

# A practical guide to the meta-analysis of rare events

Orestis Efthimiou<sup>1</sup>

<sup>1</sup>Institute of Social and Preventive Medicine, University of Bern, Switzerland.

Correspondence to [oremiou@gmail.com](mailto:oremiou@gmail.com)

## Abstract

Meta-analysing studies with low event rates is challenging as some of the standard methods for meta-analysis are not well suited to handle rare outcomes. This is more evident when some studies have zero events in one or both treatment groups. In this article, we discuss why rare events require special attention in meta-analysis, we present an overview of some approaches suitable for meta-analysing rare events and we provide practical recommendations for their use.

## Methods

We go through several models suggested in the literature for performing a rare events meta-analysis, highlighting their respective advantages and limitations. We illustrate these models using a published example from mental health. We provide the software code needed to perform all analyses in the appendix.

## Results

Different methods may give different results, and using a suboptimal approach may lead to erroneous conclusions. When data is very sparse, the choice between the available methods may have a large impact on the results. Methods that use the so-called continuity correction (e.g. adding 0.5 to the number of events and non-events in studies with zero events in one treatment group) may lead to biased estimates.

## Conclusions

Researchers should define the primary analysis a priori, in order to avoid selective reporting. A sensitivity analysis using a range of methods should be used to assess the robustness of results. Suboptimal methods such as using a continuity correction should be avoided.

## Introduction

When the outcome of interest is rare, e.g. for the case of adverse events, individual studies are often underpowered to detect treatment effects. Pooling together evidence from multiple clinical trials via a meta-analysis offers a way to increase power<sup>1</sup>. Performing such meta-analyses, however, might be a methodologically challenging task, especially when some of the studies reported no events in one or both treatment arms (we will refer to such studies as single-zero and double-zero studies respectively). The issue of rare events is in itself quite frequent; an empirical study found that 30% of a random sample of 500 Cochrane reviews contained at least one trial with zero events in one arm<sup>2</sup>. It is important to note at this point that there is no universally accepted definition of what constitutes a rare event. Probably most researchers would agree that a risk smaller than 1% would be enough to classify an event as rare. However, at the meta-analysis level, and depending on the size of the relevant studies, higher event rates might also lead to zero events if the sample sizes are sufficiently small.

Meta-analysing rare events requires special attention because standard methods are not well suited for the task. Probably the most popular method for performing a meta-analysis is the inverse variance method, either fixed-effect or using the DerSimonian and Laird random-effects model<sup>3,4</sup>. This method involves calculating a treatment effect separately from each study, along with a standard error. For binary outcomes the treatment effects are usually measured as odds ratios, risk ratios or risk differences. These study-specific estimates are then synthesized at a second level, across studies. An important aspect of the inverse variance method is that it uses a normal approximation of the true binomial likelihood (the 'large sample approximation'). This approximation does not work well when event rates are low, while for studies with no events in one or both arms, calculating treatment effects in terms of odds or risk ratios becomes impossible since it involves division by zero.

Consider for example the case of study  $i$ , which reported data in the form of the  $2 \times 2$  table shown in Table 1. The odds ratio  $OR_i$  for this study can be calculated as  $OR_i = \frac{a_i d_i}{b_i c_i}$  and the corresponding standard error, using the large sample approximation, is given by  $SE_i = \sqrt{\frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i}}$ . It is easy to see that when no events are observed in one or both treatment groups (i.e. when  $a_i = 0$  and/or  $c_i = 0$ ) then the odds ratio, the standard error, or both cannot be calculated. The same problem is present when calculating risk ratios as well. Thus, when one or more of the entries of Table 1 is zero, the inverse variance method cannot be used. Note that the problem also appears when all patients had the event in one or both treatment groups (i.e. when  $b_i = 0$  and/or  $d_i = 0$ )

One easy way to overcome the zero-cell problem in single-zero studies is to 'correct' the data. This is done by adding a fixed value (typically 0.5) to all cells of Table 1, for studies with zero events in one of their arms. This so-called 'continuity correction' bypasses the problem caused by zero events, and allows the use of the standard inverse variance methods. Another simple solution to the problem is to completely forfeit the use of odds ratios and risk ratios, and only use risk difference to measure relative

effects. This would work because the risk difference does not suffer from computational problems in the presence of zero events.

Unfortunately, both aforementioned solutions (applying a continuity correction or using the risk difference) have been found to be problematic for the meta-analysis of rare events<sup>5</sup>. Simulations showed that using the inverse variance method after applying a 0.5 continuity correction leads to excess bias in the estimated effects. In addition it has been showed that when events are rare, risk difference methods have poor statistical properties (they provide too wide intervals and have low power), which makes them unsuitable for meta-analysis<sup>5</sup>.

Double-zero studies are usually omitted from the meta-analysis (this is the default option for many statistical software). The Cochrane Handbook argues that such studies do not carry information regarding odds/risk ratios, and should therefore be excluded. Some researchers, however, pointed out that from an ethical point of view, patients in double-zero studies deserve to be included in the analyses<sup>6</sup>, while others discuss that such studies may carry information of relative treatment effects through their sample size<sup>7</sup>.

One additional issue with rare events is that, for the case of random effects meta-analysis, the estimation of the variance of random effects (heterogeneity) may be biased<sup>8</sup>, which may lead to spuriously narrow confidence intervals.

Thus, for the case of rare events meta-analysts need to resort to more advanced statistical methods. There are several alternatives that can be used to this end. In the Methods section of this paper we give a brief account of some of these methods. In the Results section we employ an example from a recent meta-analysis regarding mortality in antipsychotics<sup>9</sup> to illustrate the methods that will be discussed.

<<<TABLE 1 HERE>>>

## Methods

In this section we go through some of the methods that have been proposed for the meta-analysis of rare events.

### *I. Peto's method*

Peto's method<sup>10</sup> can only be used to estimate odds ratios. It is by definition a fixed-effects method, i.e. it cannot account for heterogeneity between trials, which is one of the limitations of this model. Peto's odds ratio follows an approach similar to the inverse variance model, but the effect estimate as well as the weight for each study are defined differently. The model incorporates evidence from single-zero studies without having to resort to continuity corrections. Double-zero studies are excluded from the analysis. Thus, the only instance when Peto's method runs into computational problems is when all studies in the meta-analysis are double-zero, i.e. when no events were observed in all studies. Simulations by Sweeting et al.<sup>11</sup> and Bradburn et al.<sup>5</sup> showed that the Peto odds ratio works reasonably well when the event is rare (< 1%), the treatment groups are balanced (i.e. there is approximately the same number of patients in the treatment and control arms within each study) and the effects are not

very large. If these conditions do not hold, Peto's method may give biased results. For this reason Cochrane does not recommend Peto as the default approach for rare-events meta-analysis<sup>1</sup>.

## **II. Mantel-Haenszel (MH) meta-analysis**

The MH method<sup>12</sup> is a different approach to fixed-effects meta-analysis. It can be used for odds ratios, risk ratios or risk differences, and it uses a different weighting scheme for each measure. The method incorporates evidence from single-zero studies without requiring continuity corrections, unless the same cell of Table 1 is zero for all studies (e.g. when  $a_i = 0$  for all  $i$ ). Thus, the method requires continuity corrections much less often. The MH method excludes from the analysis double-zero studies, unless risk difference is used. MH odds ratios have been shown to perform better than Peto's method, in cases where the latter performed poorly (e.g. when the treatment groups are unbalanced)<sup>5</sup>. Cochrane's software for meta-analyses (RevMan) uses MH as the default fixed-effect meta-analysis method. Note here that a random-effects MH approach is also implemented in RevMan 5. This however is not a 'true' MH model; it only uses the MH fixed-effect pooled result to estimate the heterogeneity, and then uses a DerSimonian and Laird (inverse variance) random-effects model. This means that this hybrid approach suffers from the usual problems that the inverse variance method faces when events are rare.

## **III. Using a 'treatment-arm' continuity correction**

As we already discussed, a simple way to bypass the complications associated with zero events is to add 0.5 to all cells of Table 1 (i.e. to  $a_i$ ,  $b_i$ ,  $c_i$  and  $d_i$ ). This method, however, has been shown to perform poorly, i.e. it may give extremely biased results, especially when groups are unbalanced<sup>11</sup>. Sweeting et al.<sup>11</sup> tried to mitigate these undesirable effects by using non-fixed corrections. In their approach, the continuity correction is different for each treatment arm of each study, and is inversely related to the size of the treatment arm. After implementing this correction to the data, standard approaches can be used for meta-analysis (e.g. inverse-variance, MH). The authors showed that a non-fixed continuity correction is preferable to the usual 0.5.

However, the use of continuity corrections has been criticized, and it has been pointed out that this essentially arbitrary correction (0.5 or any other number) can affect the results of the meta-analysis<sup>6,7</sup>.

## **IV. Logistic regression**

Logistic regression approaches use the correct binomial distribution of the data, and can be used to perform either a fixed or a random-effects meta-analysis. In this approach, single-zero studies are included without any continuity correction, while double-zero studies are excluded from the analysis. One general caveat of logistic regression is that in order to perform a random effects meta-analysis it is required to estimate the extent of heterogeneity of treatment effects, and this might be very difficult when events are rare<sup>8</sup>. Logistic regression (with unconditional binomial likelihood) has been shown to perform similarly with the MH odds-ratio without continuity correction<sup>5</sup>.

## **V. Bayesian meta-analysis**

Bayesian statistics is a branch of statistics in which the notion of probability corresponds to the state of knowledge regarding a certain phenomenon<sup>13</sup>, rather than the expected frequency of an event (which

corresponds to the so-called *frequentist* probability). Bayesian statistics are based in updating pre-existing evidence in the light of new data. In a Bayesian analysis, for each quantity of interest (e.g. the treatment effects, or the extent of heterogeneity in a meta-analysis) we assign a prior distribution. This distribution quantifies our prior knowledge regarding this quantity, along with some uncertainty. In practice, for some of the parameters of a model there might be no prior knowledge or we may wish our priors to have a minimal effect on the model's estimates. In such cases we can use 'uninformative' ('vague') distributions for these parameters.

For a Bayesian meta-analysis one can use an adaptation of the simple logistic regression model<sup>14,15</sup>. This requires specifying prior distributions for all model parameters. When data is sparse, the choice of prior distributions can be very important. Even if a prior distribution is intended to be uninformative, it might have substantial impact on the results<sup>16</sup>. This is especially true for the case of specifying priors for heterogeneity in a random-effects meta-analysis<sup>17</sup>.

This problem can be tackled by using reliable external information. For example, in a recent meta-epidemiological study<sup>18,19</sup>, Turner et al. analysed data from 15,000 binary outcome meta-analyses from the Cochrane Database of Systematic Reviews. The authors then used results to formulate informative prior distributions for the extent of between-study heterogeneity. These distributions cover 80 different settings with respect to the outcome being assessed, the nature of the interventions being compared etc. Such informative distributions can be used as priors for a random-effects Bayesian meta-analysis, to overcome the problem of having 'uninformative' priors dominating results.

#### **VI. *Beta-binomial with correlated responses***

The methods we have described up to this point (with the exception of the MH risk difference), exclude double-zero studies from the analyses. Kuss performed a simulation study<sup>7</sup> to compare several methods for meta-analysing rare events that do not exclude double-zero studies and do not use continuity corrections. Based on these simulations, Kuss concluded that a beta-binomial model with correlated responses may be the best approach to meta-analysing rare events in studies with balanced treatment groups. This approach models the probability of an event in each treatment group of each study using a binomial likelihood, and then pools this probability for each group across studies using a bivariate ('Sarmanov') beta distribution<sup>20</sup>. The beta-binomial model can be used to estimate odds/risk ratios and risk differences, and is by definition a random-effects model.

#### **VII. *Arcsine difference (AD)***

Rücker et al.<sup>21</sup> proposed the use the arcsine difference for summarizing treatment effects. Their approach calculates the treatment effect from each study separately along with its standard error. This is done using a measure called arcsine difference, which is a function of  $a_i$ ,  $b_i$ ,  $c_i$  and  $d_i$  of Table 1. The study-specific estimates of the arcsine difference are then combined using a standard fixed or random-effects approach. The advantage of this approach is that it incorporates evidence from single- and double-zero studies without requiring a continuity correction. It can also provide an estimate even in the extreme case of having only double-zero studies in the meta-analysis, i.e. when there was no event in

any study. The major disadvantage of this method is that arcsine difference is very hard to interpret. Thus, this method has not been used in practice very often. It may however be valuable as a secondary analysis, to assess the robustness of results (e.g., when another method has detected a treatment effect).

### **VIII. Other methods**

Several other models have appeared in the literature and can be used for the meta-analysis of rare events, such as an exact method based on combining confidence intervals<sup>22</sup>, a bivariate binomial-normal model<sup>16</sup>, a hypergeometric-normal model<sup>16</sup>, a Poisson-gamma model<sup>23</sup> and others. These methods have been rarely used in practice, and we will not consider them here in more detail.

#### ***Illustrative case study***

In order to illustrate the methods that we discussed in this section, we used a set of previously published data regarding mortality risk in antipsychotics<sup>9</sup>. This comprised two independent meta-analyses. The first one included 18 trials that compared long-acting injectable antipsychotics (LA-IAP) to placebo, for all-cause mortality. The event was very rare; the risk of death across both arms in all studies was around 0.2%. Only 7 events were reported in the drug arms (total 3774 patients) and 6 events in the placebo arm (2145 patients). The available data is summarized in Figure 1, where we also show the odds ratios and 95% Confidence Interval (CI) from each study, calculated using the large sample approximation. In this figure, in order to calculate odds ratios in single-zero studies we used a 0.5 continuity corrections, for illustration purposes. No odds ratios are shown for double-zero studies. The second meta-analysis compared LA-IAP to oral antipsychotics (OAP) for all-cause mortality. A total of 24 studies were included, reporting 15 deaths in LA-IAP (4059 patients) and 24 deaths in OAP (3820 patients). The total risk of death across both arms was around 0.5%. Figure 2 shows the available data.

In the original publication<sup>9</sup>, the authors performed both meta-analyses using the standard inverse-variance method, with a 0.5 continuity correction for single-zero studies. Here we re-analysed this data using a range of different methods. As we have already discussed, risk difference has been shown to perform poorly when it comes to meta-analysing rare events<sup>5</sup>, and we did not consider it here. Moreover, when the probability of an event is low, the difference between risk ratios and odds ratios becomes negligible. Thus, we only focused on odds ratios – an analysis of risk ratios would give almost identical results. For the Bayesian random-effects meta-analyses we used informative prior distributions for heterogeneity, based on the empirical study by Turner et al.<sup>19</sup>. For the first meta-analysis (LA-IAP vs placebo) we used the distribution corresponding to all-cause mortality, for pharmacological interventions vs. placebo. For the second meta-analysis (LA-IAP vs OAP) we used the distribution corresponding to all-cause mortality, for pharmacological vs. pharmacological interventions (Table IV in <sup>19</sup>).

All analyses were performed using freely available software (R<sup>24</sup> and OpenBUGS<sup>24</sup>). All software codes we used as well as some additional details regarding fitting the models are given in the Appendix.

FIGURE 1 HERE

FIGURE 2 HERE

## Results

Table 2 summarizes results from all analyses. It is evident that the choice of the method for meta-analysis is more important in the first example (LAI-AP vs placebo) as compared to the second (LAI-AP vs OAP). This is because events in the first dataset were sparser. In fact, out of the 18 studies of the LAI-AP vs placebo dataset only 1 had events in both arms – the rest were either single-zero (9 studies) or double-zero (8 studies). Conversely, in the second dataset there were 7 studies with events in both arms, 8 single-zero and 9 double-zero studies. This finding highlights a more general conclusion, i.e. the sparser the data, the larger the impact of the choice of meta-analysis method.

If we focus on the first example we can easily see that using a continuity correction (either 0.5 as in the original paper, or the treatment-arm continuity correction<sup>11</sup>) leads to narrower CIs compared to other methods, i.e. larger precision in the results. This should come as no surprise, as this method in essence imputes data in single-zero studies. Thus, this increase in precision is artificial. Moreover, it is obvious that different choices of the continuity correction lead to different results.

In the same example, the Peto method gives similar, but not identical results to Mantel-Haenszel with no continuity correction. A look at the data might convince us that Peto's method is suboptimal for this particular example, as there are studies with large imbalances, e.g. there are studies with 3:1 randomization ratio.

The fixed-effect Bayesian model gave almost identical results to the Mantel-Haenszel approach (without continuity correction). The effect of modeling random effects was rather minimal in both analyses, as the estimated value for heterogeneity was small (detailed results in the Appendix).

The beta-binomial with correlated responses model failed to converge for the first example. This highlights one of the potential disadvantages of this model, as compared to the other approaches. For the second example it gave results comparable to the rest of the methods.

Finally, in Table 2 we also show the results from the arcsine difference meta-analysis. It should be obvious that interpreting this effect measure in a clinically meaningful way can be very difficult.

TABLE 2 HERE

## Discussion

Different methods for meta-analysing rare events may lead to different conclusions, and the sparser the data the larger the differences between the results of the alternative methods. This was highlighted in a much-publicized meta-analysis regarding the effects of rosiglitazone on the risk of myocardial infarction and death. The original meta-analysis<sup>25</sup> gave results very close to the conventional threshold of 'statistical significance' (i.e.  $p\text{-value} < 0.05$ ). Subsequent meta-analyses of the same dataset using alternative methods led to (slightly) different results than the original publication<sup>23,26</sup>, which had an impact on the statistical significance of the findings.

This ambiguity stems from the fact that there is currently no clear answer as to which is the best model for meta-analysing rare events. Different models employ different assumptions, whose validity is usually difficult (or even impossible) to assess. Keeping that in mind, we highlight several general considerations that researchers can take into account when setting off to perform a meta-analysis of rare events:

1. The use of artificial continuity corrections should be avoided (with an exception perhaps for visualizing evidence<sup>7</sup>, e.g. as we did in Figures 1 and 2).
2. Risk difference is usually not the optimal effect measure to use.
3. Peto's method should not be employed when the three conditions needed are not met (event rates < 1%, balanced groups, small treatment effects). In such circumstances the Mantel-Haenszel odds-ratios without continuity correction performs better than Peto.
4. The beta-binomial model has been shown in simulations to outperform other methods in some settings<sup>7</sup>, but might suffer from issues regarding convergence of the model.
5. Bayesian meta-analysis with informative prior distributions is a good way to include random-effects in the meta-analysis. This is because it overcomes the problem of estimating heterogeneity when events are sparse.
6. Results regarding relative effects (odds/risk ratios) should always be presented along with absolute incidence rates, to put results into context<sup>27</sup>. A risk ratio of 1.5 might have very different clinical implications when the risk in the control group is 5% and when it is 0.1%.
7. Meta-analysts should avoid labelling results as statistically significant or non-significant. The use of p-values thresholds to dichotomize findings has recently attracted a lot of criticism<sup>28</sup>. Particularly for the case of rare-events meta-analysis, the use of arbitrary cut-points for p-values (such as the usual 0.05 or any other threshold<sup>29</sup>) can be even more problematic, because results might be affected by the choice of model – as was the case in the rosiglitazone example.

Finally, researchers should predefine an analysis plan a priori (e.g. at the protocol), to avoid selective use of methods. In addition, when events are rare, meta-analysts should always perform extensive sensitivity analyses using a range of alternative models, to ensure the robustness of their results. When results are very sensitive to the choice of model researchers should be particularly cautious on how they present and interpret their findings. In such cases, results should be considered exploratory and hypothesis-generating<sup>27</sup>.

**Funding:** This work has been supported by the Swiss National Science Foundation (grant title: «Enhancing methods for evaluating the comparative safety of medical interventions»)



## References

1. Higgins, J., Deeks, J. J. & Atman, D. G. Chapter 16: Special topics in statistics. In: Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* (updated March 2011). The Cochrane Collaboration, 2011. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org). (2011). Available at: Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org).
2. Vandermeer, B., Bialy, L., Hooton, N., Hartling, L., Klassen, T. P., Johnston, B. C. & Wiebe, N. Meta-analyses of safety data: a comparison of exact versus asymptotic methods. *Stat Methods Med Res* **18**, 421–432 (2009).
3. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Control Clin Trials* **7**, 177–188 (1986).
4. Mills, E. J., Thorlund, K. & Ioannidis, J. P. Demystifying trial networks and network meta-analysis. *BMJ* **346**, f2914 (2013).
5. Bradburn, M. J., Deeks, J. J., Berlin, J. A. & Russell Localio, A. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Stat Med* **26**, 53–77 (2007).
6. Keus, F., Wetterslev, J., Gluud, C., Gooszen, H. G. & van Laarhoven, C. J. H. M. Robustness assessments are needed to reduce bias in meta-analyses that include zero-event randomized trials. *Am. J. Gastroenterol.* **104**, 546–551 (2009).
7. Kuss, O. Statistical methods for meta-analyses including information from studies without any events—add nothing to nothing and succeed nevertheless. *Statist. Med.* **34**, 1097–1116 (2015).
8. Bhaumik, D. K., Amatya, A., Normand, S.-L., Greenhouse, J., Kaizar, E., Neelon, B. & Gibbons, R. D. Meta-Analysis of Rare Binary Adverse Event Data. *J Am Stat Assoc* **107**, 555–567 (2012).
9. Kishi, T., Matsunaga, S. & Iwata, N. Mortality Risk Associated With Long-acting Injectable Antipsychotics: A Systematic Review and Meta-analyses of Randomized Controlled Trials. *Schizophr Bull* **42**, 1438–1445 (2016).
10. Yusuf, S., Peto, R., Lewis, J., Collins, R. & Sleight, P. Beta blockade during and after myocardial infarction: An overview of the randomized trials. *Progress in Cardiovascular Diseases* **27**, 335–371 (1985).
11. Sweeting, M. J., Sutton, A. J. & Lambert, P. C. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Stat Med* **23**, 1351–1375 (2004).
12. Mantel, N. & Haenszel, W. Statistical Aspects of the Analysis of Data From Retrospective Studies of Disease. *J Natl Cancer Inst* **22**, 719–748 (1959).
13. Jaynes, E. T. *Bayesian Methods: General Background*. (1986).
14. Sutton, A. J. & Abrams, K. R. Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research* **10**, 277–303 (2001).
15. Smith, T. C., Spiegelhalter, D. J. & Thomas, A. Bayesian approaches to random-effects meta-analysis: a comparative study. *Stat Med* **14**, 2685–2699 (1995).

16. Stijnen, T., Hamza, T. H. & Ozdemir, P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Stat Med* **29**, 3046–3067 (2010).
17. Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R. & Jones, D. R. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statist. Med.* **24**, 2401–2428 (2005).
18. Turner, R. M., Davey, J., Clarke, M. J., Thompson, S. G. & Higgins, J. P. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *Int. J. Epidemiol.* **41**, 818–827 (2012).
19. Turner, R. M., Jackson, D., Wei, Y., Thompson, S. G. & Higgins, J. P. T. Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. *Statist. Med.* **34**, 984–998 (2015).
20. Chen, Y., Chu, H., Luo, S., Nie, L. & Chen, S. Bayesian analysis on meta-analysis of case-control studies accounting for within-study correlation. *Stat Methods Med Res* 962280211430889 (2011). doi:10.1177/0962280211430889
21. Rücker, G., Schwarzer, G., Carpenter, J. & Olkin, I. Why add anything to nothing? The arcsine difference as a measure of treatment effect in meta-analysis with zero cells. *Stat Med* **28**, 721–738 (2009).
22. Tian, L., Cai, T., Pfeffer, M. A., Piankov, N., Cremieux, P.-Y. & Wei, L. J. Exact and efficient inference procedure for meta-analysis and its application to the analysis of independent 2 x 2 tables with all available data but without artificial continuity correction. *Biostatistics* **10**, 275–281 (2009).
23. Cai, T., Parast, L. & Ryan, L. Meta-analysis for rare events. *Stat Med* **29**, 2078–2089 (2010).
24. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>, 2014).
25. Lunn, D., Spiegelhalter, D., Thomas, A. & Best, N. The BUGS project: Evolution, critique and future directions. *Statistics in Medicine* **28**, 3049–3067 (2009).
26. Nissen, S. E. & Wolski, K. Effect of Rosiglitazone on the Risk of Myocardial Infarction and Death from Cardiovascular Causes. *New England Journal of Medicine* **356**, 2457–2471 (2007).
27. Diamond, G. A., Bax, L. & Kaul, S. Uncertain effects of rosiglitazone on the risk for myocardial infarction and cardiovascular death. *Ann. Intern. Med.* **147**, 578–581 (2007).
28. Lane, P. W. Meta-analysis of incidence of rare events. *Stat Methods Med Res* **22**, 117–132 (2013).
29. Wasserstein, R. L. & Lazar, N. A. The ASA's Statement on p-Values: Context, Process, and Purpose. *The American Statistician* **70**, 129–133 (2016).
30. Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C. D., *et al.* Redefine statistical significance. *Nature Human Behaviour* 1 (2017). doi:10.1038/s41562-017-0189-z

**Table 1:** Data provided by study  $i$ , in the form of a 2x2 table

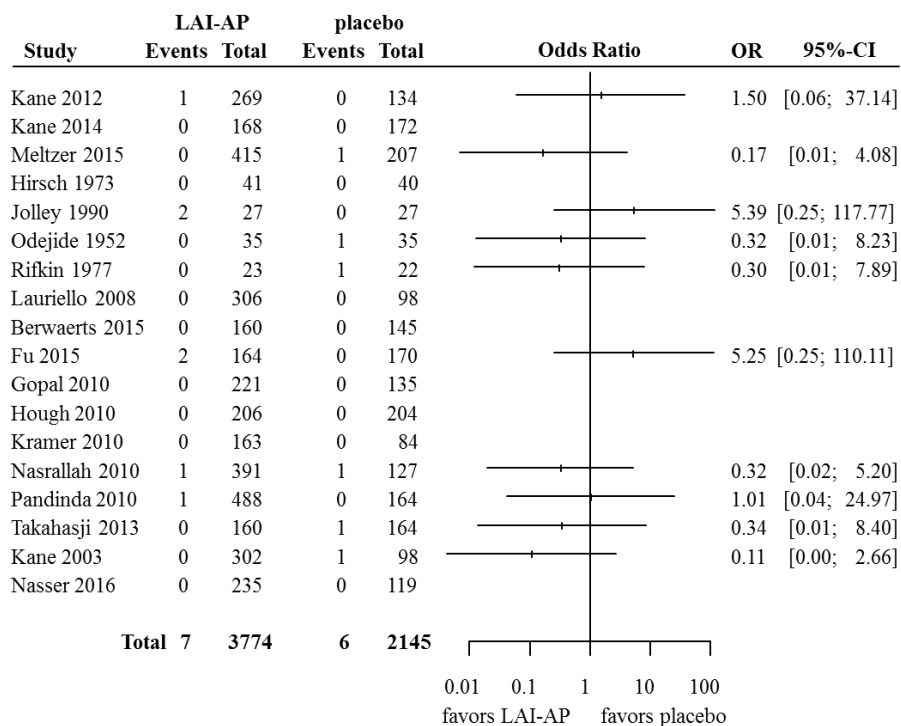
Study $i$	Events	No events
<b>Experimental</b>	$a_i$	$b_i$
<b>Control</b>	$c_i$	$d_i$

**Table 2:** Results from two sets of meta-analyses, using a range of different methods. An  $OR < 1$  or an arcsine difference  $< 0$  in the  $X$  vs  $Y$  comparison indicate that treatment  $X$  is safer (lower risk of death).

LAI-AP: long-acting injectable antipsychotics, OAP: oral antipsychotics, OR: odds ratios. CI: confidence interval. CrI: credible interval.

Method used	LAI-AP vs placebo	LAI-AP vs OAP
Inverse variance OR with 0.5 continuity correction (95% CI)	0.63 [0.24; 1.70]	0.71 [0.38; 1.34]
Inverse variance OR with treatment-arm continuity correction (95% CI)	0.70 [0.26; 1.90]	0.71 [0.38; 1.34]
Peto (95% CI)	0.75 [0.24; 2.36]	0.67 [0.35; 1.25]
Mantel-Haenszel OR with no continuity correction (95% CI)	0.77 [0.26; 2.28]	0.66 [0.34; 1.26]
Mantel-Haenszel OR with treatment-arm continuity correction (95% CI)	0.85 [0.36; 2.00]	0.71 [0.39; 1.28]
Bayesian meta-analysis, fixed effects OR (95% CrI)	0.72 [0.22, 2.27]	0.64 [0.33, 1.23]
Bayesian meta-analysis, random effects OR (95% CrI)	0.71 [0.22, 2.31]	0.63 [0.32, 1.24]
Beta-binomial model, OR (95% CI)	-	0.63 [0.29, 1.34]
Arcsine difference, fixed effects (95% CI)	-0.007 [-0.034; 0.021]	-0.019 [-0.041; 0.003]

**Figure 1:** Summary of the available studies for all-cause mortality, for long acting injectable antipsychotics (LAI-AP) vs. placebo. For illustration purposes, in order to calculate odds ratios (OR) and Confidence Intervals (CI) in single-zero studies we have used a 0.5 continuity correction. For double-zero studies no effect is shown.



**Figure 2:** Summary of the available studies for all-cause mortality, for long acting injectable antipsychotics (LAI-AP) vs. oral antipsychotics (OAP). For illustration purposes, in order to calculate odds ratios (OR) and Confidence Intervals (CI) in single-zero studies we have used a 0.5 continuity correction. For double-zero studies no effect is shown.

