NICE Guidelines Technical Support Unit

Meta-Analysis of Event Outcomes

Guideline Methodology Document 3

Version 1 (January 2021)

Caitlin Daly[1], Sumayya Anwer[2], Nicky J Welton[1], Sofia Dias[2], AE Ades[1]

[1] Population Health Sciences, Bristol Medical School, University of Bristol

[2] Centre for Reviews and Dissemination, University of York

About the NICE Guidelines Technical Support Unit

The NICE Guidelines Technical Support Unit (TSU) is a collaboration between the Universities of Bristol, Sheffield, York and Leicester. The TSU is commissioned by the Centre for Guidelines at the National Institute for Health and Clinical Excellence (NICE) to provide rapid-response technical support, methodology training, and methods research, in the context of guideline development. Please see this website for further information http://www.bristol.ac.uk/population-health-sciences/centres/cresyda/mpes/nice/

About the Guideline Methodology Document series

This series of Guideline Methodology Documents (GMDs) complements the Guide to the Methods of Technology Appraisal (1), the Guidelines Manual (2), and the NICE Decision Support Unit (DSU) Technical Support Documents (TSDs) (3-9).

The aim of the GMDs is to assist all those involved in guideline development, including guideline developers, guideline committee members, those commenting on draft guidelines during the consultation period, manufacturers, and stakeholders.

There is, of course, already a wealth of tutorial material on how to conduct systematic review and meta-analysis (10-12). The GMDs are in agreement with virtually all this material, although there are some significant differences in the way that meta-analytic methods are used.

The GMDs take the particular perspective of the guideline developer. They therefore go beyond standard treatments in which systematic review and meta-analysis tend to be seen as methods for producing "pooled" analyses that "summarise the literature". The decision context requires a focus on patients at specific points in their disease progression, methods that have particular properties regarding coherence and complete use of evidence, and procedures that are compatible with decision making under conditions of uncertainty.

The GMDs are aimed at a basic and introductory level: more advanced topics are indicated with an asterisk (*), and readers are referred elsewhere.

There are several areas of methodological uncertainty, controversy or rapid change. These are indicated in the GMDs. GMDs are extensively peer reviewed prior to publication (see acknowledgements). However, the responsibility for each GMD lies with the authors, who welcome any constructive feedback on the content, suggestions for updates and further guides. Readers should be aware that while the TSU is funded by NICE, these documents do not constitute formal NICE guidance or policy.

Acknowledgements

Contents

* indicates advanced material that may be skipped

1. INTRODUCTION: WHAT OUTCOMES ARE COVERED IN THIS GMD

This GMD makes recommendations on the synthesis of event outcomes. This includes trials that report:

- Proportions of patients who reach trial endpoints
- Numbers who reach endpoints in relation to person-time at risk
- Time-to-event outcomes with censoring

The GMD covers the standard contrast-based meta-analytic methods, using the generic inverse variance method with normal approximations as the default, and Mantel-Haenszel methods where there are zero cell counts. Reasons for synthesising rates or probabilities are discussed. The meta-analytic methods recommended here are in line with those in standard texts (10, 11), but is oriented to the decision making task of guideline developers. This requires consideration of how to deal with multiple outcomes and trials reported in multiple ways in the same meta-analysis.

The recommendations are set out for reference in Section 2. Section 3 gives an overview of the data extraction and meta-analysis process, drawing attention to decisions that guideline developers need to make before data extraction can begin.

Section 4 briefly looks at meta-analysis of trials with rare events and zero cells. Data extraction and processing procedures are set out in Section 5. Section 6 provides some worked examples of intermediate calculations that may be required, between data extraction and input into meta-analysis software. An accompanying *Data Conversion Workbook* (*GMD-3 Data Conversion Workbook.xlsx*) is available to assist in calculating the correct summary statistics for data input into Review Manager (RevMan) version 5.3, and the *metafor* package version 2.4-0 in R (version 3.6.3). The use of these packages is illustrated in a Software Appendix.

2. SUMMARY OF RECOMMENDATIONS

Recommendation 1. If event rates change over time, synthesis of log hazard ratios (LHRs) is recommended.

Recommendation 2. If event rates do not change over time, synthesis of log odds ratios (LORs) is preferred to log relative risks (LRRs), unless there is evidence that LRRs are less heterogeneous.

Recommendation 3. For trials with time-to-event data reporting hazard ratios, synthesis of log hazard ratios is recommended if the proportional hazards (PH) assumption is considered reasonable.

Recommendation 4. If there are no zero cells, generic inverse variance methods should be used, with the Paule-Mandel estimator being the recommended between-study variance estimator when employing a frequentist method for random effects models.

Recommendation 5. If there are zero cells, Mantel-Haenszel estimates should be used for fixed effect models; Bayesian methods are recommended for random effects models. Continuity corrections, where a constant is added to each cell of a 2 x 2 table, should be avoided if possible.

3. PRELIMINARY CONSIDERATIONS AND PLANNING

The process starts with a scoping review of the literature (See GMD1). The trials selected for synthesis may report outcomes in several different ways. Before data extraction begins, guideline developers will need to decide what kind of *model* will be needed to synthesise the trial evidence, in particular whether event *probabilities* or event *rates* will be modelled. Relative treatment effects can be measured in several ways: odds ratios (ORs) and relative risks (RRs) are often used for probability outcomes, and hazard ratios (HRs) for rate outcomes. However, if event rates change over time, HRs or other measures of relative effect may be required.

The scoping review should therefore answer the following questions:

- Do trials report probabilities of events that occur over time, and should follow-up time be taken into account?
- Do the trials report the number of patients who reach endpoints, or number of events in relation to time at risk; or are they time-to-event outcomes?
- Is the outcome binary, or are there multiple outcome categories? If multiple, are they mutually exclusive, or related in some other way?

The issues are discussed in more detail below.

3.1. ODDS RATIOS, RELATIVE RISKS AND HAZARD RATIOS: THE ROLE OF FOLLOW-UP TIME

Based, to some extent, on the nature of the available data, a key decision is whether the treatment is considered to alter the *probability* that the outcome occurs, or the (hazard) *rate* at which the outcome occurs. In the former case the synthesis will involve (log) odds ratios, or (log) relative risks. If treatment is considered to affect the hazard, then it will be the (log) hazard ratios that are synthesised.

Underlying the question of rates versus probabilities is whether length of follow-up impacts on the probability of the outcome. Some outcomes either occur or do not occur (a baby is either low birthweight or not). In these cases, guideline developers will need to decide between ORs and RRs as the measure of treatment effect (see below).

However, most events occur over the follow-up period, and more events are expected with longer follow-up. An event rate analysis is then the natural choice.

Recommendation 1. If event rates change over time, synthesis of log hazard ratios (LHRs) is recommended.

3.1.1. Identifying a minimum follow-up time

Many outcomes tend to occur more rapidly immediately after the trial observation period begins, and then the rate diminishes over time. This may be due to a "depletion of susceptibles" phenomenon. Almost invariably, individuals differ in their risk of reaching the endpoint, so that high risk patients will reach the endpoint first, and the hazard rate will *inevitably* decrease until only the very low risk patients remain. Guideline developers may be able to identify a minimum follow-up time by which virtually all patients who are going to reach the endpoint will have done so. Further follow-up beyond this point will not contribute materially to the data. In this situation, a probability analysis, rather than a rate analysis, is possible as long as all the trials report outcomes at some point *after* that minimum. Alternatively, it may be decided to restrict the analysis to a range of follow-up times within which the ORs or RRs remain approximately constant.

3.1.2.* When is a hazard rate analysis necessary?

If there is no restricted range of follow-up times where ORs and RRs remain roughly constant, *nor* a minimum time after which few further events occur, then an event rate analysis may be necessary, especially if formal cost-effectiveness analysis (CEA) or decision modelling is to take place. Modelling event rates is slightly more complicated as time has to be taken into account, but there may be no alternative as standard methods based on ORs and RRs can perform very poorly when events are not rare, as we describe below.

If event rates in all treatment arms are less than 10% (over the relevant time period) (13), then the relative treatment effects measured as an HR, an OR, or an RR will all be very close (as are the LHR, log odds ratio (LOR), and log relative risk (LRR)).

If event rates are not rare much more care needs to be taken. If events occur at a constant hazard rate in each arm, then the HR remains constant over time. (The HR can also remain constant when the arm rates change). But, in these circumstances, both the OR and the RR will depend on the follow-up time, and we should refer to $OR(t)$ and $RR(t)$, meaning the OR and RR at follow-up time t. For example, just after the trial onset, when follow-up time is barely greater than zero and events are "rare", $RR(t) = OR(t) = HR$, but as follow-up time is extended, the behaviour of $OR(t)$ and $RR(t)$ diverges depending on how the HR is expressed. If $HR > 1$, then $RR(t)$ *decreases* from HR to 1 as t increases, while $OR(t)$ *increases* from HR to infinity. But if $HR < 1$ then *both* $OR(t)$ and $RR(t)$ *decrease* towards zero as t increases. This is why OR and RR analyses must be abandoned when event rates change over time.

3.2. ODDS RATIOS *OR* RELATIVE RISKS

When a probability model, rather than a rate model, is required, there still remains a choice between OR and RR analyses. This is sometimes viewed as being a matter of which is easiest to interpret (11, 14). However, this is incorrect: unless the endpoint is a relatively rare event (< 5%), the OR and RR analyses represent different models of the data, which lead to different predictions, and may even lead to different recommendations.

Another consideration is that ORs are "symmetrical", in the sense that the odds ratio for survival is the reciprocal of the odds ratio for mortality. A RR for survival has no such simple relation to the RR for mortality: these represent two very different models of the data.

Ideally, the choice of OR or RR should be based on which one is the best model of the data. Investigations of large numbers of meta-analyses have failed to find marked overall differences between heterogeneity in ORs and RRs (14, 15), although there may be some disease areas where RRs better describe the impact of treatment, and others where ORs work better. Risk Differences are another possibility where treatment effects are additive, rather than multiplicative, however these are rarely used partly due to difficulties in constraining risks to lie between 0 and 1. As well as an analysis of the heterogeneity, a plot of the treatment effect against the control arm event probability can help highlight which measure is most stable in specific situations. However, because of the ambiguity in "which way round" the RR is taken, unless there are good reasons to the contrary, we recommend OR analysis as the default option.

Recommendation 2. If event rates do not change over time, synthesis of log odds ratios (LORs) is preferred to log relative risks (LRRs), unless there is evidence that LRRs are less heterogeneous.

Definitions of ORs, RRs and how to derive them from 2x2 tables are given in Section 6, Example 1.

3.2.1. *Against "synthetic" relative risks*

In the interests of transparency, it is important that the measure of relative treatment effect that is *reported* (OR or RR) is the relative effect assumed in the meta-analysis. Some investigators carry out a meta-analysis of the LORs, then, once a baseline event probability has been identified, the event probabilities in the treatment and control arms are used to calculate an RR, which is then presented in the report. This is unnecessary and confusing. In addition, the credible intervals of the "synthetic" RR cannot be calculated correctly from the outputs of standard software.

3.3. TIME TO EVENT OUTCOMES

Analysis of event rates can take several forms: standard 2x2 tables of probability outcomes can be analysed with hazard rate models using complementary log-log (cloglog) models (Section 3.3.1); hazard rates are also used for number of events in relation to time-at-risk (Section 3.3.2); and time-to-event analyses, sometimes called survival analysis (Section 3.3.3).

3.3.1. *Hazard rate analysis of probability outcomes*

If hazard rates change over time, 2x2 tables of event probability outcomes can be analysed in terms of hazard rates, using the cloglog transformation. This form of analysis assumes that the *ratio* of hazards remains the same – the *proportional hazards* (PH) assumption. Calculation of the LHR and its standard error for input into standard meta-analysis software is described in Section 6, Example 3.

3.3.2. *Hazard rate analysis of event frequencies in relation to time at risk*

This method is for trials that report the number of patients reaching the endpoint, and the total time-at-risk. The method is described in Section 6, Example 4. It can also be used in situations where the same patient can experience the endpoint event more than once, for example, number of strokes, or numbers of teeth with caries.

3.3.3. *Time-to-event analysis with censored observations ("Survival analysis")*

Time-at-risk is a way of taking account of "censoring". This is the name given to the situation where some patients do not reach the endpoint, whether because they were lost to follow-up, or a different outcome intervened, or they reached the end of the trial observation period before any endpoint occurred. The most common way of reporting the relative treatment effect from a survival analysis is the HR or LHR, usually estimated by a Cox regression. LHRs can be synthesised using the generic inverse-variance meta-analytic method. Note, however, that this form of analysis assumes that the PH assumption holds for every trial.

Recommendation 3. For trials with time-to-event data reporting hazard ratios, synthesis of log hazard ratios is recommended, if the proportional hazards (PH) assumption is considered reasonable.

3.3.4.* Time-to-event outcomes without proportional hazards

Proportional hazards is a strong assumption, especially in the context of cost-effectiveness analyses taking a life-time perspective. Survival curves crossing or converging indicate that hazards are not proportional. If the PH assumption is not accepted it may be necessary to scan in Kaplan Meier survival

curves in order to reconstruct original data. An R package (16) can be used for this, and similar software has been developed in STATA (17).

Once the survival data is extracted, a number of options are available to model and synthesise the treatment effect (18), including fractional polynomials (18, 19), separate meta-analysis of shape and scale parameters (20), models based on area under the curve (21), and models to extrapolate baseline survival beyond the trial data (22-24).

Changes in the relative treatment effect over time should be considered alongside the construction of a model for the baseline, as it is almost certain that the baseline risk or hazard will be changing over time as well (See GMD1).

3.4. * MULTIPLE EVENT OUTCOMES – COMPETING RISKS, ORDERED CATEGORIES

Multiple event outcomes occur in many different forms and synthesis of outcomes both within and between trials needs to be considered. In this GMD we consider only within-trial synthesis. Examples of between-trials synthesis are set out in GMD1.

3.4.1. Conditioning one outcome on another

A very common example of multiple outcomes arises with treatment discontinuation, which can be considered as a competing outcome alongside, for example, response to treatment, or relapse. In this case it is possible to take discontinuation vs no discontinuation as the first binary outcome, and then, among those who have not discontinued, relapse vs no relapse is a second binary outcome, *conditional on the first.* This means that two competing outcomes with a single denominator can be represented as two binary outcomes, leading to two separate and independent conventional meta-analyses (see Example 5 in Section 6).

As well as providing a solution to the problem of synthesising evidence on correlated outcomes, the conditioning approach is generally required in decision modelling, as the probability of discontinuation and the probability of response among those who continue must both be modelled. An intention-to-treat (ITT) approach loses this distinction and is thus less helpful to decision makers, while per protocol (PP) analyses normally do not consider those who discontinue. As such, the conditioning approach provides a better alternative to both the ITT and PP approaches, as it describes what would happen in practise, which is more helpful for formulating economic models.

Outcomes treated in this way will often be event outcomes which occur in time. Either, or both, outcomes may be analysed with ORs, RRs, or HRs. In the particular case of discontinuation, it may be possible to assume that discontinuation occurs relatively early, and can be considered as a probability outcome leading to an OR meta-analysis for discontinuation. The separate analysis of remission or relapse would most likely be a rate analysis.

*3.4.2. * Competing risks*

Where there are more than two competing categories, competing risks meta-analysis may be better than a conditioning solution (see schizophrenia example using WinBUGS (4)). This has been used in NICE guidelines on treatments for patients in remission from schizophrenia, where the outcomes were discontinuation for side effects, discontinuation for other reasons, and relapse (25, 26).

*3.4.3. * Ordered categories*

Another type of multiple outcome occurs with ordered categories. Common examples are the PASI (Psoriasis Area Severity Index) scores: trials may then report the PASI50, PASI75 or PASI90, the proportion of patients whose improvement relative to baseline is above 50%, 75% and 90%. The America College of Rheumatology scores (ACR) are reported in the same way. Although it is possible to treat each category as a separate binary outcome, a more robust analysis is available if all these are analysed simultaneously, based on the assumption that the treatment effect is the same at each "cut-point". Worked examples can be found in TSD2 (4). Numerous examples of these methods can be found in NICE appraisals for psoriasis, psoriatic and rheumatoid arthritis, and ankylosing spondylitis (27-30).

Note that, with both ordered categories and competing risks, separate analyses of each outcome without appropriate conditioning represents an incorrect model of the data, which ignores the negative correlations between the probabilities of being in each category. If results are to be embedded in CEAs or other decision models, these correlations must be propagated through the decision model, otherwise incorrect results will be obtained.

4. SPARSE DATA AND ZERO CELLS

The standard meta-analytic methods for frequency counts take relative effect measures (LORs, LRRs, LHRs) and their standard errors as the data, and rely on the assumptions that these have an approximately normal distribution. This assumption starts to break down when there are cell counts less than 3. When there are zero cells, estimates of treatment effect cannot be calculated.

Recommendation 4. If there are no zero cells, generic inverse variance methods should be used, with the Paule-Mandel estimator being the recommended between-study variance estimator when employing a frequentist method for random effects models.

There are several standard workarounds for zero cells counts:

1. Add 0.5, or some other small number, to every cell in the table (called a "continuity correction") and continue with the generic (inverse variance weighted) method, using fixed or random effects models
2. Mantel-Haenszel methods (31) (fixed effect only)
3. Peto method (32) (fixed effect only)
4. Add 0.5 to every cell, and use Mantel-Haenszel methods (fixed effect only)
5. *Use Bayesian methods (4), using fixed or random effects models

Options 1 and 4 produces ORs, RRs, and HRs that are biased towards 1, i.e., towards a null effect. Option 3 also biases towards the null, particularly if the relative effect is large (33). Option 4 is the default option in RevMan, STATA and *metafor* in R, although the use of a continuity correction with Mantel-Haenszel methods is both unnecessary and incorrect (34).

Unless trials with zero cell counts make up a relatively small proportion of trials, we therefore recommend option 2, or if a random effects model is indicated, option 5. Mantel-Haenszel methods are available for ORs, RRs and HRs based on events and time-at-risk. Worked examples are shown in the Software Appendix C. In the Cochrane review (35) which is used as an example, the bias in options 1, 3, and 4 is substantial, underestimating the treatment effect by almost a factor of two.

Recommendation 5. If there are zero cells, Mantel-Haenszel estimates should be used for fixed effect models; Bayesian methods are recommended for random effects models. Continuity corrections, where a constant is added to each cell of a 2 x 2 table, should be avoided if possible.

5. DATA EXTRACTION AND PROCESSING

This section sets out preferred procedures for data extraction for the cases where standard meta-analytic approaches are available, depending on whether a probability or rate model is required.

Probability Model: Pick one of the following options, choosing the earliest one possible on the list
1. 2 x 2 table, with the numerator being the number reaching the end point in each arm, and the denominators being the total number of patients.
2. LOR, LRR, and their standard errors, or confidence intervals (Appendix A)
3. OR, RR and their 95% confidence intervals (Appendix A)
4. Estimated event probability at required follow-up time in each arm, and its standard error, based on a survival analysis[1]
5. Extract required statistics from Kaplan Meier curve (Section 3.3.4)[1]

EVENT DATA: extraction for a Probability Model

---
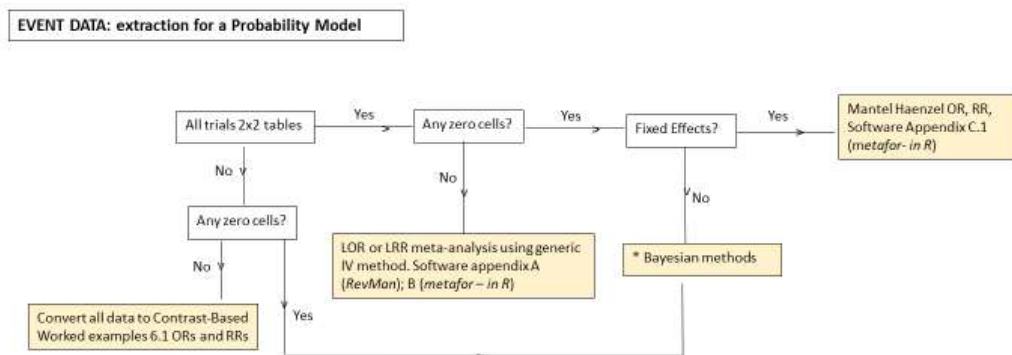
[1] If survival analysis has been performed, a rate model should be considered

Rate Model: Pick one of the following options, choosing the earliest one on the list

1. LHR and its standard error, or confidence interval (Appendix A)
2. HR and its 95% confidence intervals (Appendix A)
3. Number of events and total time at risk in each arm
4. 2 x 2 table, with the numerator being the number reaching the end point in each arm, the total number of patients, and time-at-risk. If total time at risk is not reported, use the average follow-up time multiplied by the total number of patients.

6. INTERMEDIATE CALCULATIONS AND WORKED EXAMPLES

In this section we give 5 worked examples:

1. Converting 2 x 2 tables to LOR and SE(LOR) or LRR and SE(LRR) (no zero cells)
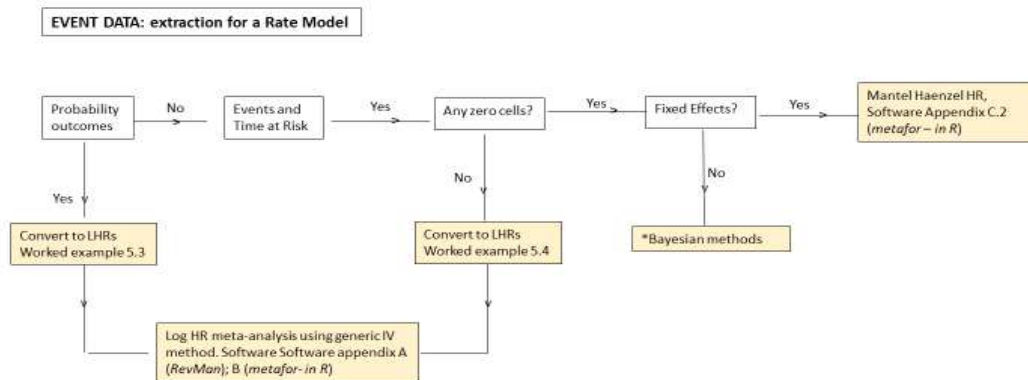2. Extraction of probabilities of an event from survival data, and its SE, and use in LOR and LRR analysis
3. Converting 2 x 2 tables with follow-up time to LHR and SE(LHR) (no zero cells)
4. Converting Poisson outcomes with person-years at risk to LHR and SE(LHR)
5. Converting two competing outcomes into two binary outcomes, one conditional on the other

The first four calculations can be carried out by the *GMD3 Data Conversion Workbook* (see Appendix B).

Example 1: Converting 2 x 2 tables to LORs or LRRs and their SEs

The results of each arm-based trial can be presented in the form of a 2 x 2 contingency table (Table 6.1).

Table 6.1. *General 2 x 2 contingency table*

	Patients reaching end-point	Patients not reaching end-point	Total
Treatment 1	a	b	$n_1(=a+b)$
Treatment 2	c	d	$n_2(=c+d)$

In Table 6.1, the number of individuals who experience an event in treatment groups 1 and 2 are represented by a and c, respectively. Similarly, the number of individuals who do not experience the event in treatment groups 1 and 2 are denoted by b and d, respectively.

Odds Ratio:

The OR comparing treatment 2 to treatment 1 (Table 6.1) for each trial is:

$$OR = \frac{bc}{ad}$$

where $\frac{a}{b}$ and $\frac{c}{d}$ are the odds of an event in an individual with treatments 1 and 2 respectively.

The standard error of the LOR is: $SE(LOR) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$

Relative Risk (Risk Ratio):

The RR of treatment 2 compared to treatment 1 (Table 6.1) for each trial is:

$$RR = \frac{cn_1}{an_2}$$

where $\frac{a}{n_1}$ and $\frac{c}{n_2}$ are the risks of an event with treatments 1 and 2 respectively.

The standard error of the LRR is: $SE(LRR) = \sqrt{\frac{1}{a} + \frac{1}{c} - \frac{1}{n_1} - \frac{1}{n_2}}$

Worked example

Table 6.2. *Arm-based event data reported in Blum 1998 (36)*

	Dyspeptic	Non-Dyspeptic
Control	130	34
Treatment	119	45

Table 6.2 shows arm-based dichotomous data for a single trial (36) which was included in a systematic review on the effect of eradicating *Helicobacter pylori* on non-ulcer dyspepsia (14). The outcome of interest is the number of patients who remain dyspeptic (i.e. the lack of success of treating non-ulcer dyspepsia).

The data are converted to the equivalent contrast-based measurement of LOR as follows:

$$OR = \frac{119 \times 34}{45 \times 130} = 0.692, \; LOR = \ln(0.692) = -0.369, \; SE(LOR) = \sqrt{\frac{1}{119} + \frac{1}{45} + \frac{1}{130} + \frac{1}{34}} = 0.260$$

The LRR is calculated as follows:

$$RR = \frac{119(130+34)}{130(119+45)} = 0.915, \; LRR = \ln(0.915) = -0.0884, \; SE(LRR) = \sqrt{\frac{1}{119} + \frac{1}{130} - \frac{1}{164} - \frac{1}{164}} = 0.062$$

Example 2: Calculating ORs and RRs from probabilities derived from Kaplan Meier curves

This example shows how data from a survival analysis can be used when it has been decided to carry out an analysis of ORs or RRs at a specified follow-up time. It assumes that the Kaplan Meier curves have been published, and that the methods in Section 3.3.4 have been used to reconstruct the KM data. Next, standard survival time software is used to calculate the event probabilities at the specified time t for each arm, $S(t,1)$ and $S(t,2)$, along with their standard errors, $se(S(t,1))$ and $se(S(t,2))$, respectively.

The LOR at time t comparing the log-odds of treatment 2 to treatment 1 is defined as:

$$LOR(t) = \ln\left(\frac{S(t,2)}{1-S(t,2)} \right) - \ln\left(\frac{S(t,1)}{1-S(t,1)} \right)$$

The standard error of the LOR is:

$$SE(LOR(t)) = \sqrt{ \left(\frac{se(S(t,1))}{S(t,1)(1-S(t,1))} \right)^2 + \left(\frac{se(S(t,2))}{S(t,2)(1-S(t,2))} \right)^2 }$$

The LRR at time *t* comparing the log-risks of treatment 2 to treatment 1 is defined as:

$$LRR(t) = \ln(S(t,2)) - \ln(S(t,1))$$

The standard error of the LRR is:

$$SE(LRR(t)) = \sqrt{ \left(\frac{se(S(t,1))}{S(t,1)} \right)^2 + \left(\frac{se(S(t,2))}{S(t,2)} \right)^2 }$$

Worked example

Table 6.3. *Event probabilities estimated from reconstructed Kaplan-Meier curve data of Albain 2009 (37)*

Study	Chemoradiotherapy		Chemotherapy + Surgery	
	Probability of Survival at 5 Years $S(5,1)$	Standard Error of Probability of Survival at 5 years $se(S(5,1))$	Probability of Survival at 5 Years $S(5,2)$	Standard Error of Probability of Survival at 5 years $se(S(5,2))$
Albain 2009	0.188	0.041	0.258	0.037

Table 6.3 shows the arm-based data for a single trial (37) which was included in a NICE guideline on non-small cell lung cancer (21). This trial presented Kaplan-Meier curves for overall survival, from which the synthetic individual patient data were extracted. Based on this synthetic data, the probability of surviving up to 5 years, along with its standard error, were estimated for each treatment using the survfit function in the *survival* package in R (38).

The data are converted to the equivalent contrast-based measurement of LOR as follows:

$$LOR = \ln\left(\frac{0.258}{1-0.258}\right) - \ln\left(\frac{0.188}{1-0.188}\right) = 0.4067,$$

$$SE(LOR) = \sqrt{\left(\frac{0.041}{0.188(1-0.188)}\right)^2 + \left(\frac{0.037}{0.258(1-0.258)}\right)^2} = 0.3309$$

The LRR is calculated as follows:

$$LRR = \ln(0.258) - \ln(0.188) = 0.3165,$$

$$SE(LRR) = \sqrt{\left(\frac{0.041}{0.188}\right)^2 + \left(\frac{0.037}{0.258}\right)^2} = 0.2610.$$

Example 3: Converting 2 x 2 tables with event data and time at risk to log hazard ratios

Table 6.4. *General data required from 2 x 2 contingency table*

	Patients reaching endpoint	Total
Treatment 1	r_1	n_1
Treatment 2	r_2	n_2

For a study reporting over a follow-up time f, the number of events observed in treatment groups 1 and 2 (with sample sizes n_1 and n_2) are denoted by r_1 and r_2, respectively (Table 6.4).

The HR comparing the hazards of treatment 2 to treatment 1 is defined as:

$$HR = \frac{\ln(1-p_2)}{\ln(1-p_1)},$$

where p_1 and p_2 are the probabilities that an event occurs in treatment arms 1 and 2 respectively,

$$\text{expressed } p_1 = \frac{r_1}{n_1} \text{ and } p_2 = \frac{r_2}{n_2}.$$

The standard error of the LHR is:

$$SE(LHR) = \sqrt{\frac{p_1}{n_1(1-p_1)[\ln(1-p_1)]^2} + \frac{p_2}{n_2(1-p_2)[\ln(1-p_2)]^2}}$$

Worked example

Table 6.5. *Arm-based event data reported in AASK (39)*

Study	Follow-up Time (years)	β-Blockers		CCB	
		New Cases	Total	New Cases	Total
AASK	3.8	70	405	32	202

Table 6.5 shows the data for a single trial from a network meta-analysis studying the incidence of diabetes in randomised controlled trials of antihypertensive drugs (39). The outcome of interest is

new cases of diabetes observed over the study follow-up period (measured in years) for two different treatments, β-Blockers and CCB.

The LHR and the corresponding standard error comparing β-Blockers to CCB are calculated as follows for $p_1 = \dfrac{32}{202} = 0.158$ and $p_2 = \dfrac{70}{405} = 0.173$:

$$HR = \frac{\ln(1-p_2)}{\ln(1-p_1)} = \frac{\ln(1-0.173)}{\ln(1-0.158)} = \frac{\ln(0.827)}{\ln(0.842)} = 1.100, \quad LHR = \ln(1.100) = 0.0955$$

$$SE(LHR) = \sqrt{\frac{0.158}{405(1-0.158)[\ln(1-0.158)]^2} + \frac{0.173}{202(1-0.173)[\ln(1-0.173)]^2}}$$

$$= \sqrt{\frac{0.158}{405(0.842)[\ln(0.842)]^2} + \frac{0.173}{202(0.827)[\ln(0.827)]^2}}$$

$$= 0.2137$$

Note that the follow-up time is not required in the computation of the LHR and its standard error.

Example 4: Converting Poisson outcomes with person-years at risk to LHR and SE(LHR)

Rate data for Poisson outcomes can be reported as the number of events observed per person-years.

Table 6.6. *General data format for trials reporting Poisson outcomes*

	Events	Person-years at risk
Treatment 1	E_1	T_1
Treatment 2	E_2	T_2

The events in treatment arms 1 and 2 are denoted by E_1 and E_2 respectively; the respective person-years observed are denoted by T_1 and T_2 (Table 6.6).

The HR comparing treatment 2 to treatment 1, and the standard error of the LHR are:

$$HR = \frac{E_2 T_1}{E_1 T_2}, \qquad SE(LHR) = \sqrt{\frac{1}{E_1} + \frac{1}{E_2}}$$

Worked example

Table 6.7. *Arm-based Poisson data reported in DART (40)*

Study	Treatment		Control	
	Number of deaths	**Person-years at risk**	**Number of deaths**	**Person-years at risk**
DART	111	1925	113	1917

Table 6.7 shows the rate data extracted from a trial that was included in a Cochrane Review (40) studying the effect of dietary fats on total mortality. Reduced fat diet interventions were compared to a control and the total mortality and the number of events observed per person years was reported.

The rate data are converted to the contrast-based LHR comparing treatment to the control as follows:

$$HR = \frac{E_T T_C}{E_C T_T} = \frac{111 \times 1917}{113 \times 1925} = 0.978, \quad LHR = \ln(0.978) = -0.022$$

$$SE\left(LHR\right) = \sqrt{\frac{1}{111} + \frac{1}{113}} = 0.134$$

Example 5: Converting data on two competing outcomes into two binary outcomes

An example of two competing outcomes might be: outcome 1 = treatment discontinuation, outcome 2 = response. In the new format there would be two independent binary outcomes: treatment discontinuation, and response conditional on no treatment discontinuation. This requires the calculation of the denominator of the outcome 2, which is the denominator of outcome 1 minus the numerator of outcome 1 (Table 6.8).

Worked example

Table 6.8. *Fictitious data illustrating competing outcomes*

	Original data: two competing outcomes			New format: two outcomes, one conditional on the other			
	Outcome 1	Outcome 2	Denominator 1	Outcome 1	Denominator 1	Outcome 2 (conditional on not outcome 1)	Denominator 2
Study	A	B	C	A	C	B	C-A
Trial 1	10	12	100	10	100	12	90
Trial 2	15	20	200	15	200	20	185

Indirect estimation of the standard error of the log-ratios (e.g., odds ratios, relative risks, hazard ratios)

Methods for deriving the standard error of a relative effect (e.g., θ = LOR, LRR, LHR), SE, from available statistics, when it is not reported directly.

Finding SE from 95% confidence interval on a LHR, LOR, LRR	$SE = \dfrac{\text{upper limit - lower limit}}{3.92}$
Finding SE from LHR, LOR, or LRR, θ, and *z*-statistic	$SE = \dfrac{|\theta|}{z}$
Finding SE from LHR, LOR, or LRR, θ, and *t*-statistic	$SE = \dfrac{|\theta|}{t}$
Finding SE from 95% CI on a HR, OR or RR	$SE = \dfrac{\ln\left(\text{upper limit}\right) - \ln\left(\text{lower limit}\right)}{3.92}$
Finding *z*-statistic from one sided *p*-value, p, corresponding to z-test	$z = \Phi^{-1}\left(1 - p\right)$, where Φ^{-1} is the inverse of the standard normal cumulative distribution function
Finding *z*-statistic from two-sided *p*-value, p, corresponding to z-test	$z = \Phi^{-1}\left(1 - \dfrac{p}{2}\right)$
Finding *t*-statistic from one sided *p*-value, p, corresponding to t-test	$t = \pm t^{-1}\left(p, df\right), df = n_1 + n_2 - 2$, where t^{-1} is the inverse of the t distribution, n_1 and n_2 are the number of patients in arm 1 and arm 2, respectively.
Finding *t*-statistic from two-sided *p*-value, p, corresponding to t-test	$t = \pm t^{-1}\left(\dfrac{p}{2}, df\right), df = n_1 + n_2 - 2$

Alternatively, the RevMan calculator may assist in these calculations. An Excel File containing the RevMan calculator may be obtained from:

http://training.cochrane.org/resource/revman-calculator

General guidance on GMD3 Data Conversion Workbook

This workbook should allow the user to input one of the prioritised sets of statistics listed in Section 5 for each trial. To keep track of data conversions, we suggest saving a workbook for each guideline. Data may be stored for multiple trials within the worksheets for this purpose. The data should be outputted in a format required for synthesis, on the contrast-level.

Table B.1. *Reference to conversions in GMD3 Data Conversion Workbook*

Procedure	Worksheet	Data Input		Data Output	
		Statistics	Column(s)	Statistics	Column(s)
Calculate LOR or LRR based on 2x2 data	LOR and LRR	Number of patients reaching endpoint	B; F	log(OR) and SE	X-Y
		Number of patients randomised to treatment group	C; G	logRR and SE	Z-AA
Calculate LOR or LRR based on event probabilities and SE	LOR and LRR	Event probability at required follow-up time	D; H	log(OR) and SE	X-Y
		Standard error of event probability	E, I	log(RR) and SE	Z-AA
Calculate LHR based on event and person-time at risk data	LHR	Number of events	B; F	log(HR) and SE	T-U
		Person-time at risk	C; G		
Calculate LHR based on 2x2 data	LHR	Number of patients reaching endpoint	D; H	log(HR) and SE	T-U
		Number of patients randomised to treatment group	E; I		
Calculate log-odds based on partial 2 x 2 data	log odds calculation	Number of patients reaching endpoint	B	logodds	G
		Number of patients randomised to treatment group	C	V(log odds)	H
				SE(log odds)	I

SOFTWARE APPENDIX

A: Odds ratios, relative risks, hazard ratios in Review Manager (RevMan) 5.3

Arm-based data are available for synthesis of OR and RR, but not HR data. HR meta-analysis must be performed using the contrast-based input of LHRs and their SEs.

A fixed effect (FE) analysis is conducted in this example. For random effects (RE) models, the only method offered by RevMan is the Der-Simonian and Laird (41).

Arm-based data

1. Right click '*Data and Analyses*'
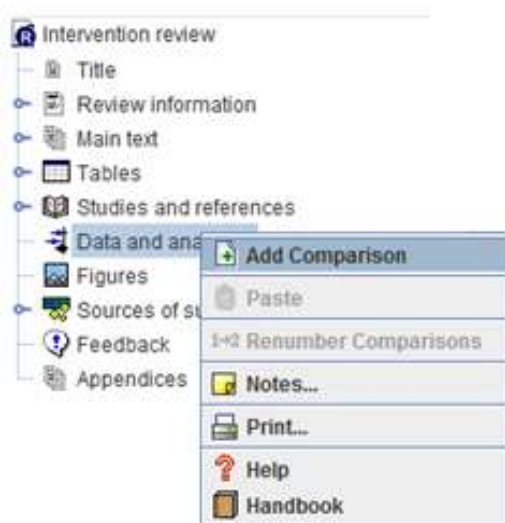
2. Select *'Add Comparison'*

3. The *New Comparison Wizard* pops up:

- In *'What name should the comparison have?*' Enter the treatments that are being compared in the meta-analysis and click *'Next>'*
- In *'What do you want to do after the wizard is closed?'* select *'Add an outcome under the new comparison'* and click *'Continue'*

4. The *New Outcome Wizard* pops up:
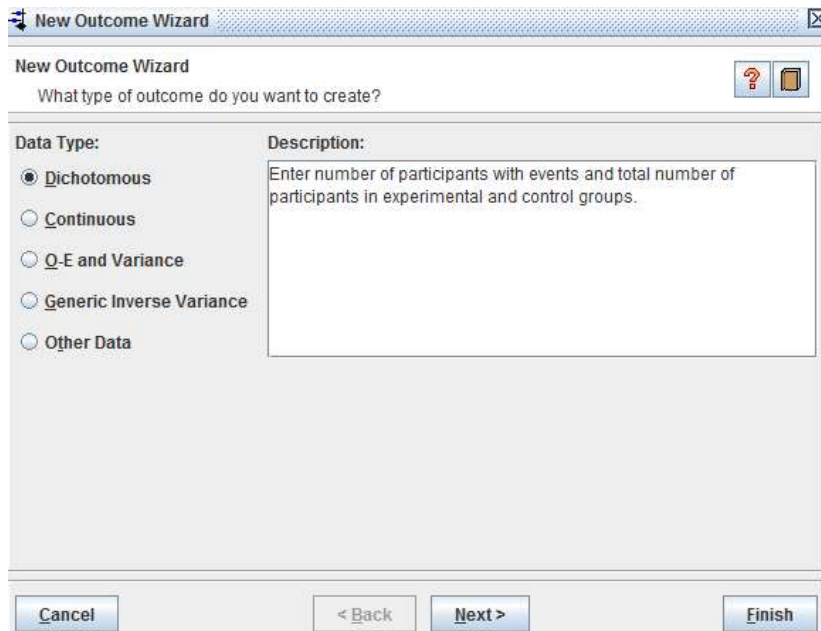
- Select 'Dichotomous' and click *'Next>'*
- For *'Name'*, enter the outcome being studied (In this example, 'Number of patients still dyspeptic') and treatment names (For this example, Group 1: Eradicating *H.pylori* and Group 2: Control). Click *'Next>'*
- The next prompt for the New Outcome Wizard allows you to specify either an Odds ratio analysis, or a Relative Risk (Risk Ratio) analysis. Select and then click *'Next>'*

- Select the desired options from those given below for *'Which analysis details do you want to use?'* and click on *'Next>'*.

- Under 'Which graph details do you want to use?', enter the graph labels and click *'Next>'*.
  (This depends on whether the outcome is desirable or not. For this example:
  Left Graph Label: Favours Treatment,
  Right Graph Label: Favours Control)
- For the final prompt 'What do you want to do after the wizard is closed?', select 'Add study data for the new outcome' and click 'Continue'.
- Select the studies required under 'Included Studies' and click *'Finish'*

5. Copy the data from an Excel file and paste it in the appropriate column in RevMan

Comparison: 1 Eradication of H.pylori, Outcome: 1.1 Number of patients still dyspeptic

Study or Subgroup	Eradicating Events	Total	Control Events	Total	Weight	Odds Ratio M-H, Fixed, 95% CI
Blum (OCAY) 1998	119	164	130	164	24.8%	0.69 [0.42, 1.15]
Koelz 1998	67	89	73	92	12.3%	0.79 [0.39, 1.59]
McColl 1998	121	154	143	154	21.3%	0.28 [0.14, 0.58]
Talley (ORCHID) 1999	101	133	111	142	18.0%	0.88 [0.50, 1.55]
Talley (USA) 1999	81	150	72	143	23.6%	1.16 [0.73, 1.83]
Total (95% CI)		690		695	100.0%	0.76 [0.59, 0.97]
Total events	489		529			
Heterogeneity: Chi² = 10.84, df = 4 (P = 0.03...						
Test for overall effect: Z = 2.16 (P = 0.03)						

Contrast-based data

RevMan carries out exactly the same "generic inverse variance" analysis on LORs, LRRs, and LHRs as it does on contrast-based continuous data. The only difference between these analyses is specifying the chosen effect measure in the relevant window (See Step 4 for more details). RevMan uses this information to correctly label the outputs: the underlying analysis is identical in every case.

1. Right click '*Data and Analyses*'

2. Select *'Add Comparison'*

3. The *New Comparison Wizard* pops up:

- In *'What name should the comparison have?'* Enter the treatments that are being compared in the meta-analysis and click *'Next>'*
- In *'What do you want to do after the wizard is closed?'* select *'Add an outcome under the new comparison'* and click *'Continue'*

4. The *New Outcome Wizard* pops up:

- Select the appropriate Data Type ('Generic Inverse Variance' in this example where we input the log-OR and SE for each included trial) and click 'Next>'
- For the second prompt, 'Name', enter the outcome being studied (In this example, 'Number of patients still dyspeptic') and treatment names (For this example, Group 1: Eradicating *H.pylori* and Group 2: Control). Click 'Next>'
- For the third prompt, "*Which analysis method do you want to use*", specify the analysis methods. This is also where the effect measure can be selected. Select "Odds Ratio" for a LogOR analysis, or "Risk Ratio" for a LogRR analysis. For a LogHR analysis, select the "Name of Effect Measure" and select "Hazard Ratio" from the drop-down box. Click 'Next>'.

- Select the desired options from those given below for 'Which analysis details do you want to use?' and click on 'Next>'.

- Under 'Which graph details do you want to use?', enter the graph labels and click 'Next>'. (This depends on whether the outcome is desirable or not. For this example:

Left Graph Label: Favours Treatment,
Right Graph Label: Favours Control)

- For the final prompt 'What do you want to do after the wizard is closed?', select 'Add study data for the new outcome' and click 'Continue'.
- Select the studies required under 'Included Studies' and click *'Finish'*

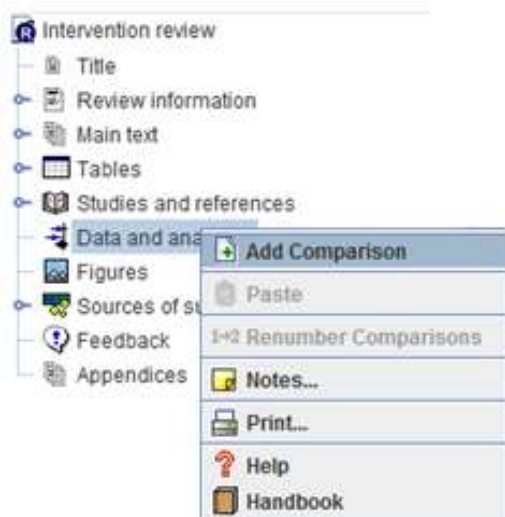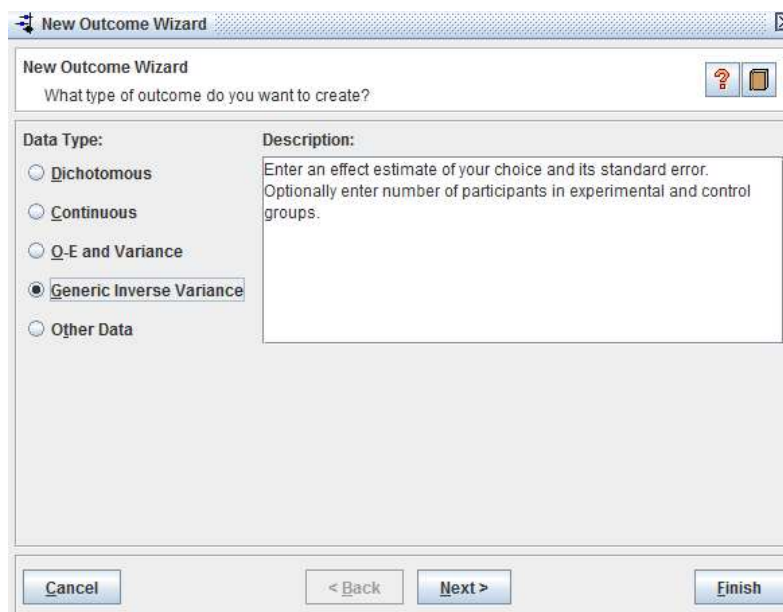5. Copy the data from an Excel file and paste it in the appropriate column in RevMan

B: ODDS RATIOS, RELATIVE RISKS, HAZARD RATIOS IN *METAFOR* VERSION 2.4-0 (R)

The *metafor* package in R can be used for meta-analysis of OR, RR, or HR, in either arm-based on contrast-based forms (42). Data can be loaded into R as an Excel file.

metafor offers a variety of meta-analysis methods (including DerSimonian and Laird (41)) to estimate the heterogeneity in a RE model (42). By default, *metafor* (version 2.4-0) uses the REML method for RE models (43); we suggest the Paule-Mandel estimator for the between-study variance, as recommended by the authors in (44).

Arm-based data

When analysing arm-based data, the Excel file consists of the study ID (Sid), the number of individuals that experienced the event (in this example, the number of people who are still dyspeptic) in the treatment arm (denoted by a) and the control arm (denoted by c). The total number of individuals in the treatment and control arms are denoted by n1 and n2 respectively (Figure 1)

	A	B	C	D	E
1	Sid	n1	a	n2	c
2	Blum (OCAY)	164	119	164	130
3	Koelz	89	67	92	73
4	McColl	154	121	154	143
5	Talley (ORCHID)	133	101	142	111
6	Talley (USA)	150	81	143	72

Figure 1. Excel spreadsheet containing arm-level data for studies comparing the removal of *H. pylori* to control in reducing dyspepsia

Study	Treatment		Control		Weight	OR [95% CI]
	Events	Total	Events	Total		
Blum (OCAY)	119	164	130	164	24.46%	0.69 [0.42, 1.15]
Koelz	67	89	73	92	13.07%	0.79 [0.39, 1.59]
McColl	121	154	143	154	12.14%	0.28 [0.14, 0.58]
Talley (ORCHID)	101	133	111	142	20.10%	0.88 [0.50, 1.55]
Talley (USA)	81	150	72	143	30.22%	1.16 [0.73, 1.83]
FE Model					100.00%	0.77 [0.60, 1.00]

Odds Ratio: 0.1   0.4   1   3

Figure 2. Meta-analysis of arm-level data comparing the effectiveness of *H.pylori* eradication in non-ulcer dyspepsia

The code that was used to conduct the analysis and create the resulting forest plot (Figure 2) is given in (Figure 3). The code is annotated to highlight other options that are available for the method of analysis, the models used and the effect measures. Output from R is displayed in blue.

```
# Set working directory
setwd(O:/)

# Load packages
library(metafor)
library(xlsx)

###############################
####   Arm-based data   ###
###############################

# Load data from the "arm-based" worksheet in an Excel file entitled "H.pylori Example.xlsx"
data.arm <- read.xlsx(file="H.pylori Example.xlsx", sheetName="arm-based", header=TRUE)

# Preview a maximum of the first 6 rows of the data
head(data.arm)

# Fit a Fixed Effect model
model.arm <- rma(ai=a,ci=c,                        # Specify the number of individuls in the each arm that experienced the event
        n1i=n1, n2i=n2,                            # Specify the total number of individuals in both treatment arms
        measure="OR",                             # Specify effect measures
        data=data.arm,                            # Specify data
        slab= Sid,                                # Specify columns that contain details of studies
        method="FE"                               # Specify the model
        )

# Display the MA results
model.arm

 # R displays a summary of the meta-analysis model. It first notes that a fixed effect model was fitted, where k = 5 studies were included.
 # The results of a Q-test for heterogeneity are displayed, followed by the pooled summary results (LOR estimate = -0.26, CI = (-0.51, -0.00)).
# Fixed-Effects Model (k = 5)
#
# I^2 (total heterogeneity / total variability):   63.03%
# H^2 (total variability / sampling variability):  2.71
#
# Test for Heterogeneity:
#   Q(df = 4) = 10.8201, p-val = 0.0287
#
# Model Results:
#
#  estimate    se    zval   pval   ci.lb   ci.ub
# -0.2553  0.1287  -1.9838  0.0473  -0.5076  -0.0031  *
#
#  ---
#  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Create a forest plot
forest(x=model.arm,                                # Specify the MA model used to construct the forest plot
    xlim=c(-7,4),                                  # Specify horizontal limits of the plot
    at=log(c(0.1, 0.4, 1, 3)),                     # Position ticks on the x-axis
    showweights=TRUE,                              # Make the study weights visible
    ilab=cbind(data.arm$a, data.arm$n1, data.arm$c, data.arm$n2),   # Specify column names of the study data
    ilab.xpos = c(-4.75,-4.0,-3, -2.25),           # Specify location for the columns added in ilab
    atransf = exp,                                 # Exponential transformation of LOR to OR
    refline=0,                                     # Draw a line for the null value at 1
    xlab= "Odds Ratio")                            # Label the x-axis

# Add labels/titles for the forest plot
text(c(-4.75, -4.0, -3, -2.25),                    # Horizontal position of the labels
    y=5.75,                                        # Vertical position of the labels
    c("Events", "Total", "Events", "Total"),       # Labels
    font=4)                                        # Font is bolded and italicised

text(c(-6.5, -4.5, -2.7, 1.8, 3),
    y=6.25,
    c("Study", "Treatment", "Control", "Weight", "OR [95% CI]"),
    font=2)
```
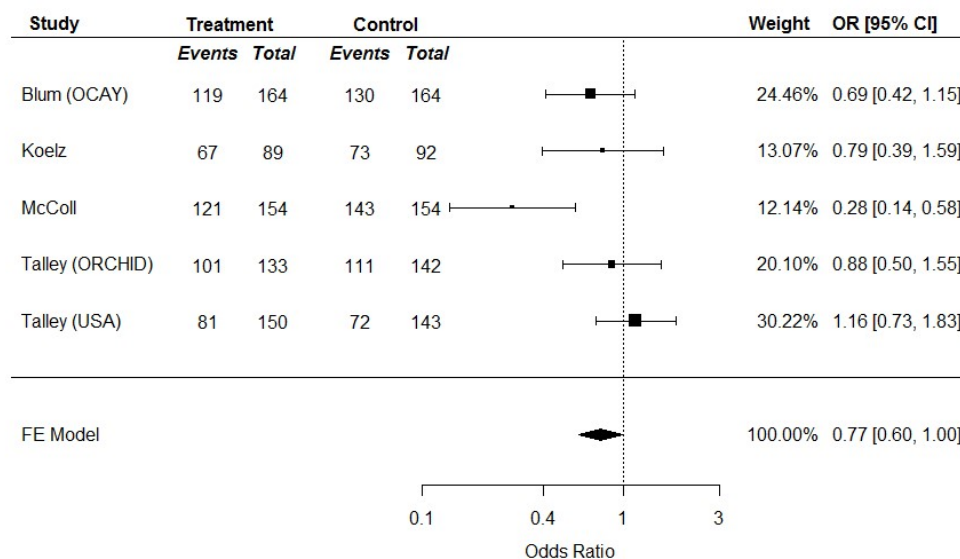
Figure 3. Code for the meta-analysis of arm-based data

The function rma() (42), which fits fixed- and random effects meta-analysis models, has the following generic form:

rma(ai, bi, ci, di, n1i, n2i, x1i, x2i, t1i, t2i, measure, method, data, slab)

Relevant arguments are assigned values depending on the type of analysis that is desired. Table S.1 details arguments used for each analysis.

Table S.1. Argument statements for specific meta-analyses

Statistic	Effect Measure	Data	Model Type
Odds Ratio (OR)	measure= "OR"	*2 x 2 Table:* ai= upper left cell ci= lower left cell bi=upper right cell di=lower right cell	*Fixed Effect:* method= "FE"
Risk Ratio/ Relative Risk (RR)	measure= "RR"	*Alternatively,* ai= upper left cell ci= lower left cell n1i= group size treatment 1 n2i= group size treatment 2	*Random Effects:* *Paule-Mandel estimator:* method= "PM"
Hazard Ratio (HR) (using the Incidence Rate Ratio, IRR)	measure = "IRR"	x1i= No. of events in group 1 x2i= No. of events in group 2 t1i= total person-time in group 1 t2i= total person-time in group 2	*Der-Simonian and Laird:* method= "DL"

Contrast-based data

For the analysis of contrast-based data, the Excel file comprises of the study ID (Sid), the log-odds ratio (log_OR) and the corresponding standard deviation (se_OR) for each included study (Figure 4).

Figure 4. Excel spreadsheet containing contrast-level data for studies comparing the removal of *H. pylori* to control in reducing dyspepsia

The code that was used to create the forest plot for the analysis (Figure 5) is given in Figure 6. The only way the rma() function for contrast-based data differs from the function used for arm-based data is the way the data are input. The argument 'yi=' is used to assign the LOR/LRR/LHR values and corresponding SE is assigned using the argument 'sei='. Output from R is displayed in blue.
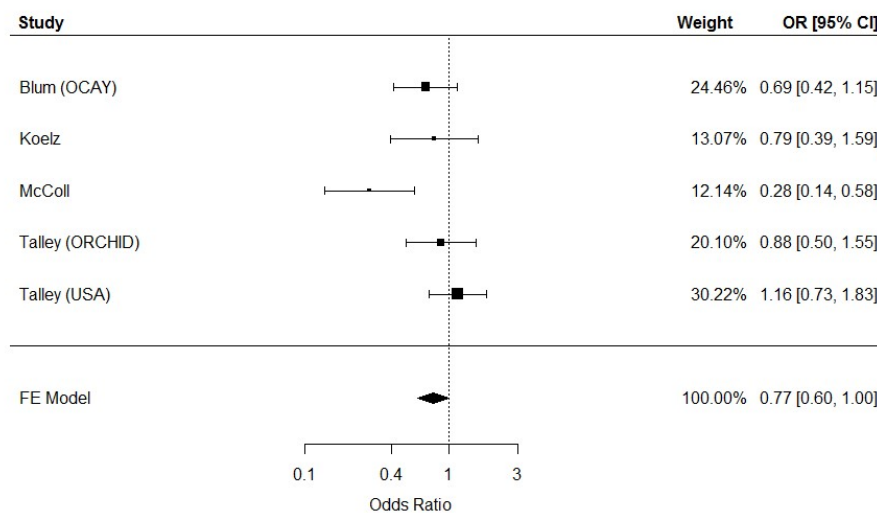
Figure 5. Meta-analysis of contrast-level data comparing the effectiveness of *H.pylori* eradication in non-ulcer dyspepsia

```
# Set working directory
setwd(O:/)

# Load packages
library(metafor)
library(xlsx)

##################################
####  Contrast -based data   ###
##################################

# Load data from the "contrast-based" worksheet in an Excel file entitled "H.pylori Example.xlsx"
data.con <- read.xlsx(file="H.pylori Example.xlsx", sheetName="contrast-based", header=TRUE)

# Preview a maximum of the first 6 rows of the data
head(data.con)

# Fit a Fixed Effect model
model.con <- rma(yi=log_OR, sei=se_OR,            # Specify the log-ratio measure and the corresponding SE
        measure="OR",                             # Specify effect measures
        data=data.con,                            # Specify data
        slab= Sid,                                # Specify columns that contain details of studies
        method="FE")                              # Specify the model to be fit

# Display the MA results
model.con

 # R displays a summary of the meta-analysis model. It first notes that a fixed effect model was fitted, where k = 5 studies were included.
 # The results of a Q-test for heterogeneity are displayed, followed by the pooled summary results (LOR estimate = -0.26, CI = (-0.51, -0.00)).
 # Fixed-Effects Model (k = 5)
 #
 # I^2 (total heterogeneity / total variability):  63.03%
 # H^2 (total variability / sampling variability):  2.71
 #
 # Test for Heterogeneity:
 #   Q(df = 4) = 10.8201, p-val = 0.0287
 #
 # Model Results:
 #
 #  estimate    se    zval   pval   ci.lb   ci.ub
 # -0.2553  0.1287  -1.9838  0.0473  -0.5076  -0.0031  *
 #
 #  ---
 #  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Create a forest plot
forest(x=model.con,                               # Specify the MA model used to construct the forest plot
    xlim=c(-7,7),                                 # Specify horizontal limits of the plot
    at=log(c(0.1, 0.4, 1, 3)),                    # Position ticks on the x-axis
    showweights=TRUE,                             # Make the study weights visible
    atransf = exp,                                # Exponential transformation of estimates and axes
    refline=0,                                    # Draw a line for the null value at 0
    xlab="Odds Ratio")                            # Label the x-axis with the effect measure

# Add labels/titles for the forest plot
text(c(-6.5, 4.1, 6.1),
    y=6.25,
    c("Study", "Weight", "OR [95% CI]"),
    font=2)
```

Figure 6. Code for the meta-analysis of contrast-based data

C.1. Probability models

The Mantel-Haenszel method is recommended when aggregating studies with sparse data, e.g. when there are zero cell counts. RevMan adds a continuity correction of 0.5 to all the cells in a 2 x 2 table where there is at least one zero cell: a procedure that is not only unnecessary but also incorrect (34). Other software such as the *metafor* package in R can compute a Mantel-Haenszel OR or RR without a continuity correction.

R: metafor

The *rma.mh()* command in the *metafor* package (version 2.4-0) in R fits fixed effect models to 2 x 2 tables using the Mantel-Haenszel method (42). The function *rma.mh()* uses exactly the same arguments as the *rma()* function for effect measures and data entry (though the option 'method=' will not work), which are described in detail in Table S.1 in the Appendix B on arm-based data with non-zero cells. By default, a continuity correction is applied to trials with zero cells in order to include them in forest plots, but the Mantel-Haenszel estimates do not include a continuity correction.

An example of sparse dichotomous data is shown in Figure 7. For each study, the number of newborns colonised by group B strep in the intrapartum antibiotic prophylaxis (IAP) and control arms are denoted *a* and *c,* respectively. The total number of mothers who gave birth in the IAP and control arms are denoted by *n1* and *n2*, respectively (35).

	A	B	C	D	E
1	Sid	a	c	n1	n2
2	Boyer 1986	0	4	85	79
3	Matorras 1991	0	3	60	65
4	Tuppurainen 1989	1	5	88	111

Figure 7. Excel spreadsheet illustrating sparse dichotomous data

The code for the forest plot for this analysis (Figure 8) is given in Figure 9. Output from R is displayed in blue.
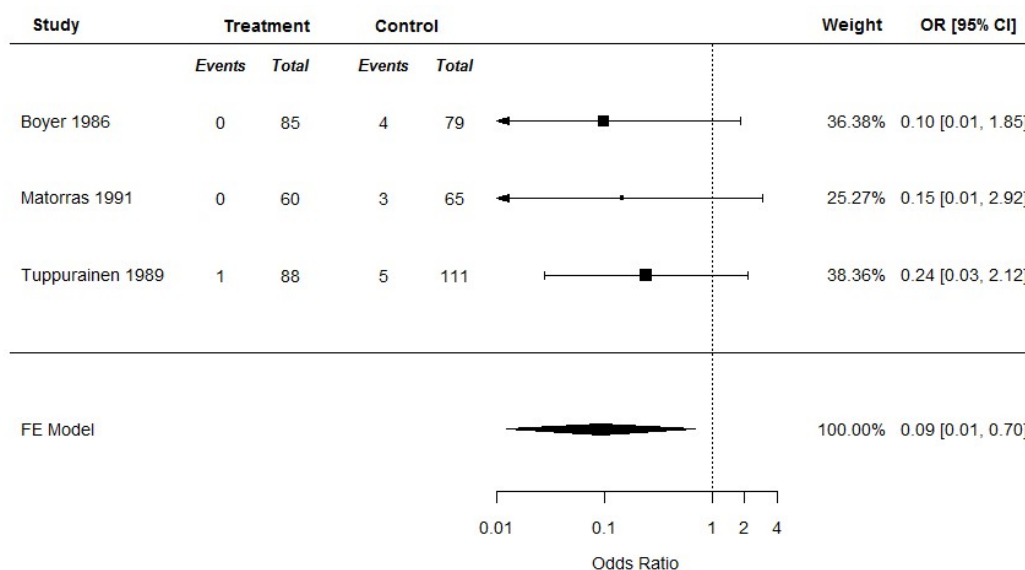

Figure 8. Meta-analysis of the effectiveness of IAP at preventing EOGBS in newborns born to colonised mothers

```
# Set working directory
setwd(O:/)

# Load packages
library(metafor)
library(xlsx)

# Load data from the "Ohlsson" worksheet in an Excel file entitled "MHEx.xlsx"
data <- read.xlsx(file="MHEx.xlsx", sheetName="Ohlsson", header=TRUE)

# Preview a maximum of the first 6 rows of the data
head(data)

####################################################
##############          Odds Ratio          ##############
####################################################

model.mh1 <- rma.mh(ai=a, ci=c,  n1i=n1, n2i=n2,          # Specify data from the 2 x 2 table
        measure="OR",                                      # Specify Effect Measures
        correct="FALSE",                                   # Do not apply a continuity correction to zero cells when computing CMH test
        data=data,                                         # Specify data
        slab= Sid)                                         # Insert labels/ study names


# Display the MA results
model.mh1

 # R displays a summary of the meta-analysis model. It first notes that a fixed effect model was fitted, where k = 3 studies were included.
 # The results of a Q-test for heterogeneity are displayed, followed by the pooled summary results (LOR estimate = -2.37, CI = (-4.38, -0.36), OR
 # estimate = 0.09, CI = (0.01, 0.70)).
 # Fixed-Effects Model (k = 3)
 #
 # I^2 (total heterogeneity / total variability):  0.00%
 # H^2 (total variability / sampling variability): 0.42
 #
 # Test for Heterogeneity:
 # Q(df = 2) = 0.8432, p-val = 0.6560
 #
 # Model Results (log scale):
 #
 #  estimate    se    zval   pval   ci.lb   ci.ub
 # -2.3702  1.0280  -2.3057  0.0211  -4.3849  -0.3554
 #
 # Model Results (OR scale):
 #
 #  estimate  ci.lb   ci.ub
 # 0.0935  0.0125  0.7009
 #
 # Cochran-Mantel-Haenszel Test:   CMH = 8.4479, df = 1, p-val = 0.0037
 # Tarone's Test for Heterogeneity: X^2 = 1.5014, df = 2, p-val = 0.4720


forest(model.mh1,
     xlim=c(-15,7),                                        # Specify horizontal limits of the plot
     #alim=c(0, 1.1),                                      # Specify limits of the actual x-axis
     at=log(c(0.01, 0.1, 1, 2, 4)),                        # Position the x-axis tick marks
     ilab=cbind(data$a, data$n1, data$c, data$n2),         # Specify column names of the study data to be included in plot
     ilab.xpos = c(-10.5, -9, -7, -5.5),                   # Specify location for the columns added in ilab
     refline=0,                                            # Positioning the vertical reference line (on log scale)
     showweights = TRUE,                                   # Include the inverse variance weights
     atransf = exp,                                        # Transform log-measures for ORs and axes
     cex=0.9,                                              # Expansion factor for the Forest Plot
     xlab= "Odds Ratio")                                   # Label the x-axis with the effect measure

# Labels and Headings
text(c(-14, -9.5, -6.5,  3, 5.5),                          # Positioning labels horizontally
    y=4.25,                                                # Positioning labels vertically
    c("Study", "Treatment", "Control", "Weight", "OR [95% CI]"),  # Headings
    font=2,                                                # Font is bolded
    cex=0.9)                                               # Expansion factor for the headings

text(c(-10.5, -9, -7, -5.5),                               # Positioning labels horizontally
    y=3.75,                                                # Positioning labels vertically
    c("Events", "Total", "Events", "Total"),              # Labels
    font=4,                                                # Font is bolded and italicised
    cex=0.8)                                               # Expansion factor for the headings
```

Figure 9. Code for meta-analysis of log odds ratios conducted using the Mantel-Haenszel method

The Mantel-Haenszel estimated OR and RR for the example above are reported in Table S.2 below. The table compares these estimates to the Mantel-Haenszel estimated OR and RR with a continuity correction of 0.5, the OR estimated using the Peto method and the OR and RR calculated using the inverse variance method with a continuity correction of 0.5.

Table S.2. Estimates for ORs and RRs, and corresponding 95% CIs for sparse data using different meta-analysis methods.

	Odds Ratio (OR)	Relative Risk/ Risk Ratio (RR)
Mantel Haenszel Method without Continuity Correction	0.094 (0.013, 0.701)	0.097 (0.014, 0.694)
Mantel Haenszel Method with Continuity Correction	0.163 (0.0367, 0.726)	0.170 (0.039, 0.735)
Peto Method	0.195 (0.065, 0.587)	
Inverse Variance Method with Continuity Correction	0.169 (0.037, 0.761)	0.177 (0.040, 0.778)

C.2. Rate models

RevMan currently does not have an option for pooling Poisson data through the Mantel-Haenszel approach, however this can be done using *metafor* in R.

R: *metafor*

An example of sparse event data is shown in Figure 10. For each study, the number of malignancies in the Etanercept (50 mg twice weekly) and placebo arms are denoted by *E1* and *E2,* respectively. The total person-years at risk in the Etanercept (50 mg twice weekly) and placebo arms are denoted by *T1* and *T2*, respectively (45).

Figure 10. Excel spreadsheet illustrating sparse event data

The code for the forest plot for this analysis (Figure 11) is given in Figure 12. Output from R is displayed in blue.

Study	Treatment		Control			IRR [95% CI]
	Events	Person-Years	Events	Person-Years		
Leonardi 2003	2	35.8	2	34.9		0.97 [0.14, 6.92]
Papp 2005	4	42.6	0	41		8.66 [0.47, 160.88]
Tyring 2006	3	68.8	1	65.9		2.87 [0.30, 27.62]
FE Model						2.90 [0.78, 10.73]

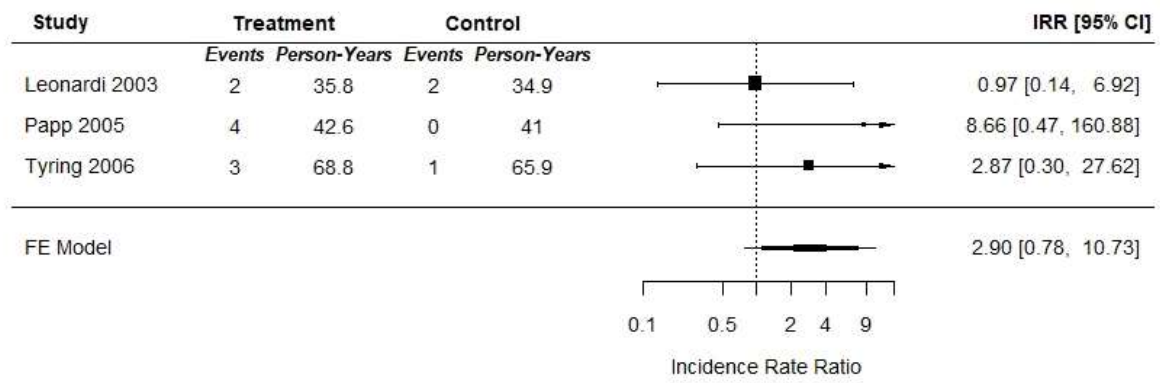0.1   0.5   2   4   9

Incidence Rate Ratio

Figure 11. Meta-analysis of the safety of Etanercept (50 mg twice weekly) in terms of incidence of malignancies in adults with psoriatic disease

```
# Set working directory
setwd(O:/)

# Load packages
library(metafor)
library(xlsx)

# Load data from the "Ohlsson" worksheet in an Excel file entitled "MHEx.xlsx"
data <- read.xlsx(file="MHEx.xlsx", sheetName="Dommasch", header=TRUE)

# Preview a maximum of the first 6 rows of the data
head(data)

##################################################
#############     Incidence Rate Ratio     #############
##################################################

model.mh1 <- rma.mh(x1i=E1, x2i=E2, t1i=T1, t2i=T2,        # Specify event and time at risk data
        measure="IRR",                                       # Specify Effect Measures
        correct="FALSE",                                     # Do not apply a continuity correction to zero cells
        data=data,                                           # Specify data
        slab= Sid)                                           # Insert labels/ study names

# Display the MA results
model.mh1

 # R displays a summary of the meta-analysis model. It first notes that a fixed effect model was fitted, where k = 3 studies were included.
 # The results of a Q-test for heterogeneity are displayed, followed by the pooled summary results (logIRR estimate = 1.06, CI = (-0.24, 2.37), IRR
 # estimate = 2.90, CI = (0.78, 10.73)).
 # Fixed-Effects Model (k = 3)
 #
 # I^2 (total heterogeneity / total variability):  0.00%
 # H^2 (total variability / sampling variability): 0.86
 #
 # Test for Heterogeneity:
 # Q(df = 2) = 1.7269, p-val = 0.4217
 #
 # Model Results (log scale):
 #
 # estimate   se        zval     pval      ci.lb     ci.ub
 # 1.0644    0.6676    1.5943   0.1109    -0.2441   2.3729
 #
 # Model Results (IRR scale):
 #
 # estimate  ci.lb      ci.ub
 # 2.8991    0.7834    10.7281
 #
 # Mantel-Haenszel Test: MH = 2.7911, df = 1, p-val = 0.0948

forest(model.mh1,
    xlim=c(-15,8),                                          # Specify horizontal limits of the plot
    #alim=c(0, 1.1),                                        # Specify limits of the actual x-axis
    at=log(c(0.1, 0.5, 1, 2, 4, 9, 16)),                   # Position the x-axis tick marks
    ilab=cbind(data$E1, data$T1, data$E2, data$T2),        # Specify column names of the study data to be included in plot
    ilab.xpos = c(-10.5, -8.5, -6.5, -4.5),                # Specify location for the columns added in ilab
    refline=0,                                             # Positioning the vertical reference line (on log scale)
    showweights = FALSE,                                   # Do not include the inverse variance weights
    atransf = exp,                                         # Transform log-measures for ORs and axes
    cex=0.9,                                               # Expansion factor for the Forest Plot
    xlab= "Incidence Rate Ratio")                          # Label the x-axis with the effect measure

# Labels and Headings
text(c(-14, -9.5, -5.5,  6.75),                            # Positioning labels horizontally
    y=4.5,                                                 #  Positioning labels vertically
    c("Study", "Treatment", "Control", "IRR [95% CI]"),    #  Headings
    font=2,                                                # Font is bolded
    cex=0.9)                                               # Expansion factor for the headings

text(c(-10.5, -9.5, -6.5, -4.5),                           # Positioning labels horizontally
    y=3.75,                                                # Positioning labels vertically
    c("Events", "Person-Years", "Events", "Person-Years"), # Labels
    font=4,                                                # Font is bolded and italicised
    cex=0.8)                                               # Expansion factor for the headings
```

Figure 12. Code for meta-analysis of log incidence rate ratios conducted using the Mantel-Haenszel method

REFERENCES

1.National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. London; 2013.

2.National Institute for Health and Care Excellence. Developing NICE guidelines: the manual. National Institute for Health and Care Excellence; 2014.

3.Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU Technical Support Document 1: Introduction to evidence synthesis for decision making. 2011.

4.Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU Technical Support Document 2: A generalised linear modelling framework for pair-wise and network meta-analysis of randomised controlled trials. 2011.

5.Dias S, Sutton AJ, Welton NJ, Ades AE. NICE DSU Technical Support Document 3: Heterogeneity: subgroups, meta-regression, bias and bias-adjustment. 2011.

6.Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. NICE DSU Technical Support Document 4: Inconsistency in networks of evidence based on randomised controlled trials. 2011.

7.Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU Technical Support Document 5: Evidence synthesis in the baseline natural history model. 2011.

8.Dias S, Sutton AJ, Welton NJ, Ades AE. NICE DSU Technical Support Document 6: Embedding evidence synthesis in probabilistic cost-effectiveness analysis: Software choices. 2011.

9.Ades AE, Caldwell DM, Reken S, Welton NJ, Sutton AJ, Dias S. NICE DSU Technical Support Document 7: Evidence synthesis of treatment efficacy in decision making: a reviewer's checklist. 2012.

10.Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis. Chichester: Wiley; 2009 2009.

11.Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane handbook for systematic reviews of interventions. 2nd ed. Chichester: John Wiley & Sons; 2019.

12.Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Methods for meta-analysis in medical research. London: Wiley; 2000.

13.Perneger TV. Estimating the relative hazard by the ratio of logarithms of event-free proportions. Contemporary Clinical Trials. 2008;29(5):762-6.

14.Deeks JJ. Issues on the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. Statistics In Medicine. 2002;21:1575-600.

15.Rhodes KM, Turner RM, Higgins JPT. Empirical evidence about inconsistency among studies in a pair-wise meta-analysis. Research Synthesis Methods. 2016;7(4):346-70.

16.Guyot P, Ades AE, Ouwens MJNM, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Medical Research Methodology. 2012;12:9.

17.Wei Y, Royston P. Reconstructing time-to-event data from published Kaplan–Meier curves. The Stata Journal. 2017;17(4):786-802.

18.Dias S, Ades AE, Welton NJ, Jansen JP, Sutton AJ. Network meta-analysis for decision making: Wiley; 2018. 488 p.

19.Jansen JP. Network meta-analysis of survival data with fractional polynomials. BMC Medical Research Methodology. 2011;11(1):1-14.

20.Ouwens MJNM, Philips Z, Jansen JP. Network meta-analysis of parametric survival curves. Research Synthesis Methods. 2010;1:258-71.

21.National Institute for Health and Care Excellence. Lung cancer: diagnosis and management. National Institute for Health and Care Excellence; 2019.

22.Guyot P, Ades AE, Beasley M, Lueza B, Pignon J-P, Welton NJ. Extrapolation of survival curves from cancer trials using external information. Medical Decision Making. 2016;37:353-66.

23.Latimer NL. Survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data: inconsistencies, limitations and a practical guide. Medical Decision Making. 2013;33:743-54.

24.Latimer NR, Abrams KR. NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching. 2014.

25.Ades AE, Mavranezouli I, Dias S, Welton NJ, Whittington C, Kendall T. Network meta-analysis with competing risk outcomes. Value in Health. 2010;13(8):976-83.

26.National Institute for Health and Care Excellence. Psychosis and schizophrenia in adults: Prevention and management. National Institute for Health and Care Excellence; 2014.

27.Loveman E, Turner D, Hartwell D, Cooper K, Clegg A. Infliximab for the treatment of adults with psoriasis. NICE Single Technology Appraisal. 2007 2007.

28.Epstein D, Sutton A. Modelling correlated clinical outcomes in health technology appraisal. Value in Health. 2011;14(6):793-9.

29.Schering Plough Ltd. Golimumab for ankylosing spondylitis. NICE Single Technology Appraisal. 2011.

30.Corbett M, Chehadah F, Biswas M, Moe-Byrne T, Palmer S, Soares M, et al. Certolizumab pegol and secukinumab for treating active psoriatic arthritis following inadequate response to disease-modifying antirheumatic drugs: A systematic review and economic evaluation. Health Technology Assessment. 2017;21(56):1-326.

31.Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute. 1959;22:719-48.

32.Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade before and after myocardial infarction: an overview of randomised trials. Progress in Cardiovascular Diseases. 1985;27:335-71.

33.Greenland S, Salvan A. Bias in the one-step method for pooling study results. Statistics In Medicine. 1990;9:247-52.

34.Fleiss JL, Berlin JA. Effect sizes for dichotomous data. In: Cooper H, Hedges LV, Valentine JC, editors. The handbook of research synthesis and meta-analysis. 2nd ed. New York: Russell Sage Foundation; 2009. p. 237-53.

35.Ohlsson A, Shah VS. Intrapartum antibiotics for known maternal Group B streptococcal colonization. Cochrane Database of Systematic Reviews. 2014(6).

36.Blum AL, Talley NJ, O'Morain C, van Zanten SV, Labenz J, Stolte M, et al. Lack of effect of treating Helicobacter pylori infection in patients with nonulcer dyspepsia. Omeprazole plus Clarithromycin and Amoxicillin Effect One Year after Treatment (OCAY) Study Group. The New England journal of medicine. 1998;339(26):1875-81.

37.Albain KS, Swann RS, Rusch VW, Turrisi AT, 3rd, Shepherd FA, Smith C, et al. Radiotherapy plus chemotherapy with or without surgical resection for stage III non-small-cell lung cancer: A phase III randomised controlled trial. Lancet (London, England). 2009;374(9687):379-86.

38.Therneau TM, Lumley T. survival: Survival analysis. R package version 2.44-1.1. 2019.

39.Elliott WJ, Meyer PM. Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis. Lancet. 2007;369(9557):201-7.

40.Hooper L, Summerbell CD, Higgins JPT, Thompson RL, Clements G, Capps N, et al. Reduced or modified dietary fat for preventing cardiovascular disease. Cochrane Database of Systematic Reviews. 2000.

41.DerSimonian R, Laird N. Meta-analysis of clinical trials. Controlled Clinical Trials. 1986;7:177-88.

42.Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. Journal of Statistical Software. 2010;36(3).

43.Viechtbauer W. Bias and efficiency of meta-analytic variance estimators in the random-effects model. Journal of Educational and Behavioral Statistics. 2005;30(3):261-93.

44.Veroniki AA, Jackson D, Viechtbauer W, Bender R, Bowden J, Knapp G, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. Research Synthesis Methods. 2016;7(1):55-79.

45.Dommasch ED, Abuabara K, Shin DB, Nguyen J, Troxel AB, Gelfand JM. The risk of infection and malignancy with tumor necrosis factor antagonists in adults with psoriatic disease: A systematic review and meta-analysis of randomized controlled trials. Journal of the American Academy of Dermatology. 2011;64(6):1035-50.