

# Challenges in meta-analyses with observational studies

Silvia Metelli,<sup>1,2</sup> Anna Chaimani <sup>1,3</sup>

<sup>1</sup>Université de Paris, Research Center of Epidemiology and Statistics (CRESS-UMR1153), INSERM, INRA, F-75004, Paris, France

<sup>2</sup>Assistance Publique - Hôpitaux de Paris (APHP), Paris, France

<sup>3</sup>Cochrane France, Paris, France

## Correspondence to

Dr Anna Chaimani, Université de Paris, Research Center of Epidemiology and Statistics Sorbonne Paris Cité (CRESS-UMR1153), INSERM, INRA, Paris 75004, France; [anna.chaimani@parisdescartes.fr](mailto:anna.chaimani@parisdescartes.fr)

Received 27 November 2019

Revised 4 February 2020

Accepted 5 February 2020

Published Online First

5 March 2020

## ABSTRACT

**Objective** Meta-analyses of observational studies are frequently published in the literature, but they are generally considered suboptimal to those involving randomised controlled trials (RCTs) only. This is due to the increased risk of biases that observational studies may entail as well as because of the high heterogeneity that might be present. In this article, we highlight aspects of meta-analyses with observational studies that need more careful consideration in comparison to meta-analyses of RCTs.

**Methods** We present an overview of recommendations from the literature with respect to how the different steps of a meta-analysis involving observational studies should be comprehensively conducted. We focus more on issues arising at the step of the quantitative synthesis, in terms of handling heterogeneity and biases. We briefly describe some sophisticated synthesis methods, which may allow for more flexible modelling approaches than common meta-analysis models. We illustrate the issues encountered in the presence of observational studies using an example from mental health, which assesses the risk of myocardial infarction in antipsychotic drug users.

**Results** The increased heterogeneity observed among studies challenges the interpretation of the diamond, while the inclusion of short exposure studies may lead to an exaggerated risk for myocardial infarction in this population.

**Conclusions** In the presence of observational study designs, prior to synthesis, investigators should carefully consider whether all studies at hand are able to answer the same clinical question. The potential for a quantitative synthesis should be guided through examination of the amount of clinical and methodological heterogeneity and assessment of possible biases.

## INTRODUCTION

Systematic reviews and meta-analyses help to establish evidence-based clinical practice and resolve contradictory research outcomes while supporting research planning and prioritisation.<sup>1</sup> Therefore, meta-analysis is being increasingly used in most medical fields with the aim to reach an overall understanding of clinical outcomes and their sources of variation. Although synthesis of randomised controlled trials (RCTs) is generally considered as the highest level of clinical evidence, there are several concerns regarding the potential of meta-analysis of observational studies to provide reliable results.<sup>2</sup> The reasons making researchers being sceptical with synthesising observational studies are

mainly related to the high risk for within-study and across-study biases, as well as to the presence of increased heterogeneity.

A major issue of observational evidence is that it is known to have limited internal validity as it is subject to both bias and confounding. Overall, observational study designs are not the most appropriate to assess the causal relationship between an intervention and an outcome as several characteristics might differ or might change over time between the different intervention groups. So, the inclusion of observational studies in a meta-analysis might introduce bias in the summary effect. To mitigate the risk of confounding and to make more comparable the different study groups, investigators usually adjust the relative effects for several characteristics that may be related to the outcome and/or to the intervention. Propensity scores (ie, the probability of treatment assignment conditional on observed baseline characteristics) are now also being used frequently in the analysis of observational studies as they likely allow reduction of confounding and selection bias. Despite the fact that these methods have the potential to produce less biased results, at the meta-analysis level they increase the methodological heterogeneity as often different studies use different analysis methods or different adjustment factors and the comparability of their results is questionable. Apart from methodological heterogeneity, clinical heterogeneity is also expected to be much higher than in meta-analyses of RCTs since observational studies are based on less stringent inclusion criteria.

An additional problem in the synthesis of observational studies is that it is always challenging or even impossible to assess the risk of bias both within and across studies. The latter is because preregistration and protocol preparation are not mandatory for observational studies, and as a result, unpublished studies or partly unpublished results cannot be identified. This leads to an increased risk of publication bias and other reporting biases such as selective outcome reporting. With respect to within-study bias, in contrast to RCTs, the lack of widely agreed quality criteria and the absence of sufficient empirical evidence to support the focus on particular study features render the assessment of the risk of bias for observational studies and their meta-analysis rather challenging. To date, more than 80 tools have been proposed to assess the credibility of observational studies, but most of them have not been used in practice. Recently, a draft Cochrane risk of bias tool for non-randomised studies was also developed that considers each observational



© Author(s) (or their employer(s)) 2020. No commercial re-use. See rights and permissions. Published by BMJ.

**To cite:** Metelli S, Chaimani A. *Evid Based Ment Health* 2020;**23**:83–87.

study as an attempt to mimic a hypothetical pragmatic randomised trial.<sup>3</sup> Nevertheless, this tool has not been finalised yet.

Despite all the above issues, observational data not only offer a valuable source of supplementary information to RCTs but also, for some clinical questions, provide the most reliable data (eg, safety of interventions and long-term outcomes) or even the only source of available evidence (eg, effectiveness of transplantation). In addition, results from observational studies might be more directly applicable to the general populations as they are designed under a more real-life setting than RCTs, which usually involve very restricted populations treated under highly standardised care.<sup>4</sup> Therefore, meta-analysis of observational data alone or in combination with RCTs (when possible) is often desirable.<sup>5</sup> However, a methodological systematic review of meta-analyses of observational studies in the field of psychiatry found several deficiencies in terms of assessing study quality, publication bias and risk for confounding, while the majority of the meta-analyses found significant heterogeneity.<sup>6</sup>

In this article, we review methods and recommendations in the literature for the different steps of a meta-analysis involving observational studies. Using an example from mental health, we present and discuss the issues that frequently arise due to the nature of these data.

## METHODS

### Searching for relevant studies

Investigators intending to perform meta-analyses with observational studies should be aware that identifying all relevant studies requires more extensive literature searches than those that usually take place in systematic reviews involving only RCTs. Lemeshow *et al*<sup>7</sup> showed that, restricting the search in common large databases such as MEDLINE may achieve a sensitivity on average between 65% and 80%, depending on the medical field. Achieving a sensitivity of about 90% required searching for observational studies in at least four databases.

### Extracting data

Poor reporting is a common issue in observational studies, and very often the required data for the meta-analysis cannot be easily obtained. Although guidance on the reporting of observational studies does exist (eg, the Strengthening the Reporting of Observational Studies in Epidemiology checklist),<sup>8</sup> following this guidance is not mandatory yet for most medical journals. Also, several studies within a systematic review might have been published before the development of such checklists. Apart from poor reporting, observational studies usually report results from several analyses, and it is not always straightforward to determine which results are more appropriate to include in a meta-analysis.<sup>9</sup> For instance, they might provide both unadjusted and adjusted results or they might have used a number of analyses adjusting for different sets of potential confounders. Since unadjusted results are most likely biased, they should not be preferred, even though they might seem less heterogeneous across studies. Instead, researchers should consider the most important potential confounders in advance (when preparing the protocol) and opt for extracting results adjusted at least for these or most of these characteristics. On top of outcome data and data on other important study characteristics, extracting information for the assessment of the risk of bias of all studies is crucial, even though existing risk of bias tools for observational studies might not be optimal.

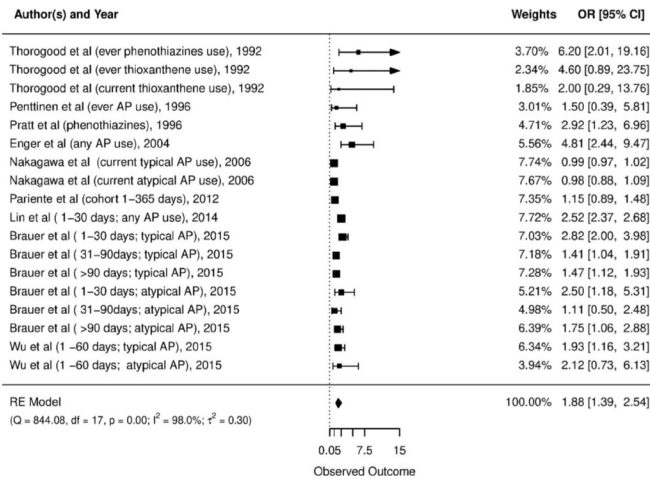
### Synthesising data and controlling for bias

Observational studies usually have larger sample sizes than RCTs and might yield highly precise results.<sup>2</sup> Given all the aforementioned issues of observational evidence (ie, bias and confounding), this phenomenon might lead to spurious inferences because usually the more precise the summary effects, the stronger the conclusions of the investigators. Further, when observational and randomised studies are synthesised using typical methods (eg, classical fixed or random effects meta-analysis), the weight of observational studies would be larger than that of the RCTs, although the latter usually give more reliable results. Thus, it is of great importance to consider very carefully the setting of each study before proceeding with the synthesis of the results, and whether it is appropriate to answer the research question of the meta-analysis.

At the stage of data synthesis, the main issues in the presence of observational studies are (1) how to accommodate the possibly large heterogeneity that may be present especially when different types of observational studies, or also RCTs, are combined in the same analysis and (2) how to account for different biases. Overall, using a fixed effect meta-analysis does not seem a reasonable approach considering that observational studies generally have very variable populations, which are followed under different conditions. In case of multiple observational designs or combination with RCTs, these discrepancies would likely be magnified. The random effects model accounts better for this apparent heterogeneity. Subgroup analysis or meta-regression by study design and by type of analysis (eg, different adjustment factors) should always take place even when there is no strong evidence for statistical heterogeneity. The risk of bias of the studies should also be considered as a potential source of variability; performing a sensitivity analysis excluding studies of lower credibility can reveal whether such studies have an impact on the summary effect.

A common misconception in meta-analyses with observational studies is that the assessment of across-study bias can take place across all studies no matter the design. Investigators should be aware that reporting bias (publication bias and selective reporting) or small-study effects would probably work differently for different study designs, and consequently, investigation and assessment of such types of biases should be made separately for different designs. Otherwise, such biases might be masked due to the variability of their magnitude and possibly their mechanism among different types of studies. For example, cohort studies are on average much larger than case-control studies or RCTs, and as a result, they often lead to more precise treatment effect estimates. Putting together these three designs in the same funnel plot would place all cohort studies at the top of the graph and all other studies at the bottom. Hence, any asymmetry would be due to heterogeneity across study designs and not to small-study effects.

Several more sophisticated methods than classical meta-analysis have also been suggested in the literature for handling heterogeneity and bias potentially more properly in the presence of multiple study designs. For example, an alternative approach considers the classification of the possible biases for each study into 'internal' and 'external' based on specific items related to the nature of the data.<sup>10 11</sup> Based on this classification, all identified sources of bias may then be used within a Bayesian modelling framework<sup>12</sup> to obtain bias-adjusted summary estimates. The combination of different meta-analytic submodels specifically designed to capture different features of different types of studies is another option,<sup>13</sup> while using hierarchical models



**Figure 1** Forest plot of the main meta-analysis for estimating the risk of myocardial infarction with antipsychotics (AP=antipsychotics, RE=random effects)

offers a convenient way to allow for variation across the results of the different designs.<sup>14</sup> Finally, the incorporation of observational evidence as prior information in a Bayesian meta-analysis of RCTs<sup>15</sup> or employing a ‘design-adjusted’ analysis<sup>16</sup> that allows studies of lower credibility to get less weight in the synthesis was originally suggested in the context of network meta-analysis but can be also applied for the case of pairwise meta-analysis. An extensive review of methods allowing combination of different study designs can be found elsewhere.<sup>17</sup>

### Reporting the findings

Reporting of meta-analyses of observational studies should follow the general principles for reporting any systematic review and meta-analysis. More emphasis should be placed, though, on the rationale for study inclusion criteria and on the way confounding bias, study quality, heterogeneity and publication

**Table 1** Results from the meta-regression model assessing the impact of exposure time

Exposure time	Log OR	SE	95% CI	P value
Intercept (1–30 days)	0.96	0.09	(0.78 to 1.14)	<0.001
Coefficient (>30 days)	-0.75	0.12	(-0.99 to to 0.52)	<0.001

The heterogeneity variance is  $\tau^2=0.02$ .

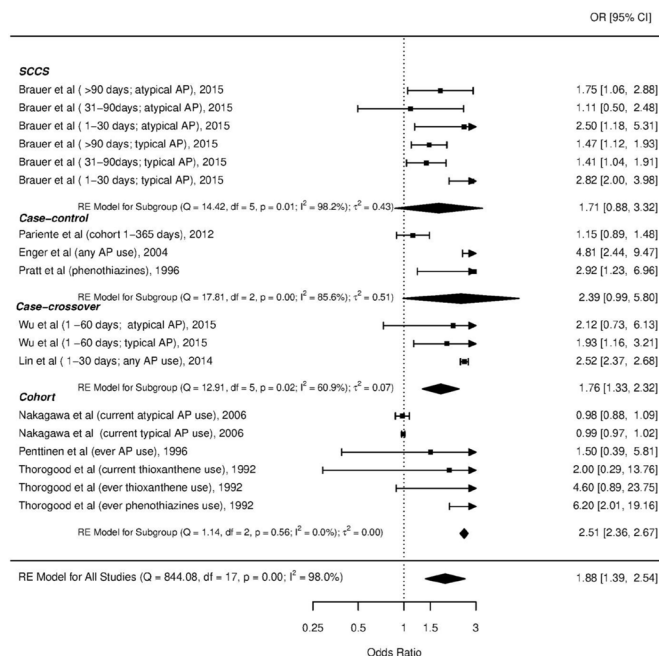
bias were assessed. Given that accounting adequately for these aspects of the data would be rather challenging, meta-analysts should consider providing alternative explanations for the observed results.<sup>18</sup>

### RESULTS

To illustrate the issues described previously, we use as example a meta-analysis of observational studies assessing the risk of myocardial infarction (MI) in antipsychotic drug users.<sup>19</sup> The dataset consists of nine studies published between 1992 and 2015: three case-control studies, two cohorts, two case-cross-over studies and one self-controlled case series. Each study has contributed more than one estimate in the original meta-analysis but not much detail is given by the authors about the different subpopulations. **Figure 1** shows the primary analysis of the original published meta-analysis. The graph shows that some studies provide very narrow confidence intervals, suggesting a strong and statistically significant association between the use of antipsychotics and MI. The diamond of the meta-analysis also resulted in an increased risk of MI with antipsychotics (OR 1.88, 95% CI 1.39 to 2.54). However, the apparent large heterogeneity ( $I^2=0.98$ ,  $\tau^2=.30$ ) implies that important discrepancies exist among studies, and consequently, it is questionable whether this summary effect is useful and meaningful.

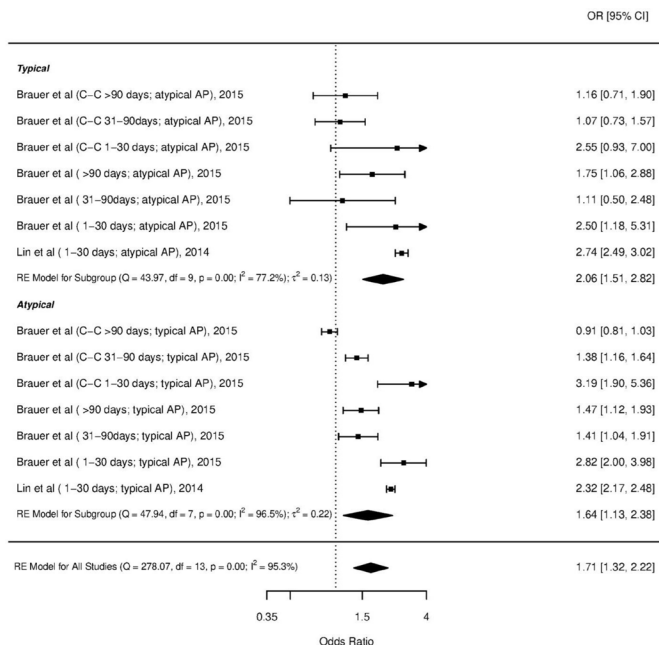
To identify possible sources of heterogeneity, we first performed a subgroup analysis by study design (**figure 2**). It is interesting that important heterogeneity is also observed within each design, implying that multiple characteristics differentiate the studies. **Figure 2** reveals some important issues in this dataset. It seems that in populations with shorter exposure, ORs tend to be larger. Indeed, performing a metaregression assessing the impact of the different exposure times in the studies (as these were reported in the supplementary material of the original article) remarkably reduced the heterogeneity and produced a significant coefficient of (in OR scale) 0.47 (95% CI 0.37 to 0.59), suggesting that the OR in studies with exposure larger than 30 days is on average half than that in the other studies (**table 1**). This constitutes a large difference and raises concerns whether short exposure studies are comparable to the other studies. In addition, it seems that within the cohort design, earlier studies tend to give larger OR estimates. The small number of cohort studies does not allow statistical assessment of the effect of publication year. However, careful examination of study characteristics and comparison between older and more recent studies could help to identify what might have changed over the years that affects the results.

We also performed a subgroup analysis by drug type (**figure 3**). Again, important heterogeneity exists within the two subgroups and particularly in the group of atypical drugs. Thus, the type of drug does not seem to explain much of the observed heterogeneity. According to the published review, the OR was larger among schizophrenic patients in comparison with other diagnostic categories, but again the heterogeneity within the design was quite large (data could not be retrieved).



**Figure 2** Subgroup analysis by study design (AP=antipsychotics, RE=random effects).

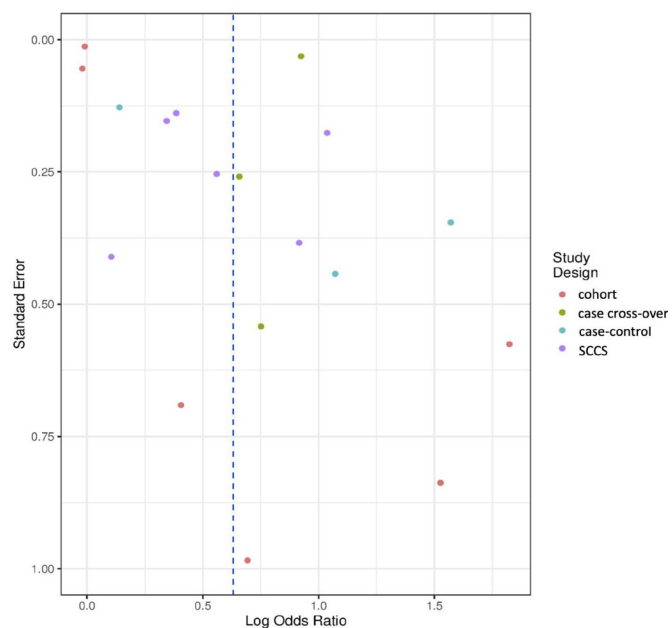




**Figure 3** Subgroup analysis by type of antipsychotic drug (typical or atypical). AP=antipsychotics, RE=random effects.

All the characteristics examined so far do not capture the potential for methodological heterogeneity across studies. This would require information on the adjustment factors and other aspects of the analysis (eg, handling missing data) that are not available in the manuscript. A sensitivity analysis where low-quality studies (as defined in the original review) were excluded resulted in a slightly decreased risk of MI but was still significant (OR 1.71, 95% CI 1.23 to 2.37). Hence, this is an indication that methodological differences might affect the results of the studies.

Finally, following the original review, we reproduced the funnel plot of all studies, but we indicated the different study designs



**Figure 4** Funnel plot of the studies indicating the different study designs with different colours. SCCS, self-controlled case series.

with different colours (figure 4). In this way, we could figure out whether the asymmetry can be explained by the different sample sizes among the different designs. A possible trend of asymmetry can be seen among cohort studies but with only six studies, asymmetry cannot be formally assessed. Furthermore, since the two groups of ‘imprecise’ (at the bottom) and ‘precise’ (at the top) estimates come from the same studies, they most probably reflect the heterogeneity already discussed, while any inference on the presence of publication bias is impossible here. However, the authors searched only three large databases, making it possible for some available studies not to have been identified.

**DISCUSSION**

Limitations of observational studies are well-known and, if not properly taken into account within a meta-analysis, they can threaten the validity of the findings. Moreover, the important heterogeneity expected among populations, settings and methodologies of different observational studies complicates even more the synthesis of their results. However, observational data may be advantageous to RCTs in terms of the amount of information, the generalisability of the findings and the evaluation of rare and long-term outcomes. Also, there are clinical outcomes for which RCTs cannot be performed due to ethical or feasibility reasons (eg, effect of smoking). As a result, meta-analyses of observational data sometimes are necessary to address questions for which randomised evidence is insufficient or absent.

Through an exemplar meta-analysis of observational studies assessing the risk of myocardial infarction in patients receiving antipsychotics, we show how ignoring discrepancies among different studies may lead to spurious conclusions. Investigators should pay special attention to investigate and address the possible biases separately for the different types of studies. It should be noted that, in systematic reviews involving several observational studies alone or in combination with RCTs, quantitative synthesis should not be by default a prominent component.<sup>2</sup> Careful examination of the possible sources for heterogeneity and risk of bias is necessary to decide whether a statistical combination of the data would give useful and meaningful results.

Interestingly, a review of the existing empirical evidence regarding the comparison of estimated treatment effects obtained from randomised and non-randomised studies has suggested that there is only little evidence for significant differences between the two.<sup>20</sup> However, clinical and methodological differences among studies are not always reflected in the form of statistical discrepancies, and consequently, the similarity of the studies of a systematic review should always be judged for the specific questions and outcomes of the review.<sup>21</sup> Ideally, sources of heterogeneity and biases are better explored when individual participant data are available, and therefore, meta-analysts should seek for such data when possible.<sup>22</sup> Nonetheless, most studies will in practice only provide aggregate data, and thus, clinical understanding of both the studies at hand and the assumptions of the different statistical approaches are important aspects for choosing plausible modelling strategies.

**Contributors** SM performed the analyses. SM and AC drafted the manuscript and both accepted its final version.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

## ORCID iD

Anna Chaimani <http://orcid.org/0000-0003-2828-6832>

## REFERENCES

- 1 Gurevitch J, Koricheva J, Nakagawa S, *et al.* Meta-Analysis and the science of research synthesis. *Nature* 2018;555:175–82.
- 2 Egger M, Schneider M, Davey Smith G. Spurious precision? meta-analysis of observational studies. *BMJ* 1998;316:140–4.
- 3 Sterne JA, Hernán MA, Reeves BC, *et al.* ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.
- 4 Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet* 2005;365:82–93.
- 5 Graham S, McDonald L, Wasiak R, *et al.* Time to really share real-world data? *F1000Res* 2018;7:1054.
- 6 Brugha TS, Matthews R, Morgan Z, *et al.* Methodology and reporting of systematic reviews and meta-analyses of observational studies in psychiatric epidemiology: systematic review. *Br J Psychiatry* 2012;200:446–53.
- 7 Lemeshow AR, Blum RE, Berlin JA, *et al.* Searching one or two databases was insufficient for meta-analysis of observational studies. *J Clin Epidemiol* 2005;58:867–73.
- 8 Cuschieri S. The STROBE guidelines. *Saudi J Anaesth* 2019;13:31–4.
- 9 Mueller M, D'Addario M, Egger M, *et al.* Methods to systematically review and meta-analyse observational studies: a systematic scoping review of recommendations. *BMC Med Res Methodol* 2018;18:44.
- 10 Thompson S, Ekelund U, Jebb S, *et al.* A proposed method of bias adjustment for meta-analyses of published observational studies. *Int J Epidemiol* 2011;40:765–77.
- 11 Eddy DM, Hasselblad V, Shachter RD. *Meta-analysis by the confidence profile method: the statistical synthesis of evidence [Internet]*. Boston : Academic Press, 1991.
- 12 Turner RM, Spiegelhalter DJ, Smith GCS, *et al.* Bias modelling in evidence synthesis. *J R Stat Soc Ser A Stat Soc* 2009;172:21–47.
- 13 Brumback BA, Holmes LB, Ryan LM. Adverse effects of chorionic villus sampling: a meta-analysis. *Stat Med* 1999;18:2163–75.
- 14 Prevost TC, Abrams KR, Jones DR. Hierarchical models in generalized synthesis of evidence: an example based on studies of breast cancer screening. *Stat Med* 2000;19:3359–76.
- 15 Schmitz S, Adams R, Walsh C. Incorporating data from various trial designs into a mixed treatment comparison model. *Stat Med* 2013;32:2935–49. 1097–0258 (Electronic).
- 16 Efthimiou O, Mavridis D, Debray TPA, *et al.* Combining randomized and non-randomized evidence in network meta-analysis. *Stat Med* 2017;36:1210–26.
- 17 Verde PE, Ohmann C. Combining randomized and non-randomized evidence in clinical research: a review of methods and applications. *Res Synth Methods* 2015;6:45–62.
- 18 Stroup DF, Berlin JA, Morton SC, *et al.* Meta-Analysis of observational studies in epidemiology: a proposal for reporting. meta-analysis of observational studies in epidemiology (moose) group. *JAMA* 2000;283:2008–12.
- 19 Yu Z-H, Jiang H-Y, Shao L, *et al.* Use of antipsychotics and risk of myocardial infarction: a systematic review and meta-analysis. *Br J Clin Pharmacol* 2016;82:624–32.
- 20 Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst Rev* 2014;4:MR000034.
- 21 Deeks JJ, Higgins JP, Altman DG. Analysing data and undertaking meta-analyses. *Cochrane Handb Syst Rev Interv* 2019:241–84.
- 22 Riley RD, Lambert PC, Abo-Zaid G. Meta-Analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;340:c221.