



Designing and analysing clinical trials in mental health: an evidence synthesis approach

Simon Wandel,¹ Satrajit Roychoudhury²

¹Novartis Pharma AG, Basel, Switzerland; ²Novartis Pharmaceuticals, East Hanover, NJ, USA

Correspondence to Dr Simon Wandel, Novartis Pharma AG, P.O. Box 4002 Basel, Switzerland; simon.wandel@novartis.com

ABSTRACT

Objective When planning a clinical study, evidence on the treatment effect is often available from previous studies. However, this evidence is mostly ignored for the analysis of the new study. This is unfortunate, since using it could lead to a smaller study without compromising power. We describe a design that addresses this issue.

Methods We use a Bayesian meta-analytic model to incorporate the available evidence in the analysis of the new study. The shrinkage estimate for the new study integrates the evidence from the other studies. At the planning phase of the study, it allows a statistically justified reduction of the sample size.

Results The design is illustrated using data from an Food and Drug Administration (FDA) review of lurasidone for the treatment of schizophrenia. Three studies inform the meta-analysis before the new study is conducted. Results from an additional phase III study, which were not available at the time of the FDA review, are then used for the actual analysis.

Conclusions In the presence of reliable and relevant evidence, the design offers a way to conduct a smaller study without compromising power. It therefore fills a gap between the assessment of evidence and its actual use in the design and analysis of studies.

INTRODUCTION

Clinical trials are a cornerstone of evidence-based medicine. The impressive number of more than 220 000 registered trials in the electronic database <http://www.clinicaltrials.gov> supports this statement, and further growth at an accelerated rate is likely. While this number shows the importance of clinical research, it also reveals that typically more than one clinical study investigating the same treatment (or interventions) is conducted. When designing a clinical study, it is therefore good practice to review the literature¹ and assess the body of evidence. This will provide information on the previous experience in the field and relevant topics such as commonly used end points, challenges with missing data and expected patient enrolment. It may also undermine the need for a new study.^{1 2} The latter is specifically important since resources to fund studies are limited and competition is high, whether in academia or industry. While all of the preceding reasons justify the need to conduct a systematic review before starting a new study, this is even more critical for statistical reasons. Defining end points, analysis sets, primary and secondary analyses and, most importantly, the sample size calculation critically depends on realistic and reliable assumptions. However, for all of these tasks, the information (data) from the review is used only indirectly in the new study. For example, an estimated treatment effect from a meta-analysis can serve as the base for the sample size calculation, yet it will not contribute to the actual analysis of the study. This is unfortunate, since it means that in the actual analysis, we ignore what we know already. Using the information in the analysis, too, could lead to a smaller study or to higher power. The question therefore is the following: is there an approach that would allow this information to be used in the actual analysis? This question is not new and has attracted interest already more than 40 years ago,^{3–5} and numerous contributions have been made since then.^{6–10} Most of the work considered the use of historical information to reduce (or even completely replace¹¹) the control group in the actual study. Such designs have been implemented successfully,^{12 13} including new drug approvals in epilepsy.¹⁴ Here, however, we are not interested in the control group alone: we aim to use the available evidence on the relative treatment effect in the analysis of our actual trial. In order to do so, we will use a Bayesian evidence synthesis design that relies on random-effects meta-analysis.^{15 16} In the following sections, we provide a primer on Bayesian meta-analysis for evidence synthesis, illustrate the approach describing the design and analysis of a study, and end with some conclusions.

A PRIMER ON BAYESIAN META-ANALYSIS FOR EVIDENCE SYNTHESIS AND SHRINKAGE ESTIMATES

Bayesian meta-analysis

Most readers may be familiar with meta-analysis conducted in the classical (frequentist) framework. However, we will use the Bayesian approach, which has an important role in evidence synthesis,¹⁶ including network meta-analysis¹⁷ and health technology assessment.¹⁸ While we cannot discuss Bayesian statistics in detail here, we refer the interested reader to an excellent introduction by Spiegelhalter *et al.*¹⁶

There are two main reasons behind our choice of the Bayesian approach: it is simpler to derive some of the quantities (estimates) we are interested in, and it is straightforward to account for uncertainty in the between-trial heterogeneity. The latter is particularly important when only few studies are included in the meta-analysis,¹⁹ which is not an uncommon case for systematic reviews.²⁰

When conducting a Bayesian meta-analysis, we undertake the same steps as for a classical one, with one exception: the actual analysis differs. For a classical meta-analysis, we use one of the many available frequentist methods.²¹ For a Bayesian meta-analysis, we simply perform a Bayesian analysis. Yet now, we have to specify what is known as prior distributions for the model parameters. Prior distributions reflect our knowledge about the parameters (the overall effect and the between-trial heterogeneity) before we have observed the data. For the overall effect, since we do not know anything about it, we choose a non-informative prior, meaning it does not favour any particular value. On the other hand, we know that the between-trial heterogeneity is usually small to substantial, and our prior reflects this accordingly.^{16 19} However, what is small or large between-trial heterogeneity?

This question cannot be answered without knowing the outcome scale. For a general definition, we therefore need an approach that connects the between-trial heterogeneity to the outcome scale in order to make it scale invariant (note that I^2 ²² has the same property). The key is to use a quantity known as the outcome standard deviation (SD), which is the SD describing the variation between participants within a study. The ratio of the between-trial heterogeneity to the outcome SD allows us to quantify the heterogeneity independently of the outcome scale; a classification is given in [table 1](#). An advantage of this approach is that we are able to communicate the typical (median) between-trial heterogeneity, as well as the uncertainty associated with it. For example, our analysis may reveal typical between-trial heterogeneity of moderate size, yet the 95% interval may cover very small to very large values.

This indicates that we are quite unsure about the true between-trial heterogeneity, which we need to keep in mind when discussing our results. This is a difference to the commonly used I^2 measure,²² which does not quantify the uncertainty of the between-trial heterogeneity.

Shrinkage estimate and effective sample size (ESS)

We now discuss a feature of a Bayesian meta-analysis that constitutes the key ingredient of the evidence synthesis design for clinical trials. In a meta-analysis, we typically summarise the results presenting the overall effect (point estimate and corresponding 95% interval). This is relevant information, yet not the only one we can retrieve from a random-effects meta-analysis. Since a random-effects meta-analysis is a hierarchical model,²³ there are additional estimates that are useful, one of which is the shrinkage estimate.¹⁶ It corresponds to the estimate of the treatment effect for a *specific* study in our meta-analysis, while accounting for the data from the *other* studies. This point is important, since it makes it clear that the shrinkage estimate will not be the same as the study-specific estimate (the study-specific estimate corresponds to the analysis of the specific study alone, shown beside each study in the forest plot). In fact, the shrinkage estimate is a weighted average of the study-specific estimate and the overall mean, with weights depending on the size of the study and the between-trial heterogeneity. Therefore, it is shrunken towards the overall mean and the 95% interval is narrower as compared with the study-specific estimate. The latter is particularly useful, showing that it borrows (incorporates) information from the other studies. If we aim at incorporating the available evidence in the analysis of our actual study, we therefore perform a meta-analysis of all studies and base our inferential statements on the shrinkage estimate for our actual study. This is exactly the approach that we use in the application.

However, even though we now have the desired estimate at hand, we still face a problem: when designing the study, we do not have its data yet, and cannot calculate the shrinkage estimate. How could we therefore derive, for example, our sample size for which the standard error (SE) of our estimate is crucial?

There is a surprisingly simple answer to that, which lies at the heart of Bayesian statistics. We perform a meta-analysis as planned, but include the actual study with missing data. With this approach, we obtain a shrinkage estimate for the actual study that represents the amount of information borrowed from the other studies. Interestingly, it corresponds exactly to the predicted effect for a new study,¹⁰ a quantity which has been recommended to be reported in addition to the overall effect for any meta-analysis.^{23–25} A useful property of the predicted effect is that it accounts for the uncertainty in the between-trial heterogeneity. Large heterogeneity between the available studies thus leads to a wide prediction interval, meaning that we can only borrow little information. Therefore, it protects against overly optimistic use of the available evidence.

Finally, we can now also quantify the amount of information that we will borrow from the other studies when using the shrinkage estimate. For that, we use the ESS. Conceptually, the ESS provides the answer to the following question: what is the equivalent number of patients that we would need to enroll in the new study in order to get the same amount of information that we now borrow from the other studies?

Table 1 Heterogeneity of a meta-analysis classified according to the ratio of the between-trial heterogeneity to the outcome SD

Heterogeneity	Ratio of the between-trial heterogeneity to the outcome SD
Small	0.0625
Moderate	0.125
Substantial	0.25
Large	0.50
Very large	1.00

Technically, there are different ways to derive the ESS, and we will use a simple approximation here.⁸ When designing the actual study, we can then derive the sample size as for a frequentist design and simply subtract the ESS from it. This highlights that incorporating existing data will, as expected, lead to a reduction in the sample size for the new study.

CASE STUDY: EVIDENCE-BASED DECISION-MAKING FOR A SCHIZOPHRENIA DRUG

In this section, we provide a hypothetical design and analysis example of schizophrenia to illustrate the evidence synthesis design. We use data from the Food and Drug Administration (FDA) review of lurasidone (lurasidone hydrochloride) for the treatment of schizophrenia.²⁶ For simplicity, we restrict our considerations to the comparison between the lurasidone 80 mg dose and placebo, based on data from two phase 2 (D1050049,^{26 27} D1050196^{26 28}) and one phase 3 (D1050229^{26 29}) study. The purpose of this example is only to illustrate the evidence synthesis design. Specifically, we do not intend to make any clinical or regulatory interpretation of the data or results; for these, we refer the interested reader to the respective regulatory documents²⁶ and the medical literature, for example, a recent review by Leucht *et al.*³⁰ The code to reproduce the example is given in web appendix 1.

An important end point to assess clinical benefit for patients with schizophrenia is the change from baseline in the Clinical Global Impressions—Severity Scale (CGI-S).³¹ The CGI-S captures mental illness of patients on a seven-point scale (1=normal to 7=among the most extremely ill patients). A negative value for change from baseline in the CGI-S score therefore implies that the patient's condition has improved over time. The change from baseline in the CGI-S at 6-weeks was the key secondary end point for all three trials; the results are summarised in table 2. Statistically significant effects in favour of lurasidone were found in studies D1050196 and D1050229, whereas no statistically significant effect was seen in study D1050049. Overall, the results of these trials appear to be homogeneous for the treatment effect, despite some variation in the placebo response. The latter, however, is not uncommon for longitudinal studies and seems to have no influence on the treatment effect, at least as far as the data in table 2 are considered.

Design of the next study

When we want to incorporate the existing information in the analysis of our next study, we need to assume similarity between all studies. It is therefore important to assess that no relevant differences with regard to factors such as patient inclusion/exclusion criteria, experience of investigators, study location, etc, exist between the available studies and the new one. When we would be willing to include the results of the new study, once available, in an updated meta-analysis, this gives us reassurance that we consider the similarity assumption as plausible. Here, we will assume that this is the case.

Table 2 Change from Baseline in CGI-S for three clinical studies of lurasidone in schizophrenia

Study	Number of patients lurasidone/placebo	CGI-S 6-week change from baseline		
		Placebo mean (SE)	Difference between lurasidone and placebo mean (SE)	p Value*
D1050049	71/72	−0.70 (0.11)	−0.09 (0.15)	0.5950
D1050196	90/90	−0.20 (0.11)	−0.41 (0.15)	0.0072†
D1050229	121/127	−1.00 (0.10)	−0.40 (0.10)	0.0340†

*Original p values from the FDA review.

†Statistically significant.

FDA, Food and Drug Administration; CGI-S, Clinical Global Impressions—Severity Scale.

In a first step, we calculate the required sample size for the new 1:1 randomised study under a classical (frequentist) design, assuming that CGI-S would be the primary outcome. Using approximate normality of the change from baseline in CGI-S at 6 weeks, we can use sample size formulas for testing a difference in means, available in standard textbooks.³² Four quantities are required: the outcome SD, the significance level, the power and the assumed true treatment effect (the alternative to power the study). From study D1050229, we find the outcome SD to be 0.787, and for the other quantities, we use a significance level of 5%, power equal to 90%, and an assumed true treatment effect of -0.32 . We then find the sample size to be 128 patients per arm; again, all calculations can be found in web appendix 1.

Since we want to incorporate the available information on the treatment effect in order to reduce the sample size, we now need to perform a meta-analysis using the data from the three available studies (table 2). As mentioned in section 2, we will use a non-informative prior for the overall effect. Additionally, the prior for the between-trial heterogeneity has a median of 0.17 and a 95% interval (0.01 to 0.56), corresponding to typical between-trial heterogeneity of moderate size, but allowing for very small to substantial values (95% interval). This appears plausible given the similarity of the included studies.

Figure 1 displays the forest plot of the meta-analysis. The analysis indicates small between-trial heterogeneity; the median is now 0.13 with a corresponding 95% interval (0.00 to 0.38). The results of the three existing studies thus appear indeed to be quite homogeneous, and we expect that this will translate into a reasonable amount of information incorporated in the new study.

As discussed before, the quantity of interest is the predicted treatment effect. In figure 1, it is presented below the overall effect. The predicted effect has the same median as the overall effect (-0.32), but a wider 95% interval (-0.81 to 0.20), because it accounts for the uncertainty in the between-trial heterogeneity. However, already graphically, it

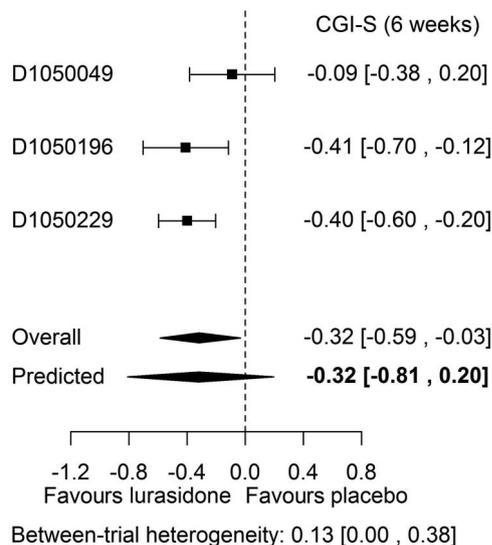


Figure 1 Meta-analysis of the 6 weeks change from baseline in Clinical Global Impressions—Severity Scale (CGI-S), lurasidone versus placebo. Shown is the difference in the 6 weeks CGI-S change from baseline (lurasidone vs placebo) with negative values favouring lurasidone. No effect corresponds to a value of zero, which is indicated by the dotted line. The straight lines with the squares are the study-specific results. The diamonds represent the overall and the predicted effect. The predicted effect (results in bold) corresponds to the effect that we would expect in a next study. For the between-trial heterogeneity, the typical (median) value and the 95% interval are given. The 0.13 indicates moderate heterogeneity, and the interval ranges from very small to substantial-to-large heterogeneity.

appears that there is substantial information on the treatment effect in the new study available. When we calculate the ESS, it turns out to be around 53 patients, which is approximately equivalent to 26 patients randomised to each arm in the new study. Therefore, we will reduce the sample size accordingly, from the original 128 patients per arm to 102 now, expecting the other studies to contribute this amount of information in the actual analysis.

Analysis of the next study

In the previous section, we have designed a hypothetical next study incorporating the available evidence. However, at the time when the FDA review of lurasidone was happening, there was actually another relevant phase III study (D1050233)^{26 33} ongoing, which had four arms, including the 80 mg lurasidone ($n=125$) and placebo ($n=120$) arms. Here, we will therefore use these data to illustrate the actual analysis. On completion, the trial showed a strong treatment effect of lurasidone 80 mg as compared with placebo (change from baseline in CGI-S for treatment vs placebo: -0.60 (-0.87 to -0.33)). This is clearly a more pronounced effect than was observed in the other three studies, and therefore we expect the corresponding shrinkage estimate to have a less extreme point estimate, since it shrinks the study-specific estimate towards the overall mean (see section ‘Shrinkage estimate and effective sample size (ESS)’). Indeed, the shrinkage estimate has a median of -0.48 and a 95% interval (-0.74 to -0.26), as shown in figure 2. It still excludes a null effect, but the treatment effect is less pronounced, which we may find more reasonable given the evidence from the other studies. The interval is also narrower than the stratified one, and reveals a $\sim 25\%$ information gain, approximately what we expected. This, however, is only the case since the analysis still indicates small

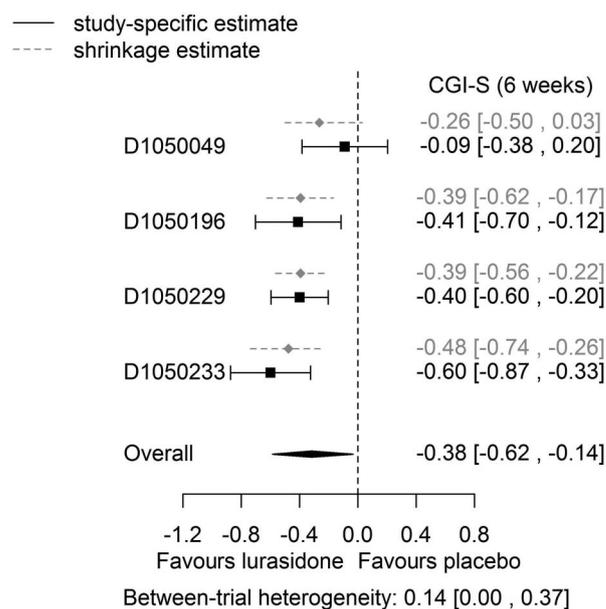


Figure 2 Meta-analysis of the 6 weeks change from baseline in Clinical Global Impressions—Severity Scale (CGI-S), lurasidone versus placebo, including studies D1050049, D1050196, D1050229 and D1050233. Shown is the difference in the 6 weeks CGI-S change from baseline (lurasidone vs placebo) with negative values favouring lurasidone. No effect corresponds to a value of zero, which is indicated by the dotted line. The straight lines with the squares are the study-specific results. The dotted grey lines with the rotated squares represent the shrinkage estimates. These correspond to the estimates for each study incorporating the data from the other studies. The diamond represents the overall effect. For the between-trial heterogeneity, the typical (median) value and the 95% interval are given. The 0.14 indicates moderate heterogeneity, and the interval ranges from very small to substantial-to-large heterogeneity.

between-trial heterogeneity. Finally, even though not relevant for the analysis of the actual study, for completeness we also show the shrinkage estimates of the other studies in figure 2.

CONCLUSIONS

Evidence-based medicine has shifted the way we evaluate new treatments and interventions. It has emphasised the need for high-quality clinical studies, as well as helped to put studies into context, by considering their results in the light of other clinical evidence.³⁴ This is a remarkable paradigm shift and the success of networks such as the Cochrane Collaboration (<http://www.cochrane.org>) show the value that this approach brings to decision-makers. Additionally, evidence-based medicine has an important role before treatment policy decisions are made. Nowadays, a systematic review should precede any clinical trial, since this allows one to assess the available evidence, justify the need for the actual study^{1 2} and support important design specifications.^{35 36} However, the next logical step, which is using the actual evidence in the analysis of a new study, is still often missing. As we have outlined, this is unfortunate, since the inclusion of evidence would result in a reduction of the sample size, which means that promising treatments (or interventions) may reach patients quicker. Importantly, the main assumption for this approach is the same as for any meta-analysis: the similarity of the studies, in our case the existing ones and the new study. If this assumption is questionable, one should carefully decide whether the proposed design is really appropriate. Additionally, more sophisticated statistical models,^{10 37} which automatically adapt to the situation when there is a substantial difference between the results of the actual study and the available evidence, may then be needed. Since we use the Bayesian approach, the choice of the prior distributions is critical, especially for the between-trial heterogeneity. Its prior distribution needs to reflect a plausible range of values; alternatively, it could be based on evidence, see, for example, the work by Rhodes *et al.*³⁸ Typically, one would also explore the operating characteristics (type I error, power) more extensively, using different assumed true scenarios for the treatment effect in the new study. Finally, we also note that there are situations where the discussed design may not be used. Whenever strict type I error control is essential, we need to consider other designs, because the type I error can be inflated (but it could equally be deflated). Similarly, whenever Bayesian approaches are excluded a priori, the design will not be accepted. It is therefore important to get clarity on these points first, before implementing the design in practice.

Competing interests SW and SR are employees of Novartis.

Provenance and peer review Not commissioned; externally peer reviewed.

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/eb-2016-102491>).

doi:10.1136/eb-2016-102491

Received 23 August 2016; Revised 7 September 2016; Accepted 8 September 2016

REFERENCES

- Carpenter LM. Is the study worth doing? *Lancet* 1993;**342**:221–3.
- Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Obstet Gynecol* 2009;**114**:1341–5.
- Pocock S. The combination of randomized and historical controls in clinical trials. *J Chronic Dis* 1976;**29**:175–88.
- Browne R. Reducing sample sizes when comparing experimental and control groups. *Arch Environ Health* 1976;**31**:169–70.
- Van Ryzin J. Designing for nonparametric Bayesian survival analysis using historical controls. *Cancer Treat Rep* 1980;**64**:503–6.
- Ryan L. Using historical controls in the analysis of developmental toxicity data. *Biometrics* 1993;**49**:1126–35.
- Simon R. Bayesian design and analysis of active control clinical trials. *Biometrics* 1999;**55**:484–7.
- Neuenschwander B, Capkun-Niggli G, Branson M, *et al.* Summarizing historical information on controls in clinical trials. *Clin Trials* 2010;**7**:5–18.
- Viele K, Berry S, Neuenschwander B, *et al.* Use of historical control data for assessing treatment effects in clinical trials. *Pharm Stat* 2014;**13**:41–54.
- Schmidli H, Gsteiger S, Roychoudhury S, *et al.* Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics* 2014;**70**:1023–32.
- French J, Wang S, Warnock B, *et al.* Historical control monotherapy design in the treatment of epilepsy. *Epilepsia* 2010;**51**:1936–43.
- Di Scala L, Kerman J, *et al.* Collection, synthesis, and interpretation of evidence: a proof-of-concept study in COPD. *Stat Med* 2013;**32**:1621–34.
- Baeten D, Baraliakos X, Braun J, *et al.* Anti-interleukin-17A monoclonal antibody secukinumab in treatment of ankylosing spondylitis: a randomised, double-blind, placebo-controlled trial. *Lancet* 2013;**382**:1705–13.
- FDA. Aptom approval letter. 2015. http://www.accessdata.fda.gov/drugsatfda_docs/applletter/2015/022416Orig1s001ltr.pdf. (accessed 17 August 2016).
- Egger M, Davey-Smith G, Altman D. *Systematic reviews in health care: meta-analysis in context*. 2nd edition. BMJ Publishing Group, 2001.
- Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian approaches to clinical trials and health-care evaluation*. Wiley & Sons, 2004.
- Mavridis D, Giannatsi M, Cipriani A, *et al.* A primer on network meta-analysis with emphasis on mental health. *Evid Based Ment Health* 2015;**18**:40–6.
- Dias S, Welton NJ, Sutton AJ, *et al.* NICE DSU Technical Support Document 1: introduction to evidence synthesis for decision making. 2011, (last updated April 2012). <http://www.nicedsu.org.uk>
- Friede T, Röver C, Wandel S, *et al.* Meta-analysis of few small studies in orphan diseases. *Res Synth Methods*. Published Online First: 30 June 2016. doi: 10.1002/jrsm.1217.
- Turner RM, Davey J, Clarke MJ, *et al.* Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *Int J Epidemiol* 2012;**41**:818–27.
- Veroniki AA, Jackson D, Viechtbauer W, *et al.* Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res Synth Methods* 2016;**7**:55–79.
- Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;**21**:1539–58.
- Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc* 2009;**172**:137–59.
- Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ* 2011;**342**:d549.
- Guddat C, Grouven U, Bender R, *et al.* A note on the graphical presentation of prediction intervals in random-effects meta-analyses. *Syst Rev* 2012;**1**:34. http://www.accessdata.fda.gov/drugsatfda_docs/nda/2010/200603orig1s000medr.pdf (accessed 29 Aug 2016).
- Potkin SG, Kimura T, Guarino J. A 6-week, double-blind, placebo- and haloperidol-controlled, phase II study of lurasidone in patients with acute schizophrenia. *Ther Adv Psychopharmacol* 2015;**5**:322–31.
- Nakamura M, Ogasa M, Guarino J, *et al.* Lurasidone in the treatment of acute schizophrenia: a double-blind, placebo-controlled trial. *J Clin Psychiatry* 2009;**70**:829–36.
- Nasrallah HA, Silva R, Phillips D, *et al.* Lurasidone for the treatment of acutely psychotic patients with schizophrenia: a 6-week, randomized, placebo-controlled study. *J Psychiatr Res* 2013;**47**:670–7.
- Leucht S, Cipriani A, Spineli L, *et al.* Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *Lancet* 2013;**382**:951–62.
- Guy W. *ECDEU Assessment Manual for Psychopharmacology—Revised (DHEW Publ No ADM 76-338)*. Rockville, MD: U.S. Department of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, NIMH Psychopharmacology Research Branch, Division of Extramural Research Programs, 1976:218–22.
- Kirkwood B, Sterne J. *Essential medical statistics*. Wiley-Blackwell, 2003.
- Loebel A, Cucchiari J, Sarma K, *et al.* Efficacy and safety of lurasidone 80 mg/day and 160 mg/day in the treatment of schizophrenia: a randomized, double-blind, placebo- and active-controlled trial. *Schizophr Res* 2013;**145**:101–9.
- Pocock SJ, Stone GW. The primary outcome fails—what next? *N Engl J Med* 2016;**375**:861–70.
- Nikolakopoulou A, Mavridis D, Salanti G. Planning future studies based on the precision of network meta-analysis results. *Stat Med* 2016;**35**:978–1000.
- Sutton AJ, Cooper NJ, Jones DR, *et al.* Evidence-based sample size calculations based upon updated meta-analysis. *Stat Med* 2007;**26**:2479–500.
- Neuenschwander B, Wandel S, Roychoudhury S, *et al.* Robust exchangeability designs for early phase clinical trials with multiple strata. *Pharm Stat* 2016;**15**:123–34.
- Rhodes KM, Turner RM, Higgins JP. Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. *J Clin Epidemiol* 2015;**68**:52–60.