studies with blindness assessment over the number of all studies collected. The changes over time will be described (tabulated or graphically displayed) and analyzed with trend analysis.

(ii) Blinding successfulness (see 2.2.1.2)

To quantify the effectiveness of blinding, data regarding the accuracy of guesses about treatment assignment will be extracted among studies reporting blindness assessment, including the number of correct guesses, the number of incorrect guesses, and the number of "don't know" answer in active and control groups. Although past research recommended an open-choice option design for guess tests, only a few studies allow a "don't know" option[17, 18]. A "don't know" response will be assigned proportionally by the number of active-treatment and control responses in each report.

If the comedication in studies with intervention of combination or adjunctive therapy would strongly affect the blinding success, such as electro-convulsive therapy, we would

exclude them from the analysis of blinding success rate.

The meta-analysis of standard Kappa statistics will be our first choice to measure blindness success rate across research on antipsychotics [19]. The formula for computing Cohen's Kappa and the corresponding standard error is expressed in Eq. 1.

$$\kappa = \frac{p_0 - p_c}{1 - p_c}$$

$$\mathrm{Var}(\kappa) = \frac{p_0(1 - p_0)}{n(1 - p_c)^2} \quad \cdots Eq.1$$

where p_0 is unblinding percent, defined as the proportion of trial persons who correctly guessed the participants' treatment allocation; p_c is chance agreement, defined as the proportion of correctly guessing the treatment assignment that would be expected by chance; n is the number of personnel included in the blindness assessment.

Since Kappa measures agreement rather than disagreement when estimating the inter-rater reliability, the explanation would be opposite while it is used to understand the blindness success. A κ value of 0 indicates the observed successful guesses rates are the same as guesses by chance, which can be meaningfully interpreted as the blinding is more successful, and a κ value of 1 reveals that the blinding is totally broken.

Since the study characteristics and participant inclusion criteria may vary in antipsychotics research, a random-effects model (Eq. 2) will be fitted to pool the κ estimates.

$$\kappa_i = \theta + \mu_i + e_i \quad \cdots Eq.2$$

where θ is the overall estimate of blinding success rate, μ_i is the between-study variation that is normally distributed with mean 0 and variance s^2, and e_i represents the random sampling error.

Heterogeneity between study-specific estimates will be tested using chi-square tests and measured with the I^2 index (a measure of the percentage of variation across the studies caused by the heterogeneity) and the common heterogeneity parameter tau. Publication bias will be evaluated through funnel plot visual analysis and the Egger's test [20].

All articles with blindness assessment will then be further separated into two groups: trials of FGA and trials of SGA. A mixed-effects model will be fitted to explore the effect of antipsychotic generation on the Cohen's κ (Eq. 3). We will assess the effects of antipsychotics categories (β_1) on κ statistics via meta-regression and derive the Cohen's κ of each category. The significance of differences in κ statistics between FGAs and

SGAs would be tested with chi-square goodness of fit test of *Q* statistics, which is used to test the homogeneity between groups.

$$\kappa_i = \beta_0 + \beta_1 Generation + \mu_i + e_i \ \cdots \ Eq.3$$

where the variance of μ_i represents the amount of residual heterogeneity, i.e., the variability of Cohen's *κ* estimates across studies that cannot be accounted for by the antidepressant generation in the mixed-effects model; β_1 stands for the effects of antidepressant generation on Cohen's *κ*. Heterogeneity was measured with I^2 and tau^2.

(iii) The relationship between the adequacy of blinding affects the effect size (see 2.2.2.1) Treatment effect sizes for each study will be calculated as Cohen's *d* statistics. For studies where SDs are not reported, we will impute SDs from other studies using Furukawa's methods[21]. To examine the relationship between effect size (Cohen's *d* statistic) and the degree of blinding success (*κ* value, the blinding success indicator), we will use the Pearson's *r* to express the correlation and t test to examine the significance of correlations. Furthermore, we will apply a meta-regression to investigate the relationship of effect sizes (Cohen's *d*) and blinding effects (Cohen's *κ*) (Eq.4).

$$D_i = \beta_0 + \beta_1 \kappa_i + \mu_i + e_i \ \cdots \ Eq.4$$

where D_i represents the effect sizes; the variance of μ_i represents the amount of residual heterogeneity, i.e., the variability of Cohen's *d* estimates across studies that cannot be accounted for by the *κ* values and antidepressant generation in the mixed-effects model; β_1 stands for the influence of blinding success degree (Cohen's *κ*) on effect sizes (Cohen's *d*).

If we cannot extract enough information to calculate the Cohen's *κ* and the corresponding standard error in most of the studies with blindness assessment, we will use the summary proportion of incorrect guesses, which indicates successful blinding, as a substitute indicator.

All tests are considered significant statistically, for p-values less than 0.05. The analyses and the correspondent graphical visualization of forest and funnel plots will be performed with R package *metaphor* [22] on the most recent version of R [23].

3. Results (image)

Figure 1. Flow chart of the study selection

Figure 2. Forest plot of blinding successfulness among (A) patients and (B) assessors.

Figure 3. Relationship between effect sizes (SMD) and the degree of blinding successfulness (kappa) of (A) patients and (B) assessors

Figure 4. Contour-enhanced funnel plot of kappa (blinding successfulness) among patients

Table 1. Characteristics of antipsychotics RCTs with blinding assessment and studies included in the meta-analysis.

Characteristics	**Trials with BA (N = ??), n (%)**	**Trials included in MA (N=?)*, n(%)**
Published year		
≤ 1999		
≥ 2000		
Sponsor		
Industry		
Non-industry		
Trial arms included		
2 arms		
3 or more arms		
Type of antipsychotics		
FGAs		
SGAs		
Blinding method		
Single		
Double		
More		
Persons blinded		
Patients		
Assessors		
Caregivers		

Investigators	
Analytical method	
Intention to treat	
Schizophrenia symptoms measure	
Observer-based	
Patient-reported	
Blinding assessed in	
Patients	
Assessors	
Timing of BA	
During trial	
End of trial	
Unclear	
Blinding ratings	
Forced choice	
(active vs. control)	
Allow ‘don’t know’ option	
Unclear	
Qualitative conclusions of BA for patients	
Reported as successful	
Reported as unsuccessful	
No conclusion reported	
Qualitative conclusions of BA for assessors	
Reported as successful	
Reported as unsuccessful	
No conclusion reported	

BA: blinding assessment; MA: meta-analysis; FGAs: first-generation antipsychotics; SGAs: second-generation antipsychotics.

Table 2. Descriptive characteristics of included studies

Details of included studies will be described here.

Table 3. Kappas, proportions of correct guesses, and treatment effect sizes of each individual study

Author (year)	No. of guesses	Kappa (95%CI)	Correct guesses (%)	Treatment effect sizes (95%CI)
Patient blinding				
Assessor blinding				

Table 4. Sensitivity analysis of the summary kappa by excluding 1) studies without clear information that assessors were blinded and 2) studies without the exact number of personnel tested with blinding assessment

Sensitivity analysis	n	Kappa (95% CI)	Heterogeneity
Patients			
Primary analysis			I^2=% τ^2= χ^2 p=
With clear information that assessors are blinded			I^2=% τ^2= χ^2 p=
With the exact number of patients for whom treatments were guessed			I^2=% τ^2= χ^2 p=

Assessors	
Primary analysis	I^2=% τ^2= χ^2 p=
With clear information that assessors are blinded	I^2=% τ^2= χ^2 p=
With the exact number of patients for whom treatments were guessed	I^2=% τ^2= χ^2 p=

References

1. Mueser, K.T. and S.R. McGurk, *Schizophrenia.* Lancet, 2004. **363**(9426): p. 2063-72.
2. Hasan, A., et al., *World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for biological treatment of schizophrenia, part 2: update 2012 on the long-term treatment of schizophrenia and management of antipsychotic-induced side effects.* World J Biol Psychiatry, 2013. **14**(1): p. 2-44.
3. Huhn, M., et al., *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, 2019. **394**(10202): p. 939-951.
4. Savovic, J., et al., *Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies.* Health Technol Assess, 2012. **16**(35): p. 1-82.
5. Savovic, J., et al., *Association Between Risk-of-Bias Assessments and Results of Randomized Trials in Cochrane Reviews: The ROBES Meta-Epidemiologic Study.* Am J Epidemiol, 2018. **187**(5): p. 1113-1122.
6. Fergusson, D., et al., *Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials.* Bmj, 2004. **328**(7437): p. 432.
7. Brauer, L.H., et al., *Haloperidol reduces smoking of both nicotine-containing and denicotinized cigarettes.* Psychopharmacology (Berl), 2001. **159**(1): p. 31-7.
8. Hutchison, K.E., et al., *Olanzapine reduces urge to drink after drinking cues and a priming dose of alcohol.* Psychopharmacology (Berl), 2001. **155**(1): p. 27-34.
9. Baethge, C., O.P. Assall, and R.J. Baldessarini, *Systematic review of blinding assessment in randomized controlled trials in schizophrenia and affective disorders 2000-2010.* Psychother Psychosom, 2013. **82**(3): p.

152-60.
10. Feighner, J.P., et al., *Diagnostic criteria for use in psychiatric research.* Arch Gen Psychiatry, 1972. **26**(1): p. 57-63.
11. Spitzer, R.L., J. Endicott, and E. Robins, *Research diagnostic criteria: rationale and reliability.* Arch Gen Psychiatry, 1978. **35**(6): p. 773-82.
12. *Diagnostic and Statistical Manual, Third Edition Revised (DSM-III-R)*. 1987, Washington, DC: American Psychiatric Association (APA).
13. *Diagnostic and Statistical Manual, Fourth Edition (DSM-IV)*. 1994, Washington, DC: American Psychiatric Association (APA).
14. *Diagnostic and Statistical Manual, Fifth Edition (DSM-5)*. 2013, Washington, DC: American Psychiatric Association (APA).
15. *International Classification of Diseases, 10th Revision (ICD-10)*. 1992, Geneva: World Health Organization (WHO).
16. Landis, J.R. and G.G. Koch, *The measurement of observer agreement for categorical data.* Biometrics, 1977. **33**(1): p. 159-74.
17. Bang, H., L. Ni, and C.E. Davis, *Assessment of blinding in clinical trials.* Control Clin Trials, 2004. **25**(2): p. 143-56.
18. Bang, H., et al., *Blinding assessment in clinical trials: A review of statistical methods and a proposal of blinding assessment protocol.* Clinical Research and Regulatory Affairs, 2010. **27**(2): p. 42-51.
19. Sun, S., *Meta-analysis of Cohen's kappa.* Health Services and Outcomes Research Methodology, 2011. **11**(3): p. 145-163.
20. Egger, M., et al., *Bias in meta-analysis detected by a simple, graphical test.* Bmj, 1997. **315**(7109): p. 629-34.
21. Furukawa, T.A., et al., *Imputing missing standard deviations in meta-analyses can provide accurate results.* J Clin Epidemiol, 2006. **59**(1): p. 7-10.
22. Viechtbauer, W., *Conducting meta-analyses in R with the metafor package.* Journal of Statistical Software, 2010. **36**(3): p. 1-48.
23. RStudio Team, *RStudio: Integrated Development for R*. 2019, RStudio, Inc.: Boston, MA.