Book title: **Oncology Clinical Trials, Second Edition**
Section or Part title (if applicable, or if known):
Chapter title: **Systematic reviews and meta-analysis**
Corresponding author for this chapter: **Claire Vale**

First author:
Name and degree(s) (as to appear on your chapter): **Claire Vale,** PhD
Institutional Affiliations (title, department, institution, city, state): Senior Research Scientist, Medical Research Council Clinical Trials Unit at UCL
Mailing address: Aviation House, 125 Kingsway, London, WC2B 6NH, UK
Phone: +44 (0)2076704723
Fax: +44 (0)20 76704685
Email address: claire.vale@ucl.ac.uk

Second author:
Name and degree(s) (as to appear on your chapter): **Sarah Burdett,** MSc
Institutional Affiliations (title, department, institution, city, state): Senior Research Scientist, Medical Research Council Clinical Trials Unit at UCL
Mailing address: Aviation House, 125 Kingsway, London, WC2B 6NH, UK
Phone:+44 (0)20 76704722
Fax: +44 (0)20 76704685
Email address: sarah.burdett@ucl.ac.uk

Third author:
Name and degree(s) (as to appear on your chapter): **David Fisher,** MSc
Institutional Affiliations (title, department, institution, city, state): Statistician, Medical Research Council Clinical Trials Unit at UCLMailing address: Aviation House, 125 Kingsway, London, WC2B 6NH, UK
Phone: 020 76704646
Fax: +44 (0)20 76704685
Email address: d.fisher@ucl.ac.uk

Fourth author:
Name and degree(s) (as to appear on your chapter): **Larysa Rydzewska,** BSc
Institutional Affiliations (title, department, institution, city, state): Research Scientist, Medical Research Council Clinical Trials Unit at UCL
Mailing address: Aviation House, 125 Kingsway, London, WC2B 6NH, UK
Phone: +44(0)20 76704721
Fax: +44 (0)20 76704685
Email address: l.rydzewska@ucl.ac.uk

Fifth author:
Name and degree(s) (as to appear on your chapter): **Jayne Tierney,** PhD
Institutional Affiliations (title, department, institution, city, state): Reader in Evidence Synthesis, Medical Research Council Clinical Trials Unit at UCL
Mailing address: Aviation House, 125 Kingsway, London, WC2B 6NH, UK
Phone:+44(0)20 76704724
Fax: +44 (0)20 76704685
Email address: jayne.tierney@ucl.ac.uk

Introduction

Every year, hundreds of thousands of research articles are published in the biomedical literature, making it near impossible for clinicians, policy makers and patients to stay abreast of the latest healthcare developments. There is therefore, a clear need for summaries of research to advise clinical practice and policy, inform patient choice and guide new research.

Traditional reviews are common in the medical literature, but can be problematic for summarising research findings. They are usually narrative, which can make it difficult to make sense of studies with conflicting results. Also, they often lack clear aims and transparent methodology, which can prevent others being able to reproduce or interpret their findings. They are commonly restricted to a particular group of studies, which may or may not be representative of the whole body of evidence. Moreover, they frequently present the opinion of individual authors, who, even with the best intentions, may not be wholly objective. Systematic reviews, in contrast (Table 1), aim to comprehensively identify and objectively appraise and collate all research studies relating to a specific research question, using explicit methodology. The methods used aim to minimise bias(1) and therefore, provide reliable findings, upon which decisions can be made.

Importantly, the objectives of the review and the nature of the trials to be included or excluded are well defined, to mimimise the risk of basing the review on a selected group of trials (Table 1). The objectives and eligibility criteria are underpinned by a comprehensive search strategy to ensure that all the relevant studies are identified. Data from the eligible studies are then extracted in a consistent way to enable an assessment of the quality of each study to be made, and the results to be synthesized. Quantitative synthesis, in the form of a meta-analysis, aids interpretation of the results and informs the conclusions of the review. The methods relating to all of these systematic review features should be clearly set out, in advance, in a protocol or other plan(2), so that they are not unduly influenced by the results of the trials. Moreover, provided the protocol is made publicly available, this also allows others to appraise or even reproduce the approach taken. Finally, the review methods, characteristics and results of each trial, and the results and interpretation of the review are presented or reported in a structured and consistent way, ideally adhering to the Preferred Reporting Items for Systematic review and Meta-Analysis (PRISMA) guidelines(3, 4).

Where new treatments only offer modest improvements in outcome compared to an existing standard, many trials lack power to detect such effects. This may be, for example, because trial sample size calculations assumed larger treatment effects, or simply because recruitment fell short of targets, but means that inconclusive

or equivocal study results are relatively common, particularly in oncology. Hence, most systematic reviews include a meta-analysis, to quantitatively combine the results of included studies and obtain an overall estimate of the treatment effect, based on all relevant trials. As a meta-analysis is based on more participants than any of the individual trials, it provides greater power to establish whether a treatment effect exists, and greater confidence in any estimate of effect. A meta-analysis carried out within the context of a good quality systematic review will together represent the most complete and reliable summary of the evidence available. Conversely, a meta-analysis that is not completed as part of a systematic review may encounter the sorts of problems that commonly affect traditional reviews (Table 1).

	Traditional review	**Systematic review**
Protocol	No	Yes
Research objective	Unclear, non-specific	Specific objective based on patients, interventions comparisons, outcomes and study type (PICOS)
Eligibility criteria for studies	Unclear, not specified or selective	Explicit inclusion and exclusion criteria based on patients, interventions comparisons, outcomes and study type (PICOS)
Search strategy	Unclear, not specified or limited	Systematic, multiple sources of studies and inclusive
Evidence synthesis	Usually narrative qualitative	Usually quantitative (meta-analysis)
Conclusions	Subjective and potentially biased or otherwise unreliable	More objective and reliable, based on results
Reproducibility	Difficult without transparent methodology	Use of transparent methodology and clear reporting enables reproducibilty

Table 1: Traditional versus systematic reviews

Systematic reviews and meta-analyses typically make use of summary or aggregate data extracted from study publications or obtained from trial investigators. However, there is also a long tradition in oncology of conducting systematic reviews based on the central collection and re-analysis of the original individual participant data (IPD)(5, 6). Notable examples occur in breast(7), colorectal(8), head and neck(9) and lung cancer(10). Although the IPD approach can bring about substantial improvements to the data and analysis quality, which can sometimes lead to disparate results(11-14), much of the methodology is equally applicable to both types of data. Therefore, in this chapter, we will use examples of systematic reviews and meta-analysis

based on aggregate data and IPD to illustrate the key principles and provide an overview of the methods for each.

Most commonly, systematic reviews and meta-analyses are used to investigate the effects of treatments or other interventions on a particular disease or condition. Such reviews, which use evidence from randomised controlled trials (RCTs)(15), have been widely used in oncology as well as other areas of medicine and the methods used in their conduct are well established. However, when randomised evidence is lacking, systematic reviews of other study designs (e.g. cohort or case control studies) can also provide evidence about treatment efficacy, albeit that such reviews are not considered such a high level of evidence(16). Systematic reviews can also be used to synthesise results of prognostic or diagnostic test accuracy studies, with methods for each being developed accordingly(17-19). Largely irrespective of the nature of the research studies being synthesised, the underlying principles of systematic review methodology hold. Therefore, although the focus of this chapter is on systematic reviews of RCTs, much of it will be of relevance to systematic reviews of other study types.

It is now well recognised that systematic reviews are an optimal way to synthesise results of primary research and in particular, to resolve or confirm uncertainty around the effects of interventions. Therefore, before embarking on any review, it is worth establishing whether a pre-existing systematic review provides a comprehensive, up-to-date and robust evaluation of the question posed or at least whether one is planned, to avoid duplication of effort(20). Search strategies are available to locate reviews in the wider medical literature(21). Additionally, the Cochrane Database of Systematic Reviews(22) and the International Prospective Register of Systematic Reviews (PROSPERO)(23), managed by the Centre for Reviews and Dissemination (CRD), maintain searchable records and protocols of planned and ongoing systematic reviews. There may be of course be merit in duplicating a systematic review, for example if a prior one is sub-standard, or where substantial new information has become available.

Through the remainder of this chapter, using a variety of examples from the oncology literature, we will describe the key features of a systematic review, namely:

- Clear objectives and eligibility criteria
- Search strategy to identify all relevant studies
- Consistent data collection across studies
- Assessment of study validity and risk of bias
- Synthesis of study characteristics and results (meta-analysis)

- Structured presentation of results

.

1. Defining the objective and eligibility criteria

First and foremost, a systematic review should have answerable and relevant objectives, which should provide a clear, precise statement about the nature of the review. This is further qualified by detailed and specific criteria that outline the studies that are to be included and excluded. Broad objectives and eligibility criteria can make for a comprehensive review that is widely generalisable, but perhaps one that is too large, labour-intensive and hard to manage. If the scope is too broad, trials addressing quite different clinical questions may be combined, making interpretation difficult(15) or inappropriate. In contrast, a review with narrow objectives and eligibility criteria, will likely include fewer trials and produce results that are easier to interpret and certainly more manageable. However, the if the review comprises too few trials, there is a risk that the results will not reveal much more than the results of any individual trial, or that they will have very limited applicability(15).

The compromise is to define objectives and eligibility criteria that are inclusive, but at the same time minimise, as far as possible, the inevitable variability between studies, such as different treatment doses or scheduling, different participants or different clinical settings. Trials that are sufficiently similar in design are more likely to have comparable results, making review results easier to interpret and more meaningful. In particular, ensuring that the interventions of interest are similar can help minimise potential statistical heterogeneity, which can blight the interpretation of meta-analysis results. One way to tackle a systematic review that is broad in scope is to subdivide it into a series of related narrower questions, to reduce inconsistencies and aid interpretation. For example, a broad systematic review of the effects of adding chemotherapy to standard care in participants with non-small cell lung cancer (NSCLC)(10), was subdivided into a series of reviews with narrower objectives. These were defined by the type of standard treatment used, ranging from surgery and radical radiotherapy(24) through to best supportive care(25) and subsequently; how chemotherapy was given – either pre-operatively(26) or as adjuvant treatment(27), which reflected the stage of disease of the included participants.

The PICOS tool(21) is a useful aid to defining clear objectives and eligibility criteria, as it sets out the population (P) of included participants; the new intervention (I) being tested; the comparison (C) against which the intervention is being compared; which outcomes (O) the intervention might affect and the study designs (S) that are eligible for the review. Note that for treatment efficacy reviews, RCTs will tend to be the default study type.

For an aggregate data systematic review in NSCLC(28), the objective was to assess whether the effect of pre-operative chemotherapy (I) improved survival and reduced recurrence rates (O) when compared against standard of care (C) for patients with non-small cell lung cancer (P). The eligibility criteria are shown in Table 2.

PICOS elements	**Definition**	**Eligibility criteria for NSCLC systematic review**(28)
Population	Defines the included participants (e.g. by age, sex, condition under investigation)	Patients with NSCLC and no prior malignancy or chemotherapy
Intervention	The research or experimental treatment	Pre-operative chemotherapy plus surgery without or without radiotherapy
Comparator (control)	The control (e.g. standard of care, placebo)	Surgery with or without radiotherapy
Outcomes	The review endpoints or outcomes of interest (e.g. survival, weight loss, quality of life)	NA
Study types	Types of studies eligible for inclusion in the review (e.g. RCTs, cohort studies etc)	Completed RCTs (S)

Table 2. Example of eligibility criteria based on PICOS

Commonly, because systematic reviews aim to assess the effect of treatment on a number of outcomes, it would be too restrictive to use outcomes to define the eligibility criteria for trials to be included in a systematic review (Table 2).

2. Identifying all relevant studies

Searching for trials

Whilst defining and applying specific eligiblity criteria guards against study selection bias (Table 3), it cannot circumvent reporting biases. These occur when the characteristics of study results determine if, when, and where they are published. Therefore, the aim is to identify and include all relevant trials, irrespective of publication status, language, or the nature of the results, through systematic and comprehensive searches for trials across multiple sources, thereby reducing the risk of falling foul of such biases.

Bibliographic databases including MEDLINE, EMBASE and the Cochrane Central Register of Controlled Trials (CENTRAL(29)), which is also populated by trial reports identified in other databases and through hand searching of journals, are important sources of (mainly) published RCTs. Validated search filters and exemplar strategies that use these filters have been developed to assist reviewers in the retrieval of RCTs for MEDLINE, CENTRAL and EMBASE(21, 30). Although searching these three databases should be considered a minimum requirement for a systematic review(30), depending on the review topic, cancer- or region-specific databases may also be useful. A recent review of published systematic reviews indexed in MEDLINE(31) showed that a

median of four electronic databases were searched by review authors. When devising search strategies and choosing search terms for these databases, again, the PICOS tool is helpful, with population, intervention, comparison and study design being the most useful(21).

Whilst these databases are undoubtedly rich sources of reported trials, there are limitations. MEDLINE and EMBASE tend to include journals with higher impact factors, published in the English language, that cover more general medical than disease-specific themes topics. Added to this, there is now substantial empirical evidence that trials with more striking results are more likely to be published; published more quickly, in the English language, in more accessible sources and multiple times, compared to those with less striking results (Table 3). Thus, limiting searches to these major bibliographic databases potentially means including only those trials with the most striking results, leading to unrepresentative, biased and so unreliable conclusions.

Type of bias	**Description**
Biases associated with study selection, availability or reporting	
Study selection bias(32)	Trials with clinically or statistically significant results are more likely to be selected for inclusion in the review
Data availability bias(32)	Data from trials with clinically or statistically significant results are more likely to be made available readily
Publication bias(33)	Trials with clinically or statistically significant results are more likely to be published
Time lag bias(34)	Trials with clinically or statistically significant results are more likely to be published quickly
Multiple publication bias(35)	Trials with clinically or statistically significant results are more likely to be published multiple times
Location bias(36)	Trials with clinically or statistically significant results are more likely to be published in readily-accessible sources
Citation bias(37)	Trials with clinically or statistically significant results are more likely to be cited
Language bias(38)	Trials with clinically or statistically significant results are more likely to be published in the English language
Outcome reporting or availability bias(39)	Trial outcomes with clinically or statistically significant results are more likely to be published.
Biases associated with study design or analysis (40)	
Participant selection bias	Trial participants on the treatment arm of trials have more favourable characteristics or prognoses than participants on the control arm (prevented by random allocation and allocation concealment)
Performance and detection bias	Trial participants on the treatment arm receive additional care or are assessed more closely for positive outcomes than participants on the control arm
Attrition bias	Trial participants are followed up, or excluded from analyses, in such a way that results on the treatment arm appear more positive

Table 3: Biases that can affect the reliability of systematic review and meta-analysis results

It has been estimated that around 10% of all trials included in Cochrane systematic reviews are identified in the grey literature(30), which includes conference proceedings, book chapters, theses and reports. However, sources of unpublished trials are often overlooked in systematic reviews, with trial registries searched in only 19% and conference proceedings in only 16% of systematic reviews identified in a recent study(31). Given that results of many trials presented at conferences are never fully published(41), and others can take some time to be reported in full, conference proceedings, in particular should not be ignored. In oncology, searching the proceedings of both general cancer conferences, for example, the American Society of Clinical Oncology (ASCO) or European Society of Medical Oncology (ESMO), and cancer-specific conferences, for example, the World Conference on Lung Cancer (WCLC) or ASCO genitourinary cancers symposium can prove very useful. Finally, further trials can often be identified by hand-searching the reference lists of the publications of eligible trials or related reviews, and by searching clinical trial registers, for example, U.S. National Institute of Health ClinicalTrials.gov and by speaking to experts working in the field. For example, in a systematic review of pre-operative chemotherapy for NSCLC (28), only half of the eligible trials identified were indexed in MEDLINE. The remainder of the trials were identified through searches of conference proceedings, CENTRAL and by hand-searching the referecne lists of relevant articles.

Screening search results for eligible trials

Once all searches have been completed, the records retrieved must be screened for relevance and eligibility. Inevitably, any search strategy will pick up many irrelevant records that need to be removed, and searching multiple sources of trials can result in duplicate records. Often, scanning the titles and abstracts is enough to filter out obvious duplicates and irrelevant records. However, detailed checking of abstracts and full-text publications is needed to properly identify trials that fulfil the eligibility criteria and objectives of the review, and also to clarify and collate multiple reports arising from the results of a single trial. It is advisable for this process to be carried out by two reviewers, who should agree on eligibility (or otherwise) of the trials identified. A screening form listing questions relating to the eligibilty criteria is very useful at this stage, both to aid the screening process and to document the decisions made. Reference management software can also facilitate the management of this stage of the review process.. The flow of trial records though this process of screening search results is commonly recorded in a flow diagram, which is a mandatory requirement of the PRISMA reporting guidelines(4). An example from a systematic review and IPD meta-analysis of pre-operative chemotherapy for NSCLC(26), is shown in Figure 1.

[Figure 1: PRISMA flow diagram for systematic review of neoadjuvant chemotherapy for non-small cell lung cancer. Modified with permission from Elsevier (Lancet, 2014, 383, 1561-1571).]

3. Consistent data collection across studies

Once a list of eligible trials has been established, key descriptive data and results need to be obtained for each included trial. Data items to be collected for each trial should be outlined in advance in the systematic review protocol. Key data items often include, but are not limited to:

- Trial identifier / reference
- Lead investigator
- Details of study design, recruitment and follow-up (**S**)
- Methods of analysis and number of patients randomised/analysed
- Baseline characteristics (e.g. age, sex, stage) of participants (**P**)
- Planned treatment details (intervention and control arms) (**I /C**)
- Outcomes assessed and their defintions, and results for each (**O**)

Most systematic reviews rely on using aggregate summary data that has been extracted from trial reports, including journal articles, conference abstracts (or associated presentations), book chapters, theses and pharmaceutical company reports. Developing a standardised data collection form helps to ensure that, as far as possible, results and other information are extracted consistently for each trial, and also provides a record of which data were available and which were not. Indeed one of the challenges for systematic reviewers is obtaining all of the necessary information needed for the systematic review. For example, trial reports may not describe the design of a trial in sufficient detail to determine the methods used or they may only provide results for some of the outcomes of interest or for a particular subgroup of participants. These problems can be magnified in conference abstracts. For example, for a systematic review of of pre-operative chemotherapy for NSCLC(28), data on survival were only available for 7 the 12 eligible trials and for disease-free survival for 3 trials, representing only 75% and 35% of all participants randomised, respectively. Although the CONSORT(42) and CONSORT for abstracts(43) guidelines for trial reporting are no doubt helping in this regard, information on unpublished trials can be even harder to establish. Trial protocols, which are increasingly available through trial institute websites or summarised in registers such as ClinicalTrials.gov, can provide information on treatments and comparators, inclusion and exclusion criteria of participants and planned trial size

and timelines. However, until such time as results become readily available, obtaining results for unpublished trials remains an issue.

Clearly, if the reporting of trials is related to results, it is important to include as much published and unpublished data as possible to limit the potential influence of reporting biases (Table 3). This may mean, for example, seeking results directly from trial investigators for any outcomes that were not reported. However, recent evidence(31) suggests that very few systematic reviewers do this. Similarly, additional information on, for example, the methods of randomisation or concealment or reasons why participants were excluded from reported analyses can be sought from trial investigators to establish the quality of the included trials.

An alternative to extracting or obtaining aggregate data is the IPD approach, which involves the central collection of the raw data for each partcipant, from all the relevant trials worldwide. Although this may demand greater resources and be more time-consuming than a review of aggregate data, it can substantially improve both the quantity and quality of data available by virtue of including more trials (published and unpublished), participants, and outcomes(6, 44, 45). For example, a systematic review of pre-operative chemotherapy for NSCLC based on IPD(26), included data on 15 trials, representing 92% of all randomised patients for both survival and disease-free survival outcomes. All participants who had been excluded from the reported analyses for these trials were re-instated in the IPD review, the results of which were based on far more data than the aggregate data review conducted previously(28). The collection of IPD also enables standardisation of outcomes such as disease-free survival across trials and allows for detailed data checking(6, 44, 45).

4. Assessing study validity and risk of bias

The validity and reliability of the results of any systematic review are determined not only by efforts undertaken to ensure that all (and only) eligible trials have been included, but also by the quality of the included studies. Whilst reporting biases (Table 3) can to some degree be circumvented by extensive searching, obtaining results based on all trials, participants and outcomes, which may or may not require the collection and re-analysis of IPD; potential biases arising from inappropriate trial design, conduct or analysis (participant selection bias, performance and selection bias, Table 3) are harder to address. Therefore, assessments of the quality, or risk of bias, of included trials are now commonplace in systematic reviews, and are a requirement of the PRISMA reporting guidelines(4). One tool that is commonly used to assess the trial validity is the Cochrane Risk of Bias tool(46), in which reviewers assess whether the risk of various biases is low, high or unclear for each of the included trials and ultimately, for the meta-analysis results overall. The tool does have some limitations.

Reviewers are most often relying on information from trial reports to make judgements about the risk of bias. If the required information is scant or missing from trial reports, then assessments can be unreliable or limited(47, 48). Contacting trial investigators for further details to inform the trial quality assessment may be important. Moreover, judgements are somewhat subjective, leading to differing assessments being made by independent reviewers(49, 50) and can only ever highlight the potential for bias and not the actual validity of a trial and its results. The potential impact of observed biases within trials on the results of the systematic review should also be evaluated. This may be done qualitatively or through sensitivity analyses(51), if a meta-analysis is planned, but there is no standard approach(52).

5. Synthesis of study characteristics and results (meta-analysis)

Results of a systematic review can be presented in a narrative way, indeed most reports of systematic reviews tabulate and summarise details of the eligible trials, including design, recruitment, treatments used and characteristics of the included patients. However, systematic reviews of interventions will generally also incorporate a quantitative synthesis of trial results in the form of a meta-analysis.

Planning meta-analyses

If a meta-analysis is to be included in a systematic review, it is important that all aspects of the analysis are planned in advance and described in detail in the review protocol or analysis plan. This ensures that the analyses are guided by the objective(s), and not motivated and modified on the basis of known trial results or by the accumulating meta-analysis results, which could introduce bias. In addition, the planning stage can help, for example, to clarify definitions of outcomes to be assessed and the data items to be obtained.

Analysis plans should include the primary and secondary outcomes, their definitions; methods for analyses of efficacy, including those for exploring the impact of trial or participant characteristics; methods for measuring and accounting for heterogeneity and methods for assessing risk of bias for included trials. Of course, developing a protocol and analysis plan does not preclude unplanned, additional exploratory analysis, which can inform or add to the main results or help generate new hypotheses. However, any post hoc analyses should be justified, and clearly described as such in any report of the results.

Outcomes and effect measures

How treatment effects are measured in individual trials and meta-analysis is largely dictated by the outcomes of interest in the review. For example, a hazard ratio (HR) would commonly be used to assess the effects of treatment on a time-to-event outcome such as survival, or an odds ratio (OR) or risk ratio (RR) (relative risk) for a dichotomous outcome such as toxicity (Table 4). In the case of an aggregate data review, the effects measures will either be extracted or estimated from published results or obtained directly from trial investigators, whereas for IPD, they can be calculated directly from the re-analysis of the collated trial data, which also allows analytical assumptions to be checked and for the analysis methods to be kept consistent across trials.

Outcome type	**Description**	**Examples of outcome type**	**Examples of appropriate statistics**
Dichotomous outcome	Whether events do or do not happen	Mortality Adverse event	Risk Ratio Odds Ratio
Continuous outcome	Whether a disease or participant measure changes	Blood pressure Pain Weight loss	Mean difference Standardised mean difference
Time-to-event outcome	Whether events do or do not happen over a period of time	Survival Time to recurrence of disease Time to relief of symptoms	Hazard ratio

Table 4. Common outcome types and effect measures used in meta-analysis

How is a meta-analysis done?

Meta-analyses are typically described as either "one-stage" or "two-stage" (53). The "two-stage" approach is more commonly used and applicable to both aggregate data or IPD meta-analyses. In the first stage, an estimate of the treatment effect and its standard error (or variance) are obtained for each trial, and in the second stage, these are combined to obtain an overall estimate of effect, which is essentially a weighted average of the effects in each trial.

$$\text{Meta-analysis effect} = \frac{\text{Sum of (trial effect x w eight)}}{\text{Sum of weights}}$$

The weight applied to each trial determines how much influence each trial has on the meta-analysis effect estimate.

For IPD meta-analyses, although the two-stage approach remains common, so-called "one-stage" approaches(53) are being used increasingly(54) and typically involve analysing the data all together in a single

regression model, while also accounting for differences between trials. The model chosen will depend on the outcome(s) of interest (Table 4), for example, a logistic regression might be used for dichotomtous outcomes, a Cox regression for time-to-event or a linear regression for continuous outcomes. Theoretically, two-stage methods can be seen as special cases of one-stage models(55-57), and studies using both simulated(58, 59) and real data(58-61) have shown that the results obtained will usually be similar, irrespective of whichever is used. However, in particular scenarios, results from two-stage meta-analysis can be biased(58, 59, 62).

Visualising meta-analysis results

The results of two-stage meta-analyses are usually presented on forest plots. For example, the results of a meta-analysis of the effect of adding docetaxel to standard care on survival in metastatic hormone sensitive prostate cancer(63) are shown in Figure 2.

[Figure 2: Effect of the addition of docetaxel to standard of care on overall survival of men with metatstatic prostate cancer. Modified with permission from Elsevier (Lancet, 2016, 17, 243-256).]

In Figure 2, the x-axis shows the range of effect sizes, in this case HRs, with the line running through 1 (or 0 on log HR scale) representing no effect of docetaxel (equivalence). The HR estimates of the effect of docetaxel for individual trials are represented by squares, and the 95% confidence intervals by the horizontal lines on either side. The size of each square is directly proportional to the amount of information (in this case events) contributed by the trial results, and the associated confidence interval shows the level of uncertainty in the estimation of the HRs. Those HRs less than 1, i.e. lying to the left of the line of no effect, indicate a reduction in the risk of a death with docetaxel, and this is the case for all trials in Figure 2. However, results of individual trials are only conventionally significant (i.e. $p<0.05$) if the 95% confidence intervals do not cross the line of no effect. Thus, only two of all the individual trials in Figure 2 show significantly improved survival with docetaxel. Note that for positive outcomes, such as resection rates, the conventions would be reversed, and beneficial effects of treatment would be represented by estimates of effect lying to the *right* of the line of equivalence.

The overall estimate for the meta-analysis is usually displayed as a diamond, with the centre corresponding to the effect size, and the edges the 95% confidence interval. Again, a meta-analysis result is only conventionally statistically significant when the confidence interval does not overlap the line of equivalence. In Figure 2, the meta-analysis HR of 0.77 represents a 23% relative reduction in the risk of death with the addition of docetaxel, and is highly statistically significant(63) reflecting the gain in power from being able to include more patients

and events. Also, as clearly seen in Figure 2, the confidence interval for the meta-analysis result (95% CI=0.68-0.87) is much narrower than for any of the individual trials, reflecting the increased confidence in this estimate of effect.

Important to note is that as forest plots were designed for use with two-stage meta-analysis models, it remains a matter of debate how they might better be used in conjunction with one-stage models, which typically do not estimate treatment effects for the individual trials(38,54).

Measuring and accounting for heterogeneity

As described previously, effect estimates from individual trials are more likely to be similar if the populations, and particularly the interventions and comparisons of those trials are similar. Nevertheless, some variability in individual trial estimates is inevitable and expected, but where these are greater than expected by chance this is termed statistical heterogeneity(64). This is usually assessed using a chi-square test known as Cochran's Q(64). A p-value of less than 0.05 or, because of the low power of the test, less than 0.10, for this test is considered suggestive of meaningful variation between effect estimates. An alternative measure is the inconsistency statistic, I^2, which describes the variability in the trial effects estimates that arise from heterogeneity rather than by chance (sampling error) alone(64). The I^2 statistic is presented as a percentage, with figures close to 0% suggestive of low heterogeneity and those close to 100% high heterogeneity(65).

Fixed effect models for meta-analysis (66, 67) assume that each trial estimates the same underlying treatment effect, and that variation in effects between trials occur by chance alone. Trials in the meta-analysis are weighted directly by the amount of information they contribute to the meta-analysis. Larger trials, including more participants and with more events, have a larger weight than smaller trials and heterogeneity is not accounted for. Alternatively, a random effects model, such as that proposed by DerSimonian and Laird(68), assumes that treatment effects will be normally distributed around a mean effect. Trials are weighted by a combination of their size, or amount of information provided, and by the underlying variability between the studies as estimated by the model. Where there is low heterogeneity, as assessed by Cochran's Q or the I^2 statistic, the results obtained will be very similar, or indeed, identical, regardless of whether a fixed effect or random effects model is employed. Conversely, where there is evidence of heterogeneity, a fixed effect model may provide an estimate of effect that is too precise (confidence intervals are too narrow), not appropriately reflecting the uncertainty associated with the meta-analysis estimate of treatment effect. Although a random

effects model will reflect the uncertainty, having wider confidence intervals; smaller trials will potentially influence the meta-analysis as much as larger, theoretically more reliable, trials.

In a systematic review of the effect of post-operative radiotherapy on survival in patients with NSCLC(69) there was some evidence of heterogeneity in the effect on survival (I^2 = 40%, P = 0.08). However, results from the fixed effect (HR = 1.18, 95% CI= 1.07 to 1.31, p=0.0001) and random effects models (HR=1.17, 1.02-1.34, p=0.02) were similar, re-assuringly demonstrating that the results were robust to the choice of model. By contrast, in an IPD meta-analysis of neoadjuvant chemotherapy prior to radiotherapy for cervical cancer(70) there was substantial evidence of heterogeneity in the effect on survival (p=0.0003, I^2=62%), and the fixed effect meta-analysis suggested no clear benefit of neoadjuvant chemotherapy (HR=1.05, 95% CI=0.93-1.19, p=0.39). However, it is perhaps inappropriate to combine such variable results using a model that assumes the same effect across trials. Although the random effects model better reflects the uncertainty of the result, with a much wider confidence interval (HR=1.11, 95% CI=0.90-1.36, p=0.32) , the estimate of effect is driven by a number of the small trials with extreme results(71). In fact neither model provides an ideal synthesis of the trials results.

Exploring heterogeneity and robustness of results

The presence of heterogeneity makes it more likely that a meta-analysis result will be non-significant and/or otherwise unclear, which inevitably hampers interpretation. Therefore, it is important to explore whether variability in characteristics of either the trials or of the included participants (so called “effect modifiers”) may help to explain differences in treatment effects (71, 72). Trial-level effect modifiers can be explored by grouping trials according to the characteristic(s) of interest, and performing separate meta-analyses within each group. Results are then compared using a test for interaction (subgroup differences)(64). Such analyses were useful in showing that the effects of neoadjuvant chemotherapy in cervical cancer(70) varied by chemotherapy cycle length (test for interaction p=0.009), with a clear detriment from giving long-cycle neoadjuvant chemotherapy before radiotherapy (HR=1.25, 95% CI 1.07-1.46), but not short-cycle chemotherapy (HR=0.83, 95% CI 0.69-1.00), helping to explain the heterogeneity identified. Alternatively, meta-regression(64) can be used to explore whether individual trial treatment effects change in response to changes in, for example, treatment doses or regimens.

The investigation of how treatment effects vary by patient characteristics is important not only in exploring heterogeneity, but also in determining which participants are more likely to benefit or are least likely to be

harmed by particular treatments. However, using meta-regression of aggregate data is usually inadequate for this purpose because it only allows investigation of interactions between trial treatment effects and summaries of patient characteristics, such as mean age or proportion of females, which may not reflect the genuine relationship between the effects and the age or sex of an individual participant(73). Therefore, full investigation of potential patient-level treatment effect modifiers is a common motivation for collecting IPD(6, 74). For example, a systematic review and IPD meta-analysis showed that the benefits of tamoxifen or the survival of women with early breast cancer were limited to those with oestrogen receptor positive tumours(75). Such findings would not have been possible without access to IPD. However, some methods of analysing these interactions between patient characteristics and treatment effects are prone to ecological bias(73) so care must be taken to analyse these appropriately(76).

Another option for exploring variation in trial results is to conduct sensitivity analyses, whereby trials with particular characteristics are included (or excluded) from a meta-analysis, to determine their influence on the overall results. Candidates for sensitivity analyses might be trials with unusual designs, or with extreme results. However, it is possible that excluding trials in such a way can introduce bias, for example, if the analyses are driven by the results or trial effects vary for reasons other than those under consideration. Sensitivity analyses performed to exclude trials considered to be of poor quality(51), can also be probematic since quality judgments will always be somewhat subjective, and potentially unreliable if based on trial reports(47, 48).

Ideally, the possible or likely sources of variation in treatment effects should be anticipated at an early stage in the systematic review process, and strategies to investigate or mitigate the effects of such heterogeneity described in detail in the protocol statistical analysis plan in advance of any analyses being conducted. Even if it transpires that there is little or moderate statistical heterogeneity, such analyses remain useful in exploring the robustness of the results to the variability that is inherent across trials in any meta-analysis. Conversely, the results of interaction tests or sensitivity analyses conducted *post hoc* should be interpreted cautiously, and regarded as hypothesis-generating rather than as conclusive.

Interpreting meta-analysis results

When interpreting the results of a meta-analysis, examining just the direction of the effect and the significance level (p-value) is insufficient. There is a need to also consider:

- the size of the effect (difference from the no effect line), to assess if it is clinically worthwhile

- the strength of evidence for the effect (the amount of eligible trials and data included and the width of the confidence interval)
- the consistency of the results of individual trials (visual inspection and heterogeneity statistics, and whether any variation can be explained by trial or patient characteristics).

Revisiting the aggregate data review of pre-operative chemotherapy for NSCLC(28), the meta-analysis suggests an improvement in survival (HR=0.82, 95% CI=0.69-0.97), and results are consistent across trials (Test for heterogeneity; p=0.98). However, the p-value is just conventionally significant (p=0.022), and the upper limit of the confidence interval is close to 1, implying that the true effect may be very small. Perhaps most importantly, some key large trials, representing 25% of the randomised patients could not be included. As inclusion of these data these could potentially alter either the size or direction of effect, the results can only be regarded as promising rather than conclusive. The potential impact of missing trials on the results of a meta-anlaysis can be indicated by the use of funnel plots. However, as reporting biases are one of a number of potential causes of funnel plot asymmetry, the interpretation of these plots is not always indicative of missing trials. Furthermore, funnel plots are not appropriate in situations where fewer than 10 trials have been included in the meta-analysis, where very few trials have statistically significant results, or where there is considerable heterogeneity(77, 78).

For the aggregate data review of the effects of adding docetaxel to standard care on survival in metastatic prostate cancer(63) represented in Figure 2, we see that the HR is substantially smaller than 1, the confidence interval is narrow with an upper limit that is also much less than 1, and the effect estimate is highly statistically significant. Thus, the meta-analysis seems to provide reliable evidence of a large treatment effect. Heterogeneity across trial results is low (Test for heterogeneity; p=0.187; I^2=37.5%) and unaffected by the choice of a fixed or random effects model providing further reassurance as to the robustness of the results. Although the meta-analysis is based on only three of the five eligible trials, it still represents 93% of all patients randomised. Thus, incorporation of data from the remaining two trials would have little impact. Taken together, this provides compelling evidence that adding docetaxel to standard care improves the survival of men with metastatic prostate cancer(79).

Is meta-analysis always appropriate?

Whilst most systematic reviews that set out to assess the effects of a treatment intervention do include a meta-analysis, there are occasions when it may not be feasible to formally synthesize the results of trials. For

example, where data are sparse, either because very few trials have been conducted in the field of interest, or those that have are either still ongoing or have not reported results. In the latter circumstances, contact with the investigators can help establish if and when results may become available, allowing better planning of the timing of any meta-analysis. Alternatively, a request for the results or IPD can help make a meta-analysis feasible.

Another obvious circumstance in which meta-analysis may be inappropriate is where there is substantial heterogeneity across trial results that cannot be explained by differences in trial or patient characteristics. This could indicate that the included trials have assessed fundamentally different questions, such that combining them in meta-analysis could lead to erroneous or invalid conclusions.

A final consideration is whether it is appropriate to combine or include the results of trials (or trial analyses) that are clearly of "poor quality" or at high risk of bias in a meta-analysis at all, as it is likely that these biases will be further compounded. This was the approach taken in a systematic review of adjuvant chemotherapy for bladder cancer(80) in which there had been serious concerns raised about the quality of the design, conduct and analysis of some of the eligible trials. If it is the analysis or reporting, rather than the conduct, of the trial that is at fault, then requesting results based on the intention-to-treat principle (that is, including all randomised participants), for all outcomes of interest, or with longer follow-up may mitigate the potential influence of these biases on the meta-analysis results. Indeed, a systematic review and meta-analysis based on IPD from many of these same trials of adjuvant chemotherapy in bladder cancer(81) was able to assess and resolve most of the issues affecting the aggregate data review. If time or resources are limited, this could be done only for those trials judged to be most at risk of bias.

Multivariate meta-analysis

Meta-analyses discussed in the bulk of this chapter are used to estimate the effect of a treatment on a specific outcome. However, multivariate meta-analysis is often used to estimate the overall effects of multiple outcomes simultaneously. For instance, in a meta-analysis in the field of cancer, a reviewer might be interested in both progression-free survival and in overall survival. The simplest approach would be to perform two independent meta-analyses, one for each outcome. However, if only the most significant outcome from each trial is reported (outcome reporting bias) then those meta-analyses will be biased. Importantly, though, outcomes such as progression-free survival and overall survival are often highly correlated; and this correlation can be used in

multivariate meta-analysis to permit so-called *borrowing of strength* from the effect in one outcome to assist in estimating the treatment effect on the other.

Network meta-analysis

Network meta-analysis is a similar concept to multivariate meta-analysis, however, in a network meta-analysis, multiple treatments may be assessed simultaneously(82). Each included trial compares two or more eligible treatments and trials should be as similar to each other as possible in terms of the included patients or settings to avoid or limit variability. The extent of available data can be presented as a web-like diagram called a network, showing where there are direct comparisons of treatment available from within one or more trials; and the relative strength of such evidence (using weighting similar to that as described for fixed- and random-effects models). A meta-analysis model is then fitted that allows information to "flow" through the network and borrowing of strength to occur between treatment comparisons, as well as providing information on any missing treatment comparisons. One common application of such a model is to rank multiple treatments in order of their efficacy compared with a common reference treatment.

6. Structured presentation of results

Appropriately conducted systematic reviews provide a comprehensive and objective summary of all available evidence. However, to enable users to to assess the validity of the methodology used, the assessments made and the interpretation, structured and transparent reporting is vital. With this in mind, in 2005, a group of systematic reviewers, editors, clinicians and others developed the PRISMA statement, which was first published in 2009(3, 4). The aim of the PRISMA statement and accompanying checklist was to improve the reporting of systematic reviews and meta-analyses. Published alongside the original statement was a paper giving further explanation to review authors on good practice in reporting, using a series of examples for each item on the checklist(83). The checklist details 27 items that should be included in every systematic review and meta-analysis report as well as a flow diagram that shows the flow of information through the review process (Figure 1). The statement has now been endorsed and taken up by a number of organisations including the Cochrane Collaboration, Council of Science Editors and the World Association of Medical Editors, as well as several hundred journals that publish systematic reviews across healthcare. Since the launch of the original PRISMA statement, a number of extensions have been developed to facilitate the reporting of different types or systematic reviews (e.g. IPD meta-analyses(84)) as well as the reporting of abstracts(85) and protocols(86) of systematic reviews.

Impact of systematic reviews and meta-analyses

Clear and detailed reporting of results(4, 86) allows for ease of understanding and facilitates the use of systematic reviews and meta-analysis. Because they are generally considered to be the highest level of evidence, and often resolve uncertainty about a particular healthcare intervention, systematic reviews have a key role in informing evidence-based clinical guideline recommendations and thus influencing clinical practice. In particular, systematic reviews based on IPD can often help guideline developers to make more nuanced recommendations by targeting particular treatments to those individuals who might benefit the most(87). However, the potential impact of a systematic review can be further enhanced through timely publication. It should be noted that publication in medical journals alone may not maximise the potential impact of a systematic review, which may require making the publication freely available or engaging directly with guideline developers or patient advocacy organisations.

Systematic reviews and meta-analyses also have great potential to influence different stages of clinical research. Systematic reviews can inform the design(88), explain the rationale for(89-91), or directly influence new trials(90, 91). For example providing sufficient justification for the initiation of a new trial by identifying gaps in the existing evidence or informing the design and/or conduct of an ongoing trial by considering the external evidence. Because they often provide more detailed and reliable results, and a greater depth of understanding, well-conducted IPD meta-analyses thus have even greater potential to inform the design, conduct, analysis and reporting of new or ongoing clinical trials(92) (Figure 3). The collective approach of systematic reviews based on IPD can also speed up the design and launch of new trials initiated by members of the collaborative group(63). Systematic reviews can also help place the results of a clinical trial in the context of the results of other related trials and many journals now require this as part of the reporting standards for clinical trials(43).

[Figure 3: Impact of systematic reviews and meta-analysis throughout the clinical research process]

Summary

Systematic reviews and meta-analyses provide reliable and up to date evidence in oncology. Providing they are well designed, properly conducted, follow rigourous methodology, and are clearly reported, in accordance with appropriate guidelines, their results have the power and precision to guide clinical practice and policy. In addition, they can highlight where uncertainties about treatments persist, providing justification for further clinical research and informing future trials.

References

1. Green S, Higgins JPT, Alderson P, Clarke M, Mulrow CD, Oxman AD. Chapter 1: Introduction. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0: The Cochrane Collaboration, 2011.

2. Green S, Higgins JPT. Chapter 2: Preparing a Cochrane review. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0: The Cochrane Collaboration, 2011.

3. Preferred Reporting Items for Systematic Reviews and Meta-Analyses. http://prisma-statement.org/.
4. Moher D, Liberati A, Tetzlaff J, Altman D, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med*, 2009; 6(7), e1000097. doi:1000010.1001371/journal.pmed.1000097.

5. Stewart LA, Clarke MJ, on behalf of the Cochrane Working Party Group on Meta-analysis using Individual Patient Data. Practical methodology of meta-analyses (overviews) using updated individual patient data. *Statistics in Medicine*, 1995; 14, 2057-2079.

6. Stewart LA, Tierney JF. To IPD or Not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Evaluation and the Health Professions*, 2002; 25(1), 76-97.

7. Early Breast Cancer Trialists Collaborative Group. Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer. *New England Journal of Medicine*, 1988; 319(26), 1681-1692.

8. Meta-analysis Group in Cancer. Toxicity of fluorouracil in patients with advanced colorectal cancer: effect of administration scehdule and prognostic factors. *Journal of Clinical Oncology*, 1998; 16(11), 3537-3541.

9. Pignon JP, Bourhis J, Domenge C, Designé L, on behalf of the MACH-NC Collaborative Group. Chemotherapy added to locoregional treatment for head and neck squamous cell carcinoma: three meta-analyses of updated individual data. *Lancet*, 2000; 355(9208), 949-955.

10. Non-small Cell Lung Cancer Collaborative Group Chemotherapy in non-small cell lung cancer: a meta-analysis using updated data on individual patients from 52 randomised clinical trials. *BMJ*, 1995; 311, 899-909.

11. Duchateau L, Pignon J-P, Bijnens L, Bertin S, Bourhis J, Sylvester R. Individual patient-versus literature-based meta-analysis of survival data: Time to event and event rate at a particular time can make a difference, an example based on head and neck cancer. *Controlled Clinical Trials*, 2001; 22(5), 538-547.

12. Jeng GT, Scott JR, Burmeister LF. A comparison of meta-analytic results using literature versus individual patient data: paternal cell immunization for recurrent miscarriage. *Journal of the American Medical Association*, 1995; 274, 840-836.

13. Stewart LA, Parmar MKB. Meta-analysis of the literature or of individual patient data: is there a difference? *Lancet*, 1993; 341, 418-422.

14. Collaboration EHT. Value of updating a systematic review in surgery using individual patient data. *British Journal of Surgery*, 2004; 91, 495-499.

15. O'Connor D, Green S, Higgins JPT. Chapter 5: Defining the review question and developing criteria for including studies. In: Higgins JPT, Green S, eds. Cochrane Handbook of Systematic Reviews of Intervention. Version 5.1.0: The Cochrane Collaboration, 2011.

16. Schünemann HJ, Oxman AD, Vist GE, Higgins JPT, Deeks JJ, Glasziou P, et al. Chapter 12: Interpreting results and drawing conclusions. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0: The Cochrane Collaboration, 2011.

17. Altman DG. Systematic reviews of evaluations of prognostic variables. *BMJ*, 2001; 323 224-228.

18. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM, Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. *Annals of Internal Medicine*, 2008; 149(12), 889-897.

19. Debray TP, Riley RD, Rovers MM, Reitsma JB, Moons KG, Cochrane IPD Meta-analysis Methods group. Individual participant data (IPD) meta-analyses of diagnostic and prognostic modeling studies: guidance on their use. *PLoS Med*, 2015; 12(10), e1001886.

20. Moher D. The problem of duplicate systematic reviews. *BMJ*, 2013; 347, f5040 doi: 5010.1136/bmj.f5040.

21. Centre for Reviews and Dissemination. Systematic Reviews: CRD's guidance for undertaking reviews in health care. York, UK: CRD, University of York, 2009 2009.

22. Cochrane Database of Systematic Reviews http://www.thecochranelibrary.com

23. PROSPERO: International prospective register of systematic reviews http://www.crd.york.ac.uk/PROSPERO/prospero.asp.

24. Le Pechoux C, Burdett S, Auperin A. Individual patient data (IPD) meta-analysis (MA) of chemotherapy (CT) in locally advanced non-small cell lung cancer (NSCLC). *Journal of Thoracic Oncology*, 2008; 3(4, Supplement 1), S20, 35IN.

25. NSCLC Meta-analyses Collaborative Group. Chemotherapy in addition to supportive care improves survival in advanced non-small-cell lung cancer: a systematic review and meta-analysis of individual patient data from 16 randomized controlled trials. *Journal of Clinical Oncology*, 2008; 26(28), 4617-4625.

26. NSCLC Meta-analysis Collaborative Group. Preoperative chemotherapy for non-small cell lung cancer: a systematic review and meta-analysis of individual participant data. *Lancet*, 2014; 383, 1561-1571.

27. NSCLC Meta-analyses Collaborative Group. Adjuvant chemotherapy, with or without postoperative radiotherapy, in operable non-small cell lung cancer: two meta-analyses of individual patient data. *Lancet*, 2010; 375(9722), 1267-1277.

28. Burdett S, Stewart L, Rydzewska L. A systematic review and meta-analysis of the literature: chemotherapy and surgery versus surgery alone in non-small cell lung cancer. *Journal of Thoracic Oncology*, 2006; 1(7), 611-621.

29. Cochrane Central Register of Controlled Trials (CENTRAL). Available from: http://www.cochranelibrary.com/about/central-landing-page.html.

30. Lefebvre C, Manheimer E, Glanville J. Chapter 6: Searching for studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0: The Cochrane Collaboration, 2011.

31. Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, et al. Epidemiology and reporting charateristics of systematics reviews of biomedical research: A cross-sectional study. *PLoS Med*, 2016; 13(5), e1002028.

32. Ahmed I, Sutton AJ, Riley RD. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. *BMJ*, 2012; 344, d7762.

33. Dickersin K. Publication bias: recognising the problem, understanding its origins and scope, and preventing harm. In: Rothstein H, Sutton A, Borenstein M, eds. Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments. Chichester: John Wiley & Sons Ltd, 2005: 261-286.

34. Hopewell S, Clarke M, Stewart L, Tierney J. Time to publication for results of clinical trials. *Cochrane Database of Systematic Reviews*, 2007; (2), Art. No.: MR000011. DOI: 000010.001002/14651858.MR14000011.pub14651852.

35. von Elm E, Poglia G, Walder B, Tramèr MR. Different patterns of duplicate publication: an analysis of articles used in systematic reviews. *JAMA*, 2004; 291(8), 974-980.

36. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet*, 1991; 337(8746), 867-872.

37. Gøtzsche PC. Reference bias in reports of drug trials. *BMJ*, 1987; 295, 654–656.

38. Juni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *International Journal of Epidemiology*, 2002; 31(1), 115-123.

39. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ*, 2010; 340, c365.

40. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. : The Cochrane Collaboration, 2011.

41. Scherer RW, Langenberg P, von Elm E. Full publication of results initially presented in abstracts. *Cochrane Database of Systematic Reviews*, 2007; (Issue 2), MR000005.

42. Altman DG. Better reporting of randomised controlled trials: the CONSORT statement. *BMJ*, 1996; 313, 570-571.

43. Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, et al. CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet*, 2008; 371(9609), 281-283.

44. Stewart L, Tierney J, Burdett S. Do systematic reviews based on individual patient data offer a means of circumventing biases associated with trial publications? In: Rothstein H, Sutton A, Borenstein M, eds. Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments. Chichester: John Wiley & Sons, 2005: 261-286.

45. Tierney JF, Vale CL, Riley R, Tudur Smith C, Stewart LA, Clarke M, et al. Individual participant data (IPD) meta-analyses of randomised controlled trials: Guidance on their use. *PLoS Med*, 2015; 12(7), e1001855.

46. Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, 2011; 343, d5928.

47. Vale CL, Tierney JF, Burdett S. Can trial quality be reliably assessed from published reports of cancer trials: evaluation of risk of bias assessments in systematic reviews. *BMJ*, 2013; 346, f1798.

48. Mhaskar R, Djulbegovic B, Magazin A, Soares HP, Kumar A. Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. *Journal of Clinical Epidemiology*, 2012; 65(6), 602-609.

49. Hartling L, Bond K, Vandermeer B, Seida J, Dryden DM, Rowe BH. Applying the risk of bias tool in a systematic review of combination long-acting beta-agonists and inhaled corticosteroids for persistent asthma. *PLoS One*, 2011; 6(2), e17242.

50. Jordan VM, Lensen SF, Farquhar C. There were large discrepancies in risk of bias tool judgements when a randomized controlled trial appeared in more than one systematic review. *Journal of Clinical Epidemiology*, 2016

51. Sterne JAC, Egger M, Moher D. Chapter 10: Addressing reporting biases. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Intervention. Version 5.1.0: The Cochrane Collaboration, 2011

52. Jorgensen L, Paludan-Muller AS, Laursen DR, Savovic J, Boutron I, Sterne JA, et al. Evaluation of the Cochrane tool for assessing risk of bias in randomized clinical trials: overview of published comments and analysis of user practice in Cochrane and non-Cochrane reviews. *Syst Rev*, 2016; 5(80)

53. Simmonds MC, Higgins JPT, Stewart LA, Tierney JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomised trials - a review of methods used in practice. *Clinical Trials*, 2005; 2(3), 209-217.

54. Simmonds M, Stewart G, Stewart L. A decade of individual participant data meta-analyses: A review of current practice. *Contemp Clin Trials*, 2015

55. Olkin I, Sampson A. Comparison of meta-analysis versus analysis of variance of individual patient data. *Biometrics*, 1998; 54, 317-322.

56. Mathew T, Nordstrom K. On the equivalence of meta-analysis using literature and using individual patient data. *Biometrics*, 1999; 55(4), 1221-1223.

57. Mathew T, Nordstrom K. Comparison of one-step and two-step meta-analysis models using individual patient data. *Biom J*, 2010; 52(2), 271-287.

58. Tudur Smith C, Williamson PR. A comparison of methods for fixed effects meta-analysis of individual patient data with time to event outcomes. *Clin Trials*, 2007; 4(6), 621-630.

59. Bowden J, Tierney JF, Simmonds M, Copas AJ. Individual patient data meta-analysis of time-to-event outcomes: one-stage versus two-stage approaches for estimating the hazard ratio under a random effects model. *Res Synth Methods*, 2011; 2(3), 150-162.

60. Stewart GB, Altman DG, Askie LM, Duley L, Simmonds MC, Stewart LA. Statistical analysis of individual participant data meta-analyses: a comparison of methods and recommendations for practice. *PLoS One*, 2012; 7(10), e46042.

61. Koopman L, van der Heijden GJ, Hoes AW, Grobbee DE, Rovers MM. Empirical comparison of subgroup effects in conventional and individual patient data meta-analyses. *International Journal of Technology Assessment in Health Care*, 2008; 24(3), 358-361.

62. Stijnen T, Hamza TH, Ozdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Statistics in Medicine*, 2010; 29(29), 3046-3067.

63. Vale CL, Burdett S, Rydzewska LH, Albiges L, Clarke NW, Fisher D, et al. Addition of docetaxel or bisphosphonates to standard of care in men with localised or metastatic, hormone-sensitive prostate cancer: a systematic review and meta-analyses of aggregate data. *Lancet Oncol*, 2016; 17(2), 243-256.

64. Deeks JJ, Higgins JPT, Altman DG. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0: The Cochrane Collaboration, 2011.

65. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*, 2003; 327, 557-560.

66. Yusuf S, Peto R, J.A. L, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Progress in Cardiovascular Diseases*, 1985; 27, 335-371.

67. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 1959; 22(4), 719-748.

68. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials*, 1986; 7, 177-188.

69. PORT Meta-analysis Trialists Group. Postoperative radiotherapy for non-small cell lung cancer. *Cochrane Database of Systematic Reviews*, 2005; (2), Art. No.: CD002142. DOI: 002110.001002/14651858.CD14002142.pub14651852.

70. Neoadjuvant Chemotherapy for Cervix Cancer Meta-analysis (NACCCMA) Collaboration. Neoadjuvant chemotherapy for locally advanced cervical cancer: A systematic review and meta-analysis of individual patient data from 21 randomised trials. *European Journal of Cancer*, 2003; 39(17), 2470-2486.

71. Bowden J, Tierney JF, Copas AJ, Burdett S. Quantifying, displaying and accounting for heterogeneity in the meta-analysis of RCTs using standard and generalised Q statistics. *BMC Med Res Methodol*, 2011; 11, 41.

72. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 2002; 21, 1539-1558.

73. Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI, Anti-Lymphocyte Antibody Induction Therapy Study G. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Statistics in Medicine*, 2002; 21(3), 371-387.

74. Stewart LA, Tierney JF, Clarke M, on behalf of the Cochrane Individual Patient Data Meta-analysis Methods Group. Reviews of individual patient data. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 510 (updated March 2011). Chichester, UK: Wiley-Blackwell, 2011: 547-558.

75. Early Breast Cancer Trialists' Collaborative G, Clarke M, Coates AS, Darby SC, Davies C, Gelber RD, et al. Adjuvant chemotherapy in oestrogen-receptor-poor breast cancer: patient-level meta-analysis of randomised trials. *Lancet*, 2008; 371(9606), 29-40.

76. Fisher DJ, Copas AJ, Tierney JF, Parmar MKB. A critical review of methods for the assessment of patient-level interactions in individual patient data (IPD) meta-analysis of randomised trials, and guidance for practitioners. *Journal of Clinical Epidemiology*, 2011; 64, 949-967.

77. Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, 2011; 343, d4002.

78. Ioannidis JP, Trikalinos TA. The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. *CMAJ*, 2007; 176(8), 1091-1096.

79. National Health Service England. Clinical commissioning policy statement: docetaxel in combination with androgen deprivation therapy for the treatment of hormone naïve metastatic prostate cancer. Jan 20, 2016. Available from: https://www.england.nhs.uk/wp-content/uploads/2016/01/b15psa-docetaxel

80. Sylvester R, Sternberg C. The role of adjuvant combination chemotherapy after cystectomy in locally advanced bladder cancer: what we do not know and why. *Annals of Oncology*, 2000; 11, 851-856.

81. Advanced Bladder Cancer (ABC) Meta-analysis Collaboration. Adjuvant chemotherapy in invasive bladder cancer: a systematic review and meta-analysis of individual patient data. *European Urology*, 2005; 48(2), 189-201.

82. Cipriani A, Higgins JP, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. *Annals of Internal Medicine*, 2013; 159(2), 130-137.

83. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*, 2009; 339, b2700.

84. Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Individual Participant Data: The PRISMA-IPD Statement. *JAMA*, 2015; 313(16), 1657-1665.

85. Beller EM, Glasziou PP, Altman DG, Hopewell S, Bastian H, Chalmers I, et al. PRISMA for Abstracts: Reporting Systematic Reviews in Journal and Conference Abstracts. *PLoS Med*, 2013; 10(4), e1001419. doi:1001410.1001371/journal.pmed.1001419.

86. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*, 2015; 4, 1.

87. Vale CL, Rydzewska LHM, Rovers MM, Emberson JR, Gueyffier F, Stewart LA. Uptake of systematic reviews and meta-analyses based on individual participant data in clinical practice guidelines: descriptive study. *BMJ*, 2015; 350, h1088.

88. Clarke M. Doing new research? Don't forget the old. *PLoS Med*, 2004; 1(2), e35.

89. Cooper NJ, Jones DR, Sutton AJ. The use of systematic reviews when designing studies. *Clin Trials*, 2005; 2(3), 260-264.

90. Clarke M, Hopewell S, Chalmers I. Clinical trials should begin and end with systematic reviews of relevant evidence: 12 years and waiting. *Lancet*, 2010; 376(9734), 20-21.

91. Jones AP, Conroy E, Williamson PR, Clarke M, Gamble C. The use of systematic reviews in the planning, design and conduct of randomised trials: a retrospective cohort of NIHR HTA funded trials. *BMC Med Res Methodol*, 2013; 13, 50.

92. Tierney JF, Pignon J-P, Gueffyier F, Clarke M, Askie L, Vale CL, et al. How individual participant data meta-analyses can influence trial design and conduct *Journal of Clinical Epidemiology*, 2015; 68(11), 1325-1335.