ELSEVIER

Journal of Clinical Epidemiology 64 (2011) 949–967

Journal of Clinical Epidemiology

A critical review of methods for the assessment of patient-level interactions in individual participant data meta-analysis of randomized trials, and guidance for practitioners

D.J. Fisher*, A.J. Copas, J.F. Tierney, M.K.B. Parmar

Medical Research Council Clinical Trials Unit, London NW1 2DA, UK

Accepted 24 November 2010

Abstract

Objective: Treatments may be more effective in some patients than others, and individual participant data (IPD) meta-analysis of randomized trials provides perhaps the best method of investigating treatment-covariate interactions. Various methods are used; we provide a comprehensive critique and develop guidance on method selection.

Study Design and Setting: We searched MEDLINE to identify all frequentist methods and appraised them for simplicity, risk of bias, and power. IPD data sets were reanalyzed.

Results: Four methodological categories were identified: PWT: pooling of within-trial covariate interactions; OSM: ''one-stage'' model with a treatment-covariate interaction term; TDCS: testing for difference between covariate subgroups in their pooled treatment effects; and CWA: combining PWT with meta-regression. Distinguishing across- and within-trial information is important, as the former may be subject to ecological bias. A strategy is proposed for method selection in different circumstances; PWT or CWA are natural first steps. The OSM method allows for more complex analyses; TDCS should be avoided. Our reanalysis shows that different methods can lead to substantively different findings.

Conclusion: The choice of method for investigating interactions in IPD meta-analysis is driven mainly by whether across-trial information is considered for inclusion, a decision, which depends on balancing possible improvement in power with an increased risk of bias.

Keywords: Meta-analysis; IPD; RCT; Interaction; Subgroup; Methodology

1. Introduction

The main aim of meta-analysis is to combine estimates of a particular effect across independent studies—frequently the effect of a treatment assessed in randomized controlled trials (RCTs)—to obtain a summary estimate of effect and standard error. Because of the inevitable differences in trial design and population, some variation in effect is to be expected. The widely used random-effect model of DerSimonian and Laird [1] was proposed to account for this. There are, however, statistical and clinical benefits to be had from investigating *how* factors influence the treatment effect—in other words, the nature of interactions between such factors and treatment. In meta-analysis of aggregate data (from publications or supplied by investigators), trial-level heterogeneity is often explored using subgroup analyses or meta-regression [2]. However, to assess *patient*-level heterogeneity in this framework, individual patients must be assigned trial-level average values, which is inefficient and risks bias because of the ''ecological fallacy'' [3].

We may overcome some of these problems by obtaining individual participant data (IPD) for each trial. Treatment, outcome, and covariate measurements are then known for each patient within each trial, representing the best opportunity to explore treatment-covariate interactions. More powerful and flexible analyses can be done and, in particular, patient-level heterogeneity in the treatment effect can be separated from trial-level heterogeneity and investigated directly. Various methods of accomplishing this are used in practice and, as previously observed [4], most IPD literature concentrates either on documenting these methods or on proposing

* Corresponding author. Tel.: +44-207-670-4646; fax: +44-207-670-4949.

E-mail address: df@ctu.mrc.ac.uk (D.J. Fisher).

0895-4356/$ - see front matter
doi: 10.1016/j.jclinepi.2010.11.016

What is new?

Key findings

- Methodology for analyzing treatment by patient-level covariate interactions in individual participant data (IPD) meta-analysis is not yet fully established, and practical guidance is limited.
- A common approach, that estimates treatment effects within covariate subgroups and tests the differences between these subgroup estimates, is shown to be at risk of bias and should be discouraged.
- An approach that estimates a ''within-trial'' patient-level interaction only is recommended. ''Across-trial'' information is at risk of bias, and if it is to be used to supplement the ''within-trial'' information it should be done with caution.

What this adds to what was known?

- Although methodology in this area is discussed elsewhere, this is generally of a technical nature without a clear presentation of each distinct methodological approach.
- Clear guidance is given to help systematic reviewers assess whether a treatment effect varies across patient-level covariates using appropriate methodology.

What is the implication, what should change now?

- Estimating treatment effects within covariate subgroups and testing the differences between these subgroup estimates should not be used further. Instead, the guidance given in this article should be followed to select an appropriate method.
- If the data are unsuitable for assessing patient-level covariate interactions (e.g., sparse categories or large heterogeneity across trials), reviewers should state this rather than attempt to draw conclusions, which would likely be unreliable.

new statistically advanced techniques (e.g., [5]). A recent article by Thompson et al. [6] discusses most of these methods for IPD meta-analysis of time-to-event data, but we focus specifically on treatment-covariate interaction estimation in trials, describing the relevant issues in greater detail and including in our critique one further commonly used method.

In this article, we identify the principal published methods to assess treatment-covariate interactions, and critically appraise them to summarize their advantages and disadvantages. We then develop guidance on which approach might be used in different circumstances. We apply the methods to IPD data sets comparing interventions for cancer to illustrate their use in practice, before making some concluding remarks. Because these data sets all relate to time-to-event outcome measures, this article will focus on such outcome measures while also providing broader guidance.

2. Methods in the literature to analyze treatment-covariate interactions

A literature search of Medline (1966–2009) was performed to identify the main approaches for analyzing patient-level treatment-covariate interactions, both proposed in theory and used in practice, and also previous methodological reviews (see Appendix A). However, our search was not planned to quantify the number of reviews using these approaches in practice. We limited ourselves to frequentist approaches as these are most commonly used by systematic reviewers. Four categories of methods were identified, all of which may be fitted using standard statistical software unless otherwise stated. For simplicity, in this section we consider two competing treatments and a patient covariate that is binary, continuous, or ordered categorical. We assume that if the covariate is not binary then the interaction can be adequately represented by a linear interaction term. A mathematical presentation of the four methods is provided in Appendix B.

2.1. Pooling of within-trial covariate interactions (PWT)

This approach is based on the familiar weighted-average meta-analysis of aggregate data, but here we pool treatment-covariate interaction effects instead of main study effects [7]. Interactions are estimated independently within each trial using regression models or contingency tables, and are pooled using inverse-variance meta-analysis. Like the equivalent technique for main effects, this method can be described as ''*two-stage*'' [8] in that two separate sets of calculations must be carried out, the second using the output of the first.

2.2. "One-stage" model with a covariate interaction term (OSM)

An alternative approach is to estimate the treatment-covariate interaction using all available data in a single model containing terms representing trial membership, treatment, covariate, and treatment-covariate interaction [9], as detailed in Appendix B. Alternatively, the interaction term may be replaced by two independent terms, representing across- and within-trial interaction effects [5]. The treatment effect may be fixed or random, but the latter is computationally difficult for time-to-event data (see next section).

2.3. Testing for treatment effect differences across covariate subgroups (TDCS)

This next method, commonly applied (e.g., [10–14]), arose as a simple way of analyzing the effect of *categorical* covariates on the treatment effect, and indeed cannot be used with continuous covariates. This is a two-stage approach in which inverse-variance pooled treatment effects are first estimated within each patient subgroup, and secondly tested for differences using a trend line or ANOVA. Both the within-subgroup treatment estimates and the resulting interaction effect are estimated almost exclusively using fixed effects in practice, although this is not imperative. Interaction and trend in treatment effect across levels of a patient-level categorical covariate are tested in a similar manner to the partitioned Q test for subgroups used in aggregate-data meta-analyses [15].

2.4. "Manually" combining separately calculated within- and across-trial effects (CWA)

The PWT method estimates a "within-trial" effect. In contrast, a meta-regression fits a trend line through trial-level treatment effects and covariate means, and estimates an independent "across-trial" effect. The possibility suggests itself, therefore, to apply both methods and combine the two interaction effect estimates "manually" so as to use the maximum amount of information in the data [16]. To do this, an inverse-variance meta-analysis with either fixed or random effects might be used. Alternatively, the analyst may weight the two estimates according to the perceived quality of each (see next section).

3. Critique of the methods

We next critique these methods with respect to simplicity, efficiency, and potential for bias, with particular emphasis on whether across-trial evidence of covariate interaction should be included. The evidence for an interaction from within trials, for example using PWT, may be considered to be of higher quality because within each trial the patients with different values of covariate are recruited and managed within the confines of a single trial protocol. By contrast, the apparent evidence for a patient-level treatment-covariate interaction obtained across trials, as in a meta-regression, is highly susceptible to ecological bias, because trials may differ in many respects other than simply the distribution of that covariate [17]. Whether or not to include such "across-trial" evidence will be discussed later. Although we have limited ourselves to frequentist approaches, any of the methods can also be used in a Bayesian framework—in fact, doing so offers some advantages such as a more straightforward estimation of random effects. Our critique, summarized in Table 1, concentrates on methods as currently available from mainstream statistical packages, rather than their theoretical potential.

3.1. Estimating interaction effects using within-trial information only

Arguably the simplest method conceptually and practically is PWT. As this method uses within-trial information only, trials with little variation in the covariate contribute only modest information to the estimate. Further, where the covariate is categorical, trials with data in only one subgroup must be discarded.

By contrast, the most computationally intensive and potentially complex approach reviewed here is the use of OSM. These models are sometimes used in epidemiological studies where the data have a multilevel structure, but are currently not routinely used for meta-analysis. They allow much greater flexibility and maximize the power available in the data, and could be considered as a "gold standard" because two-stage models can be obtained from them as special cases [5]. OSM permits multiple parameters to be estimated simultaneously, allowing the possibility of performing complex analyses and testing a wider range of hypotheses. We can parameterize the model to estimate the within-trial interaction effect specifically (in its more basic form, the model instead estimates a single interaction coefficient equivalent to a weighted average of within- and across-trial effects). However, there are significant computational challenges in the fitting of random treatment effects to time-to-event outcome data [5], and furthermore, with very large data sets the computational time is likely to be greater than for PWT [6].

Estimating trial heterogeneity can cause problems with convergence, and where such heterogeneity exists, the failure of fixed-effect OSM to represent it can lead to inappropriate standard errors. Generally speaking, such models require greater care to be taken to avoid misrepresentation of the data, and are more likely to require extensive statistical expertize. Some reviewers have stated a preference for OSM on the basis that use of two-stage methodology negates some of the advantages that IPD provides [18], although theoretically there is no reason for the power of OSM to significantly exceed that of PWT if across-trial information is not used [5].

3.2. Including across-trial information

The use of across-trial information is controversial and, given that within-trial interaction effects are easily estimated in isolation (such as with PWT), many analysts may never use it. However, in circumstances in which more power is desired and the increased risk of bias is considered acceptable, for example if the analysis is hypothesis generating, the addition of across-trial information is a possible solution. It can be shown that, in circumstances where PWT has low power, a meta-regression (across-trial information only) can provide greater power [19]. This is the motivation behind the CWA method, in which the PWT estimate and meta-regression slope are examined and

Table 1
Summary table of methods discussed in this article

Method	PWT	OSM	TDCS	CWA
Description	Pooling of within-trial patient-level interactions	One-stage model with covariate interaction term, and either fixed or random treatment effect.	Calculating pooled treatment effects within covariate subgroups and testing for differences between these.	PWT combined with a meta-regression.
Within-trial, across-trial information or both?	Within-trial information only	Within- and across-trial effects estimated separately; can be combined.	Implicit combination of within- and across-trial information, which cannot be separated.	Within- and across-trial effects estimated separately; can be combined.
Difficulty of use	Low. Straightforward to apply	High for time-to-event outcomes with heterogeneity between trials, as random-effect models not well developed. Moderate for other outcome types or if random effects not required. Requires statistical modeling expertize and statistical software such as Stata.	Low, unless random effects are used, in which case moderate.	Low. Although there are three steps (PWT; trial-level meta-regression; calculation of weighted average) they are each straightforward.
Key subjective decisions to be made	None	There are relatively more distributional assumptions to be made, which may be partly subjective. Need to decide if across-trial information will be considered and if so the criteria for inclusion.	None, as the combination of across- and within-trial information is implicit.	Need to decide if across-trial information will be considered and if so the criteria for inclusion. Also, need to decide the relative weighting of the two if estimates combined.
Bias and power	Bias is minimized but power may be low, particularly if variation is largely between rather than within trials.	Increased complexity has potential for bias from model misspecification; but assuming model is a good fit, power should be maximized. If across-trial information is included then power may increase further but with a corresponding increased risk of bias.	Power is moderate, because all information is used, but may not be maximized because not all correlation is accounted for. The risk of bias is high because of lack of knowledge of individual across- and within-trial estimates, and standard errors for the interaction effect may be inappropriate.	If across-trial information is discarded, power and risk of bias are identical to PWT. Otherwise, inclusion of across-trial information may increase power but at an increased risk of bias.

Abbreviations: PWT, pooling of within-trial covariate interaction; OSM, ''one-stage'' model with a covariate interaction term; TDCS, testing for treatment effect differences across covariate subgroups; CWA, ''manually'' combining separately calculated within- and across-trial effects.

potentially combined by means of a further weighted average. This approach has been discussed elsewhere for meta-analysis of randomized trials [16] or epidemiological data [6], and simulation has shown that CWA can indeed increase power relative to PWT, especially where data is scarce [20]. However, PWT and CWA often give markedly differing estimates of effect [21] because of ecological bias. Consequently, most analysts prefer to use within-trial information alone [7,16], although some consider that combining evidence is not unreasonable [20] because the allocation to the patient covariate is not randomized (excepting factorial designs where another treatment is considered as the covariate), and so even within-trial information may be somewhat biased. A similar practice is sometimes used in meta-analysis of main effects, where cohort or case-control studies can provide evidence in circumstances where there are few, if any, randomized trials. The CWA approach gives the analyst the freedom to study the two effects separately and, further, the potential to choose the weightings in the final combination step, allowing the potential ''down weighting'' of across-trial evidence to reflect its susceptibility to bias. This is analogous with the

approach sometimes used in meta-analyses of observational studies to deal with lower quality studies, although it is of course open to bias from the analyst if the weighting is not specified in advance. In any case, the selection of the weights is inevitably subjective. Because the use of across-trial information is controversial, it should be decided a priori whether to consider it for inclusion in the analysis, and the reasons behind the decision should be critically discussed as part of the analysis.

In addition to the CWA approach, there are further options available to the reviewer. Firstly, OSM can be parameterized such that within-trial and across-trial effects are estimated simultaneously by a single combined interaction term (see Appendix B). This is a simpler model to fit than the separate within- and across-trial parameterization, and may also have greater power. However, the challenges with OSM described above still remain [6].

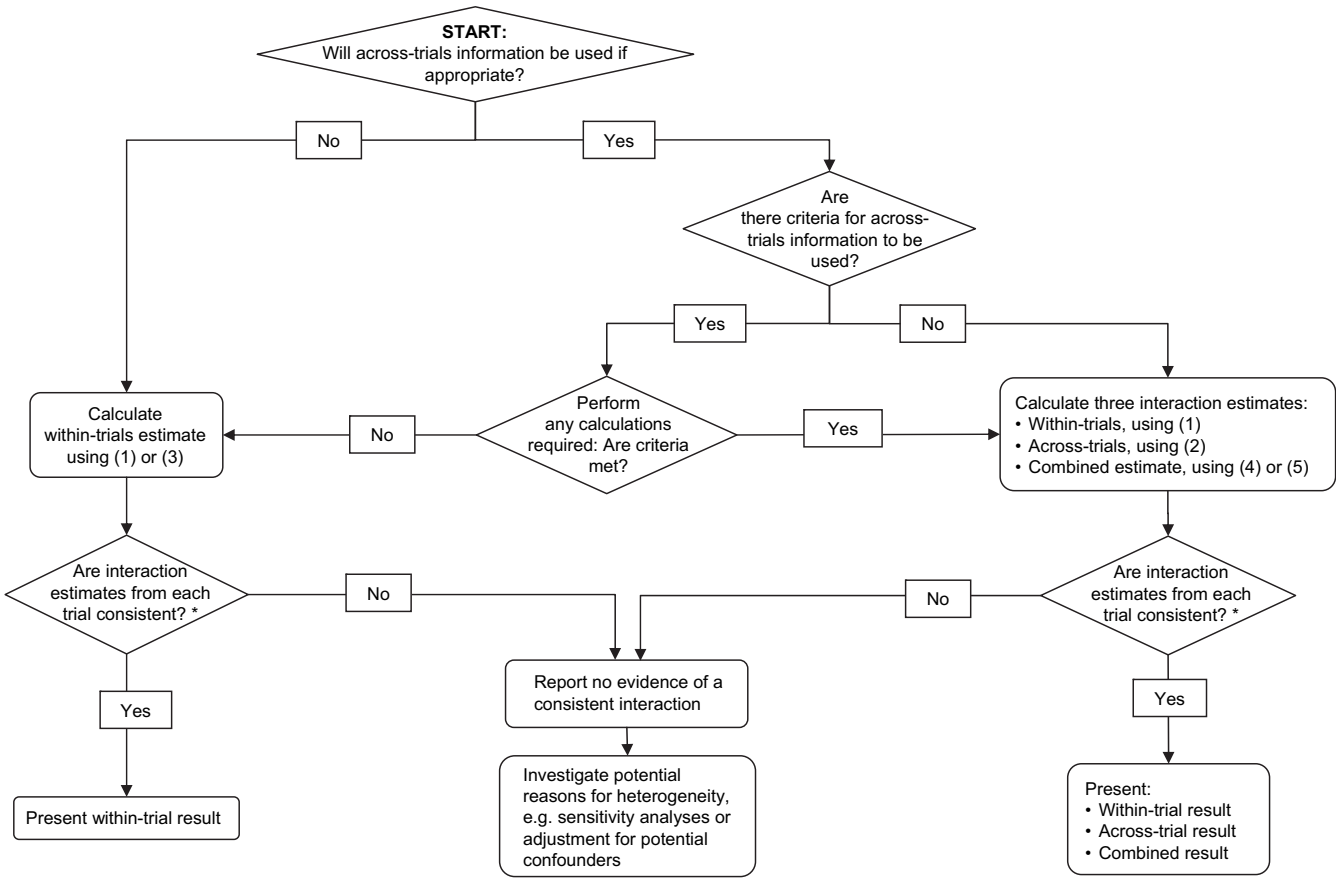
The TDCS method might appear to be an attractive way to analyze interactions, particularly with categorical covariates. It is simple to implement in its most common form, which makes use of the Peto odds ratio method [22] or the ''log-rank'' modification to this for time-to-event data [8], neither of which rely on iterative procedures and require no specialist software. These approximate the maximum likelihood estimates of the odds ratio and Cox hazard ratio (HR), respectively [22–24]. There are, however, serious drawbacks. Measuring within-subgroup effects implicitly involves combining patient data across trials; therefore, some across-trial information concerning the interaction must be used. But trials can contribute patients to more than one subgroup, so some within-trial information must also be used. Hence, an indeterminate mixture of within- and across-trial information will be present, with no way to recover estimates of the separate effects or therefore assess the risk of bias from the inclusion of across-trial information [16,19], in contrast to the other methods in this section. Further, within-trial correlations arising from trials supplying data to more than one covariate subgroup will not be accounted for, leading to a loss of efficiency and inappropriate standard errors. Lastly, heterogeneity in treatment effects across trials is usually ignored. These problems could partly be addressed by adding complexity, for instance by adding random effects to the within-subgroup meta-analyses, the across-subgroup trend, or both, as appropriate. However, these suggestions have not been formally tested, and such added complexity negates the original simplicity of the approach. We therefore do not recommend the TDCS method for formal testing of patient-level treatment-covariate interactions.

In conclusion, then, the choice of approach is governed chiefly by whether the use of across-trial information is considered and secondly by the resources available to the analyst together with the complexity of the modeling required. This latter may vary from estimation of a simple scalar interaction term, representing a solely relative effect, through to that of an absolute, nonlinear interaction effect adjusted for other covariates. We therefore propose a strategy for method selection.

4. A proposed strategy for method selection

Our proposed strategy is illustrated in Fig. 1. Firstly, an a priori decision should be made as to whether to consider including across-trial information. If within-trial information only is to be used, for example if a prior hypothesis is being tested and minimizing bias is of primary importance, we suggest the PWT method as it is straightforward to implement and interpret while having a low risk of bias. The choice of whether to fit fixed or random effects to the interaction is subject to the same arguments as discussed elsewhere for main (treatment) effects [25]. In particular, we caution against the automatic use of random effects where there is evidence of heterogeneity across interaction effects (e.g., a large I^2 or an obvious outlier). In such situations further investigation may be required, such as a check for model misspecification by inspecting the form of the interaction within each trial separately. Another possible source of heterogeneity in the interaction is confounding: if a second known covariate also interacts with treatment, and if the association between these two covariates differs across trials, then the original interaction may be altered. This may be solved through adjustment for the second covariate, if known and available. This is simply performed both for PWT (where additional terms are added to the model for each trial) and for OSM. If heterogeneity remains, it is probably best to conclude that the interaction is inconsistent across trials.

If the inclusion of across-trial information is being considered, we recommend two possible approaches. Firstly, estimates of the within trials, across trials, and combined effects could be presented, for which the CWA method is a convenient choice. Graphical presentation is greatly beneficial here (see next section), and Riley et al. [16] demonstrate the use of plots to informally judge the possible degree of ecological bias. Secondly, formal criteria could be specified for the inclusion of across-trial information, whereby the combined and within-trial estimates would be presented if the criteria are met, and the within-trial estimate alone if not. We do not propose any specific criteria of our own, although we note that clearly the risk of appreciable bias is lower when the within- and across-trial effect estimates are similar. Formal tests of difference between estimates (i.e., of ecological bias) may be used to quantify this, although statistical significance is unlikely to be reached, because of imprecision, even if the underlying bias is substantial. Simmonds and Higgins [19] have developed statistics to quantify the relative power of within- vs. across-trial interaction information, but not specifically to reflect ecological bias. Whichever approach is taken the within-trial effect estimate should be presented, and if the combined estimate is also presented then a suitable note



Key: (1) PWT; (2) meta-regression; (3) OSM within-only; (4) OSM combined; (5) CWA

* e.g. non-significant heterogeneity; fixed- and random-effect results consistent

Fig. 1. Flow diagram of recommended approaches.

of caution regarding ecological bias should be included. If criteria are used for the inclusion of the across-trial information, these need to be clearly stated and pursued objectively.

The advice given so far is aimed at reviewers who wish to avoid complex statistical modeling, especially those analyzing time-to-event data where extra methodological difficulties exist. However for those with suitable expertize and resources, OSM provides a powerful and versatile alternative in many situations. Finally, although the TDCS method is not recommended, the objection is not to the production of meta-analytic results by subgroup per se (which has potential graphical use—see next section), but to the statistical comparison of such results across subgroups.

5. Suggestions for graphical presentation

If the covariate is categorical (either ordered or not), a forest plot of within-subgroup estimates remains probably the clearest means of displaying the treatment effect by subgroup. Underneath the subgroup effects graphic, we suggest adding one or more interaction effects (within, across, or combined), each represented by a symbol such as a clear circle that cannot be confused with the subgroup estimates nor with a standard meta-analysis pooled estimate (generally represented with black squares and a clear diamond, respectively). These interactions should be derived from an appropriate model such as PWT, CWA, or OSM. If across-trial information concerning the interaction is considered then displaying interaction estimates simultaneously for within- and across-trial effects gives clear and immediate representations of their size, direction, and precision, thus providing great benefit in assessing the risk of ecological bias. Because of the issues affecting the TDCS method described earlier, the subgroup-specific estimates will not necessarily be compatible with the interaction estimate, although this fact is unlikely to be problematic in practice [6].

Figs. 2, 3a, and 4 demonstrate this approach, although they include output from many methods for illustration. This is a new approach to graphical presentation, rarely if ever seen at present, and we suggest it here as the default presentational style for future treatment-covariate interaction analyses using IPD. Assuming the treatment effect is

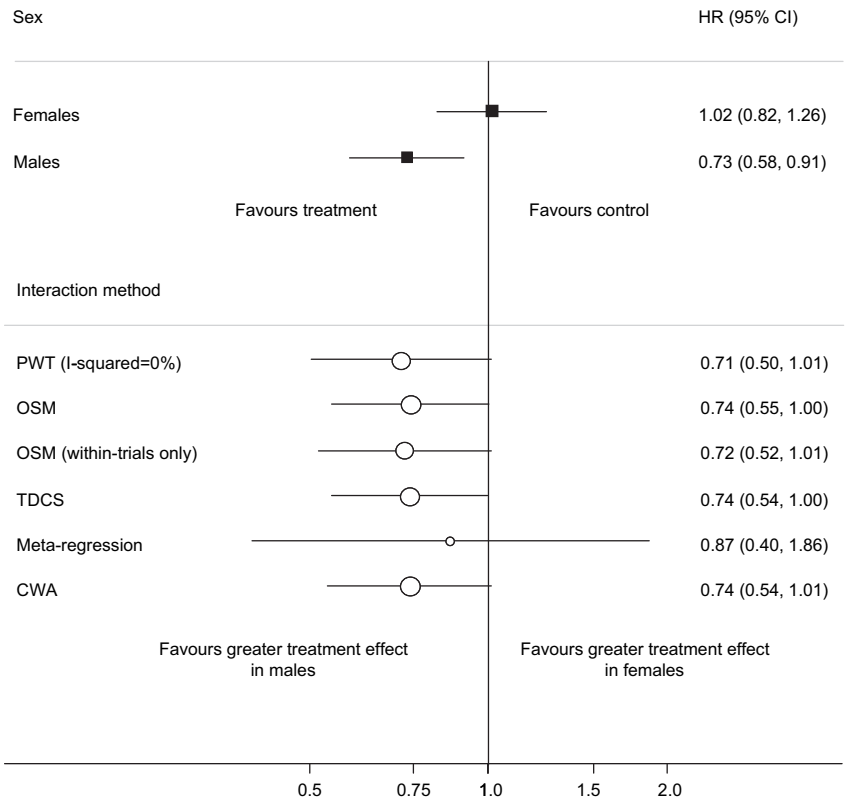


Fig. 2. Treatment effect by sex: sarcoma data set. Independently pooled results for male and female subgroups, and covariate interaction estimates from each of the methods under discussion.

estimated as a hazard or odds ratio, then the interaction estimate is a ratio of ratios, on the same scale and with a similar interpretation. An interaction with a binary covariate is the ratio of the hazard or odds ratios within the two categories. For multiple ordered categories or continuous factors, the interaction effect represents the proportional increase in hazard or odds ratio per category or unit of the covariate.

We have two suggestions for continuous covariates. If PWT is used, a forest plot of within-trial interaction coefficients could be presented along with the pooled interaction estimate. Alternatively, a continuous plot of covariate against treatment effect with confidence limits could be generated using OSM (for linear interactions; see Fig. 3b) or fractional polynomials [26].

6. Case studies

We examined a total of 19 covariates within 5 IPD meta-analyses looking at the effects of treatments on survival (time to death) in the field of cancer [10–14]. Overall, few significant treatment-covariate interactions were found (data not shown). Three case studies were selected to illustrate some of the issues described above [10–12] and to demonstrate a variety of covariate data types (binary, categorical, continuous). Each analysis was originally performed using the TDCS method, by means of the ''log-rank'' modification for time-to-event data [8] to the Peto odds ratio method [22]. Treatment-covariate interactions were reanalyzed and our results compared with those published. For illustrative purposes, we present estimates of the interaction from all four methods we have presented and also from a meta-regression, but highlight the methods that could be selected if either across-trial information is excluded or included without a set of criteria (Fig. 1). The combining of within- and across-trial estimates with CWA is done using fixed-effects inverse-variance weighting only.

6.1. Binary covariate (sarcoma data set [10]; treatment interaction with sex)

Fig. 2 shows that the results from all the methods considered here give broadly consistent results: some evidence of a qualitative interaction effect, suggesting that the treatment is more effective in men than in women. The across-trial effect has a relatively wide confidence interval (CI), indicating low power, and the combined estimate (using OSM or CWA) is almost identical to the within-trial estimate alone. Because heterogeneity in the interaction estimates is low (PWT $I^2 = 0\%$), we suggest PWT should be applied using fixed effects giving an HR = 0.71, 95% CI: 0.50, 1.01, $P = 0.058$. The conclusion is little changed from

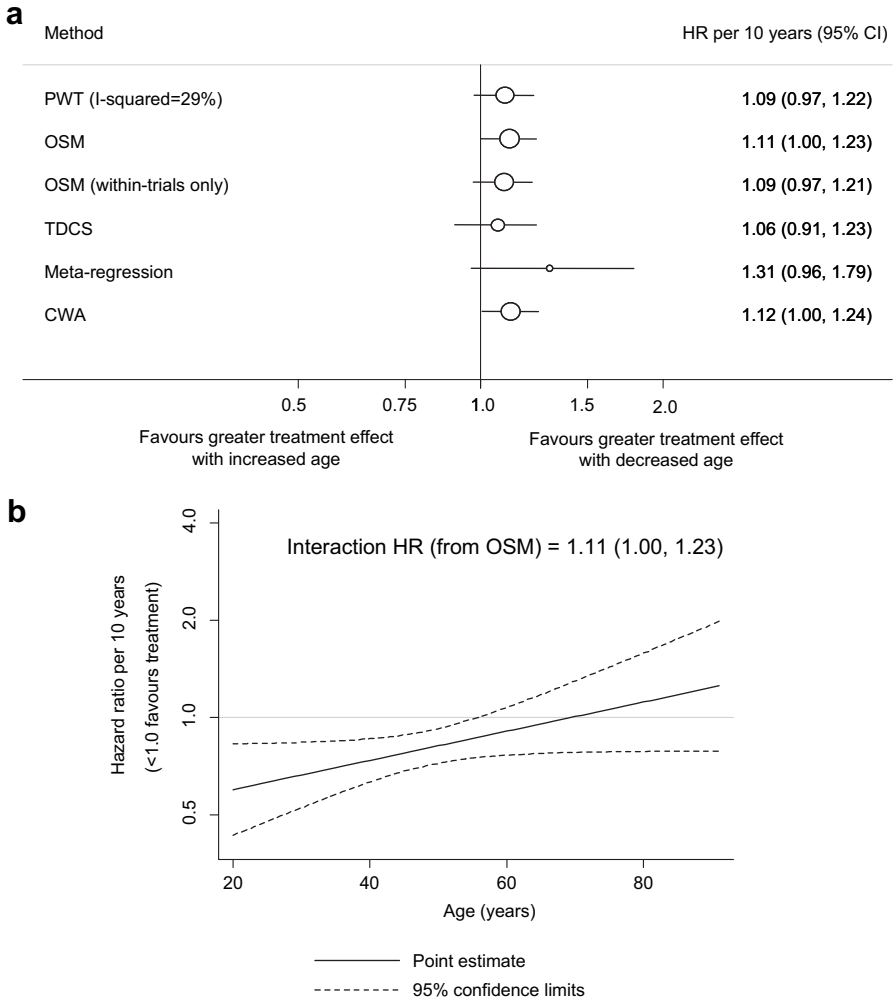


Fig. 3. a. Treatment effect by age: cervix data set. Covariate interaction estimates from each of the methods under discussion. b. Absolute treatment effect by age: cervix data set. Estimated hazard ratio and 95% confidence interval for absolute treatment effect by age at randomization (in units of 10 years).

the original analysis using TDCS, where the interaction HR was 0.74 (95% CI: 0.54, 1.00; $P = 0.049$).

6.2. Continuous covariate (cervical cancer data set [11]; treatment interaction with age)

When originally analyzed, the TDCS method was used which required that age be categorized. No linear trend was seen across age categories ($P = 0.48$). If instead age is used as a continuous linear factor, Fig. 3a shows that the magnitude of the (fixed-effect) within-trial estimate is small, whereas that of the across-trial effect is noticeably larger. The within-trial effect has noticeably greater precision. The combined estimate by CWA is a modest effect significant at the 5% level (HR = 1.12; 95% CI: 1.00, 1.24; $P = 0.044$). However, there is possible trial heterogeneity in the within-trial interaction (using PWA: $I^2 = 29\%$; $P = 0.12$), which arguably merits further investigation.

Assuming that the risk of ecological bias is acceptable, fixed-effect OSM using within- and across-trial information may be used to estimate the absolute effect of treatment by age which may then be shown in a continuous plot with confidence limits (Fig. 3b). Where feasible plots of this type have great value, showing in this case that treatment effect declines with age and may only be beneficial to patients aged under 55.

6.3. Ordered categorical covariate (PORT data set [12]; treatment interaction with nodal status)

The original analysis using TDCS showed reasonable evidence of a linear trend in treatment effect across nodal status categories ($P = 0.016$), suggesting that the treatment may be harmful for those with zero or one affected with lymph nodes. Several trials have few patients in one or more nodal status categories, and others do not obviously show a linear trend in isolation, so further investigation might be warranted. However, trial heterogeneity in the interactions is low (PWT $I^2 = 16\%$; $P = 0.31$), and the within-trial interaction HR (using fixed-effect PWT) shows no effect (HR = 0.91; 95% CI: 0.74, 1.11; $P = 0.34$). The across-trial HR is of noticeably greater magnitude, with

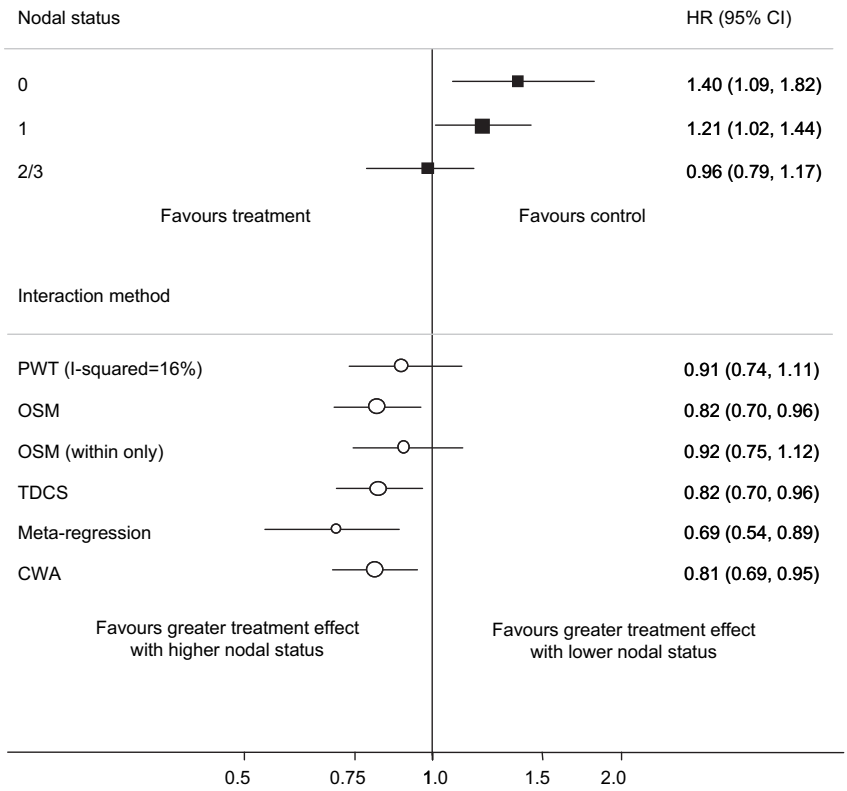


Fig. 4. Treatment effect by nodal status: PORT data set. Independently pooled results for each nodal stage subgroup, and covariate interaction estimates from each of the methods under discussion.

similar precision (HR = 0.69; 95% CI: 0.54, 0.89; $P = 0.003$ using meta-regression, Fig. 4). The combined estimate (by CWA) is significant at the 5% level (HR = 0.81; 95% CI: 0.69, 0.95; $P = .009$). The discrepancy between the PWT analysis and the original analysis using TDCS (or CWA) exemplifies that the inclusion and exclusion of across-trial information can lead to markedly different conclusions. It also illustrates the importance of a priori judgment as to whether across-trial information should be included, and the need for caution concerning possible ecological bias if combined estimates are reported.

7. Discussion

A well-established benefit of collecting IPD is the ability to investigate heterogeneity in treatment effects at the patient level, but various methods of analysis exist and no single method can be recommended for all scenarios. We suggest the choice may naturally be influenced by the objectives of the analysis, the data structure, and the resources and expertize available. Four main approaches have been presented here, each of which is either commonly used in practice or suggested in the literature; all can be fitted (at least at a basic level) using standard statistical software such as Stata. Their characteristics have been compared and discussed, and they have also been applied to real data: a selection of IPD meta-analysis data sets with time-to-event outcomes in the field of cancer.

Our primary conclusion is that method selection should be led initially by whether or not across-trial information is considered for inclusion in the analysis, and that this decision is likely to have the greatest impact on the subsequent findings. This will be a largely subjective decision, influenced by whether the risk of bias from its inclusion is considered to be offset by the potential gain in power. A risk of bias may be considered acceptable, for example, when many interactions are investigated to be identified for later confirmatory studies. Conversely, if an interaction has already been highlighted and definitive confirmation is required then it may be more natural to include only within-trial information concerning the interaction. It is always preferable to decide on the role of the across-trial information and any criteria for its inclusion in advance of analysis.

We have given special prominence to the methods we have termed PWT and CWA, the pooling of within-trial effects possibly combined with a meta-regression, viewing these as natural first steps for an analysis. They are simple to apply and naturally separate the within- and across-trial information. We acknowledge that ‘‘one-stage’’ (OSM) approach will be attractive for some analysts, particularly those with greater resources and expertize. It may be that in due course, when some technical aspects of methodology and software have been addressed and suitable tutorial articles become available, this method will naturally become the dominant approach to IPD meta-analysis. This is not

the case yet, however. Further, particular difficulties remain such as estimating random treatment effects with time-to-event outcomes.

The method we have termed TDCS has been commonly used in the past, for analysis of time-to-event outcomes such as disease survival [27,28] or binary outcomes such as vascular event incidence [29,30]. The reason for its ubiquity appears historical, it being an extension of the Peto ''log-rank'' method for estimation of main effect HRs. This was favored in the past, due probably to its use of noniterative procedures able to be carried out by hand or in nonstatistical packages such as Excel. Indeed, its extension to *trial*-level covariate interaction analysis is methodologically sound and remains an option for those same reasons. However, although TDCS has the ability to detect *patient*-level covariate interactions, its estimates may contain both within- and across-trial information but does not allow the separation of the two sources, so that estimates are at risk of ecological bias.

In this article, we have assumed that we wish to test for interaction between a single treatment effect and a covariate, which is continuous or binary, or an ordered categorical variable whose interaction is to be represented by a linear slope coefficient. These are the simplest scenarios, in which the interactions are represented by a single parameter. In practice, however, the form of the data or the desires of the reviewer may be more complex. Firstly, there are various reasons why a meta-analysis estimating more than one parameter—that is, a multivariate meta-analysis [31,32]—might be desired. One such reason is the estimation of *absolute* effects (i.e., the treatment effect for an individual patient based on their covariate value) where a second parameter is required to represent the baseline treatment effect. Alternatively, the interaction between covariate and treatment may itself require more than one parameter because either the covariate is continuous or ordered but causes a complex nonlinear interaction, or the covariate is an unordered categorical covariate such as histological disease subtype. Of the methods critiqued here, only OSM naturally allows multiple parameters to be estimated simultaneously, and we believe that appropriate models can be formulated that allow separate estimation of within- and across-trial effects for most, if not all, forms of multiple-parameter interactions. The separation of the effects for an unordered categorical covariate is a complex problem in its own right and deserves separate exploration. However, accounting for trial heterogeneity if the outcome is time-to-event may be problematic. Although PWT does not allow such a range of options, it nevertheless allows nonlinearity in the interaction to be tested by comparing a quadratic to the equivalent linear effect, because this requires only one extra parameter. If the evidence for nonlinearity is weak, this is sometimes taken as reassurance that a linear interaction is an adequate summary, although the power to detect such an effect will often be limited. Alternatively, work is ongoing on strategies for combining within-trial interactions for continuous variables estimated using fractional polynomials [33].

A potentially important issue is imbalance in the covariate across the trials in the analysis. For example, if some trials have few or no patients in some covariate categories because of selective reporting [34] this may introduce bias into the within-trial estimate. Similarly, if some trials contain continuous covariate measurements largely from one section of the distribution, then bias may occur if the resulting interaction is not representative of the true effect. There may be little or no possibility of reliably investigating this within the data, and we suggest this would prove a fruitful area for further research. Informally, the analyst might perform a sensitivity analysis by excluding trials where the range of covariate values is limited. Covariate interaction tests cannot be guaranteed to be free from bias or confounding because they are not randomized. However, we feel bias and confounding is minimized by the use of RCTs and by following the guidance given here.

Another area for further work might be to explore an approach within CWA to include across-trial information only if the point estimate for the across-trial interaction effect is sufficiently similar to the within-trial effect point estimate by a priori defined criteria, that is, if the two effect estimates are *consistent*. This idea may be formalized using models incorporating data-driven estimates of consistency. A similar idea is sometimes used in multiple treatment comparison meta-analyses, which use a Bayesian approach [35], although the concept could also be implemented in a frequentist framework.

Overall, we examined a total of 19 covariates within 5 IPD meta-analyses (data not shown). We detected relatively few significant treatment-covariate interactions, which in our experience is a common finding. Nevertheless there were instances, such as that detailed in case study 3, where within- and across-trial effects differed, thus exhibiting ecological bias. In such circumstances, different methods may produce dissimilar estimates and so the selection of the method for analysis, following the guidance we present, may have a major impact on the findings.

Acknowledgments

The authors would like to thank the Medical Research Council for funding this research. They would also like to thank the Sarcoma Meta-analysis Collaboration (SMAC), the Chemoradiotherapy for Cervical Cancer Meta-analysis Collaboration (CCCMAC), and the PORT Meta-analysis Trialists Group (PMT) who brought together the IPD for the meta-analyses used in the case studies. They would also like to thank all the groups that contributed to these meta-analyses for permission to use data from their trials for this research. The contents of this publication and the methods used are, however, the sole responsibility of the authors and do not necessarily represent the views of the meta-analyses

collaborative groups or the trial groups listed. SMAC: Scandinavian Sarcoma Group, Sweden—SSG 81/1; Dana-Faber Cancer Institute, USA—DFCI 78/080; Intergroup Sarcoma Committee, USA—EST 1782; MD Anderson Cancer Center, USA; European Organization for Research and Treatment of Cancer, Belgium—EORTC 62771; National Cancer Institute, USA—NCI SB-S4, NCI SB-S5, NCI SB-S6; Mayo Clinic, USA—#75-77-01; Eastern Cooperative Oncology Group, USA—EST 2377; Istituti Ortopedici Rizzoli, Italy; Institut Bergonie, France; Gynecologic Oncology Group, USA—GOG 20; Swiss Group for Clinical Cancer Research, Switzerland—SAKK 57/87. CCCMAC: Gynecologic Oncology Group, USA—GOG0123, GOG0165; Yale University School of Medicine, USA—YALE HIC 5566; Cross Cancer Institute and University of Alberta, Canada—NCICCTG CX.2; Instituto de Radiología y Centro de Lucha Contra el Cancer, Uruguay—QTCON; University Medical Center Groningen and University of Groningen, Netherlands—DUY-KWF-CYKO 9407; Institute for Oncology and Radiology of Serbia—Project 1683; Toronto Sunnybrook Cancer Center, Canada; First Teaching Hospital, China; Acy'badem Oncology and Neurological Science Hospital, Turkey; Sanjay Gandhi Postgraduate Institute of Medical Sciences, India; Chiang Mai University, Thailand; and University of Yamanashi, Japan. PMT: Institut Jules Bordet and Hôpitaux St Pierre et Erasme, Belgium; Chinese Academy of Medical Sciences, China; European Organization for Research and Treatment of Cancer, Belgium—EORTC 08861; Groupe d'Etude et de Traitement des Cancers Bronchiques, France—GETCB 04CB86, GETCB 05CB88; Lung Cancer Study Group, USA—LCSG 773; Hôpital Calmette, CHRU, France; MRC Clinical Trials Unit, UK—MRC LU11; Institute of Oncology, Slovenia.

Appendix A

Results of literature search

MEDLINE search terms

meta AND (analys OR regression*) AND (subgroup* OR interaction* OR covariate* OR heterogen* OR (effect AND modif*))*

Limits: Year of publication 1966–2009; Title only; English language only.

The limitation to ''title only'' was used to filter out the large number of practical meta-analyses with heterogeneity assessments. It was felt that relevant methodological papers, whether theoretical or reviews of current use, should include the desired search terms in their titles.

Results (140 items)

1. Chen Y, Pei J. Factors influencing the association between CYP17 T34C polymorphism and the risk of breast cancer: meta-regression and subgroup analysis. Breast Cancer Res Treat 2009 [Epub ahead of print].
2. Maher BS, Reimers MA, Riley BP, Kendler KS. Allelic heterogeneity in genetic association meta-analysis: an application to DTNBP1 and schizophrenia. Hum Hered 2010;69:71–9 [Epub 2009 Dec 4].
3. Hennekens CH, Demets D. The need for large-scale randomized evidence without undue emphasis on small trials, meta-analyses, or subgroup analyses. JAMA 2009;302:2361–2.
4. Fiocco M, Putter H, van Houwelingen JC. Meta-analysis of pairs of survival curves under heterogeneity: a Poisson correlated gamma-frailty approach. Stat Med 2009;28:3782–97.
5. Heid IM, Huth C, Loos RJ, Kronenberg F, Adamkova V, Anand SS, et al. Meta-analysis of the INSIG2 association with obesity including 74,345 individuals: does heterogeneity of estimates relate to study design? PLoS Genet 2009;5:e1000694.
6. Nakaoka H, Inoue I. Meta-analysis of genetic association studies: methodologies, between-study heterogeneity and winner's curse. J Hum Genet 2009;54:615–23.
7. Keefe SM, Strom BL. On the importance of heterogeneity in meta-analysis. Clin Trials 2009;6:443–4.
8. Copur MS. Impact of older age on the efficacy of newer adjuvant chemotherapy regimens in colon cancer, a subgroup analysis of a meta-analysis: Practice changing? Certainly not; Hypothesis generating? Perhaps. Clin Colorectal Cancer 2009;8:190–1.
9. Baker WL, White CM, Cappelleri JC, Kluger J, Coleman CI; Health Outcomes, Policy, and Economics (HOPE) Collaborative Group. Understanding heterogeneity in meta-analysis: the role of meta-regression. Int J Clin Pract 2009;63:1426–34.
10. Sanga S, Broom BM, Cristini V, Edgerton ME. Gene expression meta-analysis supports existence of molecular apocrine breast cancer with a role for androgen receptor and implies interactions with ErbB family. BMC Med Genomics 2009;2:59.
11. Virgili G, Conti AA, Moja L, Gensini GF, Gusinu R. Heterogeneity and meta-analyses: do study results truly differ? Intern Emerg Med 2009;4:423–7.
12. Sievenpiper JL, Carleton AJ, Chatha S, Jiang HY, de Souza RJ, Beyene J, et al. Heterogeneous effects of fructose on blood lipids in individuals with type 2 diabetes: systematic review and meta-analysis of experimental trials in humans. Diabetes Care 2009;32:1930–7.
13. Raimondi S, Botteri E, Iodice S, Lowenfels AB, Maisonneuve P. Gene-smoking interaction on colorectal adenoma and cancer risk: review and meta-analysis. Mutat Res 2009;670:6–14.
14. Risch N, Herrell R, Lehner T, Liang KY, Eaves L, Hoh J, et al. Interaction between the serotonin

transporter gene (5-HTTLPR), stressful life events, and risk of depression: a meta-analysis. JAMA 2009;301:2462−71.
15. Bai J, Dai J, Yu H, Shen H, Chen F. Cigarette smoking, MDM2 SNP309, gene-environment interactions, and lung cancer risk: a meta-analysis. J Toxicol Environ Health A 2009;72:677−82.
16. Gong YG, Wu CN, Xing QH, Zhao XZ, Zhu J, He L. A two-method meta-analysis of Neuregulin 1(NRG1) association and heterogeneity in schizophrenia. Schizophr Res 2009;111:109−14.
17. Coory MD. Comment on: Heterogeneity in meta-analysis should be expected and appropriately quantified. Int J Epidemiol 2010;39:932; author reply 933.
18. Alles MC, Gardiner-Garden M, Nott DJ, Wang Y, Foekens JA, Sutherland RL, et al. Meta-analysis and gene set enrichment relative to er status reveal elevated activity of MYC and E2F in the ''basal'' breast cancer subgroup. PLoS One 2009;4:e4710.
19. Salanti G, Marinho V, Higgins JP. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. J Clin Epidemiol 2009;62:857−64.
20. Calura E, Cagnin S, Raffaello A, Laveder P, Lanfranchi G, Romualdi C. Meta-analysis of expression signatures of muscle atrophy: gene interaction networks in early and late stages. BMC Genomics 2008;9:630.
21. Malling HJ, Thomsen AB, Andersen JS. Heterogeneity can impair the results of Cochrane meta-analyses despite accordance with statistical guidelines. Allergy 2008;63:1643−5.
22. Ioannidis JP. Interpretation of tests of heterogeneity and bias in meta-analysis. J Eval Clin Pract 2008;14:951−7.
23. Yu M, Kim S, Wang Z, Hall S, Li L. A Bayesian meta-analysis on published sample mean and variance pharmacokinetic data with application to drug-drug interaction prediction. J Biopharm Stat 2008;18:1063−83.
24. Mastandrea P. Some heterogeneity factors affecting the B-type natriuretic peptides outcome: a meta-analysis. Clin Chem Lab Med 2008;46:1687−95.
25. Wilson SR, Rack JJ, Shi X, Norris AM. Comparing physically abusive, neglectful, and non-maltreating parents during interactions with their children: a meta-analysis of observational studies. Child Abuse Negl 2008;32:897−911.
26. Kotnis A, Kannan S, Sarin R, Mulherkar R. Case-control study and meta-analysis of SULT1A1 Arg213His polymorphism for gene, ethnicity and environment interaction for cancer risk. Br J Cancer 2008;99:1340−7.
27. Sethi S. The problems of meta-analysis for antibiotic treatment of chronic obstructive pulmonary disease, a heterogeneous disease: a commentary on Puhan et al. BMC Med 2008;6:29.
28. Higgins JP. Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. Int J Epidemiol 2008;37:1158−60.
29. Biggerstaff BJ, Jackson D. The exact distribution of Cochran's heterogeneity statistic in one-way random effects meta-analysis. Stat Med 2008;27: 6093−110.
30. Reade MC, Delaney A, Bailey MJ, Angus DC. Bench-to-bedside review: avoiding pitfalls in critical care meta-analysis–funnel plots, risk estimates, types of heterogeneity, baseline risk and the ecologic fallacy. Crit Care 2008;12:220.
31. Bancroft BA, Baker NJ, Blaustein AR. A meta-analysis of the effects of ultraviolet B radiation and its synergistic interactions with pH, contaminants, and disease on amphibian survival. Conserv Biol 2008;22:987−96.
32. 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. Int J Technol Assess Health Care 2008;24:358−61.
33. Kober H, Barrett LF, Joseph J, Bliss-Moreau E, Lindquist K, Wager TD. Functional grouping and cortical-subcortical interactions in emotion: a meta-analysis of neuroimaging studies. Neuroimage 2008;42:998−1031.
34. Bagnardi V, Zatonski W, Scotti L, La Vecchia C, Corrao G. Does drinking pattern modify the effect of alcohol on the risk of coronary heart disease? Evidence from a meta-analysis. J Epidemiol Community Health 2008;62:615−9.
35. Dean MA, Roth PL, Bobko P. Ethnic and gender subgroup differences in assessment center ratings: a meta-analysis. J Appl Psychol 2008;93:685−91.
36. Patsopoulos NA, Evangelou E, Ioannidis JP. Sensitivity of between-study heterogeneity in meta-analysis: proposed metrics and empirical evaluation. Int J Epidemiol 2008;37:1148−57.
37. Iskedjian M, Einarson TR. Meta-analyses of cisapride, omeprazole and ranitidine in the treatment of GORD: implications for treating patient subgroups. Clin Drug Investig 1998;16:9−18.
38. Furlan AD, Tomlinson G, Jadad AA, Bombardier C. Examining heterogeneity in meta-analysis: comparing results of randomized trials and nonrandomized studies of interventions for low back pain. Spine (Phila Pa 1976) 2008;33:339−48.
39. Xu H, Platt RW, Luo ZC, Wei S, Fraser WD. Exploring heterogeneity in meta-analyses: needs, resources and challenges. Paediatr Perinat Epidemiol 2008;22 Suppl 1:18−28.
40. Bakker PR, van Harten PN, van Os J. Antipsychotic-induced tardive dyskinesia and polymorphic

variations in COMT, DRD2, CYP1A2 and MnSOD genes: a meta-analysis of pharmacogenetic interactions. Mol Psychiatry 2008;13:544–56.
41. Rondeau V, Michiels S, Liquet B, Pignon JP. Investigating trial and treatment heterogeneity in an individual patient data meta-analysis of survival data by means of the penalized maximum likelihood approach. Stat Med 2008;27:1894–910.
42. Gonzales DA, Norsworthy KJ, Kern SJ, Banks S, Sieving PC, Star RA, et al. A meta-analysis of N-acetylcysteine in contrast-induced nephrotoxicity: unsupervised clustering to resolve heterogeneity. BMC Med 2007;5:32.
43. Ioannidis JP, Patsopoulos NA, Evangelou E. Uncertainty in heterogeneity estimates in meta-analyses. BMJ 2007;335:914–6.
44. Taylor A, Kim-Cohen J. Meta-analysis of gene-environment interactions in developmental psychopathology. Dev Psychopathol 2007;19:1029–37.
45. Loke YK, Trivedi AN, Singh S. Meta-analysis: gastrointestinal bleeding due to interaction between selective serotonin uptake inhibitors and non-steroidal anti-inflammatory drugs. Aliment Pharmacol Ther 2008;27:31–40.
46. Koopman L, van der Heijden GJ, Glasziou PP, Grobbee DE, Rovers MM. A systematic review of analytical methods used to study subgroups in (individual patient data) meta-analyses. J Clin Epidemiol 2007; 60:1002–9.
47. Martin JL, Sainz-Pardo M, Furukawa TA, Martín-Sánchez E, Seoane T, Galán C. Benzodiazepines in generalized anxiety disorder: heterogeneity of outcomes based on a systematic review and meta-analysis of clinical trials. J Psychopharmacol 2007;21:774–82.
48. Ioannidis JP, Patsopoulos NA, Evangelou E. Heterogeneity in meta-analyses of genome-wide association investigations. PLoS One 2007;2:e841.
49. Novak S, Nemeth WC. How clinically relevant is a meta-analysis of electrical nerve stimulation when based on heterogeneous disease states? Pain 2007;131:228–9.
50. Del Rio Vilas VJ, Hopp P, Nunes T, Ru G, Sivam K, Ortiz-Pelaez A. Explaining the heterogeneous scrapie surveillance figures across Europe: a meta-regression approach. BMC Vet Res 2007;3:13.
51. Gilbody S, Lightfoot T, Sheldon T. Is low folate a risk factor for depression? A meta-analysis and exploration of heterogeneity. J Epidemiol Community Health 2007;61:631–7.
52. Viechtbauer W. Hypothesis tests for population heterogeneity in meta-analysis. Br J Math Stat Psychol 2007;60:29–60.
53. Dohoo I, Stryhn H, Sanchez J. Evaluation of underlying risk as a source of heterogeneity in meta-analyses: a simulation study of Bayesian and frequentist implementations of three models. Prev Vet Med 2007;81:38–55.
54. Haghdoost AA, Alexander N. Systematic review and meta-analysis of the interaction between Plasmodium falciparum and Plasmodium vivax in humans. J Vector Borne Dis 2007;44:33–43.
55. Li L, Yu M, Chin R, Lucksiri A, Flockhart DA, Hall SD. Drug-drug interaction prediction: a Bayesian meta-analysis approach. Stat Med 2007;26:3700–21.
56. Domschke K, Deckert J, O'donovan MC, Glatt SJ. Meta-analysis of COMT val158met in panic disorder: ethnic heterogeneity and gender specificity. Am J Med Genet B Neuropsychiatr Genet 2007;144B:667–73.
57. Thomas R, Zimmer-Gembeck MJ. Behavioral outcomes of Parent-Child Interaction Therapy and Triple P-Positive Parenting Program: a review and meta-analysis. J Abnorm Child Psychol 2007;35: 475–95.
58. Keijzer MB, Borm GF, Blom HJ, Bos GM, Rosendaal FR, den Heijer M. No interaction between factor V Leiden and hyperhomocysteinemia or MTHFR 677TT genotype in venous thrombosis. Results of a meta-analysis of published studies and a large case-only study. Thromb Haemost 2007;97:32–7.
59. Simmonds MC, Higgins JP. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. Stat Med 2007;26:2982–99.
60. Shrier I, Platt RW, Steele RJ. Mega-trials vs. meta-analysis: precision vs. heterogeneity? Contemp Clin Trials 2007;28:324–8.
61. Caro JJ, Xenakis JG, Huybrechts KF. Ignoring study design heterogeneity in meta-analyses. Arch Intern Med 2006;166:2046; author reply 2046–7.
62. Kuhnert R, Böhning D. A comparison of three different models for estimating relative risk in meta-analysis of clinical trials under unobserved heterogeneity. Stat Med 2007;26:2277–96.
63. Mittlböck M, Heinzl H. A simulation study comparing properties of heterogeneity measures in meta-analyses. Stat Med 2006;25:4321–33.
64. Göhler A, Januzzi JL, Worrell SS, Osterziel KJ, Gazelle GS, Dietz R, et al. A systematic meta-analysis of the efficacy and heterogeneity of disease management programs in congestive heart failure. J Card Fail 2006;12:554–67.
65. Zintzaras E, Kitsios G, Harrison GA, Laivuori H, Kivinen K, Kere J, et al. Heterogeneity-based genome search meta-analysis for preeclampsia. Hum Genet 2006;120:360–70.
66. Boccia S, De Lauretis A, Gianfagna F, van Duijn CM, Ricciardi G. CYP2E1PstI/RsaI polymorphism and interaction with tobacco, alcohol and GSTs in gastric cancer susceptibility: a meta-analysis of the literature. Carcinogenesis 2007;28:101–6.

67. Bösch H, Steinkamp F, Boller E. Examining psychokinesis: the interaction of human intention with random number generators—a meta-analysis. Psychol Bull 2006;132:497−523.
68. Kim-Cohen J, Caspi A, Taylor A, Williams B, Newcombe R, Craig IW, et al. MAOA, maltreatment, and gene-environment interaction predicting children's mental health: new evidence and a meta-analysis. Mol Psychiatry 2006;11:903−13.
69. Beyer WE. Heterogeneity of case definitions used in vaccine effectiveness studies and its impact on meta-analysis. Vaccine 2006;24:6602−4.
70. Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? Psychol Methods 2006;11:193−206.
71. Knapp G, Biggerstaff BJ, Hartung J. Assessing the amount of heterogeneity in random-effects meta-analysis. Biom J 2006;48:271−85.
72. Kriston L, Harms A, Berner MM. A meta-regression analysis of treatment effect modifiers in trials with flexible-dose oral sildenafil for erectile dysfunction in broad-spectrum populations. Int J Impot Res 2006;18:559−65.
73. Lewis CM, Levinson DF. Testing for genetic heterogeneity in the genome search meta-analysis method. Genet Epidemiol 2006;30:348−55.
74. Viechtbauer W. Confidence intervals for the amount of heterogeneity in meta-analysis. Stat Med 2007; 26:37−52.
75. Jackson D. The power of the standard test for the presence of heterogeneity in meta-analysis. Stat Med 2006;25:2688−99.
76. Abrams KR, Gillies CL, Lambert PC. Meta-analysis of heterogeneously reported trials assessing change from baseline. Stat Med 2005;24:3823−44.
77. Trikalinos TA, Karvouni A, Zintzaras E, Ylisaukko-oja T, Peltonen L, Järvelä I, et al. A heterogeneity-based genome search meta-analysis for autism-spectrum disorders. Mol Psychiatry 2006;11:29−36.
78. Smith CT, Williamson PR, Marson AG. An overview of methods and empirical comparison of aggregate data and individual patient data results for investigating heterogeneity in meta-analysis of time-to-event outcomes. J Eval Clin Pract 2005;11:468−78.
79. De Laurentiis M, Arpino G, Massarelli E, Ruggiero A, Carlomagno C, Ciardiello F, et al. A meta-analysis on the interaction between HER-2 expression and response to endocrine treatment in advanced breast cancer. Clin Cancer Res 2005;11:4741−8.
80. Zintzaras E, Ioannidis JP. HEGESMA: genome search meta-analysis and heterogeneity testing. Bioinformatics 2005;21:3672−3.
81. Böhning D. Meta-analysis: a unifying meta-likelihood approach framing unobserved heterogeneity, study covariates, publication bias, and study quality. Methods Inf Med 2005;44:127−35.
82. Batterham MJ. Investigating heterogeneity in studies of resting energy expenditure in persons with HIV/AIDS: a meta-analysis. Am J Clin Nutr 2005;81: 702−13.
83. Michiels S, Baujat B, Mahé C, Sargent DJ, Pignon JP. Random effects survival models gave a better understanding of heterogeneity in individual patient data meta-analyses. J Clin Epidemiol 2005;58: 238−45.
84. Smith CT, Williamson PR, Marson AG. Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. Stat Med 2005;24:1307−19.
85. Zeiger JS, Beaty TH, Liang KY. Oral clefts, maternal smoking, and TGFA: a meta-analysis of gene-environment interaction. Cleft Palate Craniofac J 2005;42:58−63.
86. Teramukai S, Matsuyama Y, Mizuno S, Sakamoto J. Individual patient-level and study-level meta-analysis for investigating modifiers of treatment effect. Jpn J Clin Oncol 2004;34:717−21.
87. Huang JQ, Zheng GF, Irvine EJ, Karlberg J. Assessing heterogeneity in meta-analyses of Helicobacter pylori infection-related clinical studies: a critical appraisal. Chin J Dig Dis 2004;5:126−33.
88. Zintzaras E, Ioannidis JP. Heterogeneity testing in meta-analysis of genome searches. Genet Epidemiol 2005;28:123−37.
89. Schmid CH, Stark PC, Berlin JA, Landais P, Lau J. Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. J Clin Epidemiol 2004; 57:683−97.
90. Faddy SC. Significant statistical heterogeneity in a meta-analysis of the usefulness of acetylcysteine for prevention of contrast nephropathy. Am J Cardiol 2004;94:414.
91. Villanueva EV, Zavarsek S. Evaluating heterogeneity in cumulative meta-analyses. BMC Med Res Methodol 2004;4:18.
92. Park EW, Tudiver F, Schultz JK, Campbell T. Does enhancing partner support and interaction improve smoking cessation? A meta-analysis. Ann Fam Med 2004;2:170−4.
93. Bigger JT. Issues in subgroup analyses and meta-analyses of clinical trials. J Cardiovasc Electrophysiol 2003;14(9 Suppl):S6−8.
94. Knapp G, Hartung J. Improved tests for a random effects meta-regression with a single covariate. Stat Med 2003;22:2693−710.
95. Schluchter MD. Publication bias and heterogeneity in the relationship between systolic blood pressure, birth weight, and catch-up growth: a meta-analysis. J Hypertens 2003;21:273−9.

96. Baujat B, Mahé C, Pignon JP, Hill C. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. Stat Med 2002;21:2641–52.
97. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21:1539–58.
98. Takkouche B, Etminan M, Caamaño F, Rochon PA. Interaction between aspirin and ACE inhibitors: resolving discrepancies using a meta-analysis. Drug Saf 2002;25:373–8.
99. Kirsch I, Scoboria A. Apples, oranges, and placebos: heterogeneity in a meta-analysis of placebo effects. Adv Mind Body Med 2001;17:307–9; discussion 312–8.
100. Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI; Anti-Lymphocyte Antibody Induction Therapy Study Group. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. Stat Med 2002;21:371–87.
101. Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. J Clin Epidemiol 2002;55:86–94.
102. Petitti DB. Approaches to heterogeneity in meta-analysis. Stat Med 2001;20:3625–33.
103. Lehmann DJ, Williams J, McBroom J, Smith AD. Using meta-analysis to explain the diversity of results in genetic studies of late-onset Alzheimer's disease and to identify high-risk subgroups. Neuroscience 2001; 108:541–54.
104. Pandit JJ. A meta-analysis of the effect of low-dose volatile anaesthetics on the ventilatory response to hypoxia: analysis for sources of heterogeneity. Adv Exp Med Biol 2001;499:309–14.
105. Song F, Sheldon TA, Sutton AJ, Abrams KR, Jones DR. Methods for exploring heterogeneity in meta-analysis. Eval Health Prof 2001;24:126–51.
106. Clemmens D. The relationship between social support and adolescent mothers' interactions with their infants: a meta-analysis. J Obstet Gynecol Neonatal Nurs 2001;30:410–20.
107. Dwyer T, Couper D, Walter SD. Sources of heterogeneity in the meta-analysis of observational studies: the example of SIDS and sleeping position. J Clin Epidemiol 2001;54:440–7.
108. Trikalinos TA, Ioannidis JP. Predictive modeling and heterogeneity of baseline risk in meta-analysis of individual patient data. J Clin Epidemiol 2001;54: 245–52.
109. Gurevitch J, Morrison JA, Hedges LV. The Interaction between Competition and Predation: A Meta-analysis of Field Experiments. Am Nat 2000;155:435–453.
110. Costa-Bouzas J, Takkouche B, Cadarso-Suárez C, Spiegelman D. HEpiMA: software for the identification of heterogeneity in meta-analysis. Comput Methods Programs Biomed 2001;64: 101–7.
111. Hahn S, Williamson PR, Hutton JL, Garner P, Flynn EV. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. Stat Med 2000;19:3325–36.
112. Engels EA, Schmid CH, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. Stat Med 2000;19:1707–28.
113. Marcus PM, Hayes RB, Vineis P, Garcia-Closas M, Caporaso NE, Autrup H, et al. Cigarette smoking, N-acetyltransferase 2 acetylation status, and bladder cancer risk: a case-series meta-analysis of a gene-environment interaction. Cancer Epidemiol Biomarkers Prev 2000;9:461–7.
114. Cates C. Will eradication of Helicobacter pylori improve symptoms of non-ulcer dyspepsia? Studies included in meta-analysis had heterogenous, not homogenous, results. BMJ 2000;320:1208.
115. Gu C, Province M, Rao DC. Meta-analysis of genetic linkage to quantitative trait loci with study-specific covariates: a mixed-effects model. Genet Epidemiol 1999;17 Suppl 1:S599–604.
116. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. Stat Med 1999;18:2693–708.
117. Song F. Exploring heterogeneity in meta-analysis: is the L'Abbé plot useful? J Clin Epidemiol 1999;52: 725–30.
118. Gueyffier F, Bulpitt C, Boissel JP, Schron E, Ekbom T, Fagard R, et al. Antihypertensive drugs in very old people: a subgroup meta-analysis of randomised controlled trials. INDANA Group. Lancet 1999;353:793–6.
119. Takkouche B, Cadarso-Suárez C, Spiegelman D. Evaluation of old and new tests of heterogeneity in epidemiologic meta-analysis. Am J Epidemiol 1999;150:206–15.
120. Bernsen RM, Tasche MJ, Nagelkerke NJ. Variation in baseline risk as an explanation of heterogeneity in meta-analysis by S. D. Walter, Statistics in Medicine, 16, 2883-2900 (1997). Stat Med 1999;18:233–8.
121. van Houwelingen H, Senn S. Investigating underlying risk as a source of heterogeneity in meta-analysis. Stat Med 1999;18:110–5.
122. Boutitie F, Gueyffier F, Pocock SJ, Boissel JP. Assessing treatment-time interaction in clinical trials with time to event data: a meta-analysis of hypertension trials. Stat Med 1998;17:2883–903.
123. Berry SM. Understanding and testing for heterogeneity across 2 × 2 tables: application to meta-analysis. Stat Med 1998;17:2353–69.
124. Dubertret C, Gorwood P, Ades J, Feingold J, Schwartz JC, Sokoloff P. Meta-analysis of DRD3

gene and schizophrenia: ethnic heterogeneity and significant association in Caucasians. Am J Med Genet 1998;81:318−22.
125. Hardy RJ, Thompson SG. Detecting and describing heterogeneity in meta-analysis. Stat Med 1998; 17:841−56.
126. Stuck AE, Rubenstein LZ, Wieland D. Bias in meta-analysis detected by a simple, graphical test. Asymmetry detected in funnel plot was probably due to true heterogeneity. BMJ 1998;316:469; author reply 470−1.
127. Walter SD. Variation in baseline risk as an explanation of heterogeneity in meta-analysis. Stat Med 1997;16:2883−900.
128. Thompson SG, Smith TC, Sharp SJ. Investigating underlying risk as a source of heterogeneity in meta-analysis. Stat Med 1997;16:2741−58.
129. Biggerstaff BJ, Tweedie RL. Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis. Stat Med 1997;16: 753−68.
130. Hotopf M, Hardy R, Lewis G. Discontinuation rates of SSRIs and tricyclic antidepressants: a meta-analysis and investigation of heterogeneity. Br J Psychiatry 1997;170:120−7.
131. Pladevall-Vila M, Delclos GL, Varas C, Guyer H, Brugués-Tarradellas J, Anglada-Arisa A. Controversy of oral contraceptives and risk of rheumatoid arthritis: meta-analysis of conflicting studies and review of conflicting meta-analyses with special emphasis on analysis of heterogeneity. Am J Epidemiol 1996;144:1−14.
132. Beck CT. The effects of postpartum depression on maternal-infant interaction: a meta-analysis. Nurs Res 1995;44:298−304.
133. Berlin JA. Invited commentary: benefits of heterogeneity in meta-analysis of data from epidemiologic studies. Am J Epidemiol 1995;142:383−7.
134. Colditz GA, Burdick E, Mosteller F. Heterogeneity in meta-analysis of data from epidemiologic studies: a commentary. Am J Epidemiol 1995;142: 371−82.
135. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. BMJ 1994;309:1351−5.
136. Hunink MG, Wong JB. Meta-analysis of failure-time data with adjustment for covariates. Med Decis Making 1994;14:59−70.
137. Parent M, LeBel M. Meta-analysis of quinolone-theophylline interactions. DICP 1991;25:191−4.
138. Ramos JL, Michan C, Rojo F, Dwyer D, Timmis K. Signal-regulator interactions. Genetic analysis of the effector binding site of xylS, the benzoate-activated positive regulator of Pseudomonas TOL plasmid meta-cleavage pathway operon. J Mol Biol 1990; 211:373−82.
139. Holtzman JL, Weeks CE, Kvam DC, Berry DA, Mottonen L, Ekholm BP, et al. Identification of drug interactions by meta-analysis of premarketing trials: the effect of smoking on the pharmacokinetics and dosage requirements for flecainide acetate. Clin Pharmacol Ther 1989;46:1−8.
140. Turley MA. A meta-analysis of informing mothers concerning the sensory and perceptual capabilities of their infants: the effects on maternal-infant interaction. Matern Child Nurs J 1985;14:183−97.

Appendix B

Mathematical description of the four methods discussed

For ease of comparison, each approach is presented as the fitting of a model relating a linear function of treatment x and covariate z to either an outcome Y (continuous, binary, or count) or a relative hazard (if the data are time to event). For generalized linear models (GLMs), this will be written as $\mu = g^{-1}(x, z)$, where $\mu = \mathrm{E(Y)}$. For time-to-event data, this becomes $h_j(t)/h_0(t) = \psi(x_j, z_j)$, where $h_j(t)$ is the hazard function for individual j and $h_0(t)$ is the baseline hazard.

For simplicity, the trial, treatment, covariate, and interaction effects are labeled as α, β, γ, and δ, respectively, in each approach, although these parameters have different values and interpretations across approaches. We use indices i for trial, j for patient and, where applicable, k for covariate level. Subscripts W and A are used to denote within- and across-trial effects, respectively. For simplicity, in this section we consider just two competing treatments (so x is binary) and a patient covariate z that is either binary, continuous, or ordered categorical. We assume that if z is not binary then the interaction can be adequately represented by a linear term so that δ is scalar.

Pooling of within-trial covariate interactions (PWT)

First, the following model should be fitted within each trial i separately:

$$\text{GLM}: \quad g(\mu_{ij}) = \alpha_i + \beta_i x_{ij} + \gamma_i z_{ij} + \delta_i x_{ij} z_{ij} \tag{1a}$$

$$\text{Time to event}: \quad \psi^{-1}\left[h_j(t)/h_0(t)\right] = \alpha_i + \beta_i x_{ij} + \gamma_i z_{ij} + \delta_i x_{ij} z_{ij} \tag{1b}$$

The resulting treatment-covariate interaction effect estimates $\widehat{\delta}_i$ are then pooled using inverse-variance meta-analysis (fixed effects or random effects):

$$\widehat{\delta}_W = \frac{\sum_i w_i \widehat{\delta}_i}{\sum_i w_i}; \quad se\left(\widehat{\delta}_W\right) = \frac{1}{\sqrt{\sum_i w_i}} \tag{2}$$

for weights w_i defined as either $w_i = 1/se\left(\widehat{\delta}_i\right)^2$ (fixed effects) or $w_i = 1/\left(se\left(\widehat{\delta}_i\right)^2 + \tau^2\right)$ (random effects). τ^2 is the

across-trial variance, estimated as follows, where k is the number of studies and Q is defined as in Equation (4):

$$\tau^2 = \frac{Q-(k-1)}{\sum_i w_i - \left(\frac{\sum_i w_i^2}{\sum_i w_i}\right)} \tag{3}$$

Heterogeneity is assessed as for standard inverse-variance meta-analysis, that is:

$$Q = \sum_i w_i \left(\widehat{\delta}_i - \widehat{\delta}\right)^2 \tag{4}$$

distributed as χ^2 on $k-1$ degrees of freedom under the null hypothesis of no heterogeneity.

One-stage models with a covariate interaction term (OSM)

Treatment-covariate interaction is estimated using all available data in a single model. In its most usual form, it can be written as follows:

$$\text{GLM}: \quad g\left(\mu_{ij}\right) = \alpha_i + \beta x_{ij} + \gamma z_{ij} + \delta x_{ij} z_{ij} \tag{5a}$$

$$\text{Time to event}: \quad \psi^{-1}\left[h_j(t)/h_0(t)\right] = \alpha_i + \beta x_{ij} + \gamma z_{ij} + \delta x_{ij} z_{ij} \tag{5b}$$

The trial effect α_i consists of a vector of dummy variables representing trial membership—that is, as a "fixed" effect. The treatment effect β can either be "common" (i.e., a fixed value common to all trials) or "random," in which case $\beta = \beta_0 + u_i$ where $u_i \sim N(0, \tau^2)$. γ and δ are considered as "common" effects. Note that any of the coefficients could potentially be treated as common, fixed, or random depending on model assumptions. The model presented here is that which may be considered the most basic yet useful.

The interaction δ in Equations (5a) or (5b) will be a weighted average of across- and within-trial effects, but there is an alternative formulation of this model, described by Simmonds [1], that enables them to be separated. If we calculate trial mean covariate values $\overline{z}_i$ (i.e., the values that would be used for performing a meta-regression), then we can fit (for a GLM or time-to-event model, respectively):

$$g\left(\mu_{ij}\right) = \alpha_i + \beta x_{ij} + \gamma z_{ij} + \delta_W x_{ij}\left(z_{ij} - \overline{z}_i\right) + \delta_A x_{ij} \overline{z}_i \tag{6a}$$

$$\psi^{-1}\left[h_j(t)/h_0(t)\right] = \alpha_i + \beta x_{ij} + \gamma z_{ij} + \delta_W x_{ij}\left(z_{ij} - \overline{z}_i\right) + \delta_A x_{ij} \overline{z}_i \tag{6b}$$

The trial, treatment, and covariate coefficients α_i, β, and γ retain a similar interpretation to Equations (5a) or (5b), but δ is replaced by δ_W and δ_A. $\delta_A - \delta_W$ represents the "excess" across-trial relationship that exists over and above the within-trial relationship—that is, a quantification of the ecological bias [1]. Having estimated δ_W and δ_A separately using Equations (6a) or (6b), we may consider "combining" these estimates by presenting the estimate of interaction $\widehat{\delta}$ from Equations (5a) or (5b) if suitable criteria are satisfied.

To assess heterogeneity here, each term in the model needs to be allowed to vary with trial membership using fixed effects. That is (for a GLM or time-to-event model, respectively):

$$g\left(\mu_{ij}\right) = \alpha_i + \beta_0 x_{ij} + \beta_{1i} x_{ij} + \gamma_0 z_{ij} + \gamma_{1i} z_{ij} + \delta_0 x_{ij} z_{ij} + \delta_{1i} x_{ij} z_{ij} \tag{7a}$$

$$\psi^{-1}\left[h_j(t)/h_0(t)\right] = \alpha_i + \beta_0 x_{ij} + \beta_{1i} x_{ij} + \gamma_0 z_{ij} + \gamma_{1i} z_{ij} + \delta_0 x_{ij} z_{ij} + \delta_{1i} x_{ij} z_{ij} \tag{7b}$$

where β_0, γ_0, and δ_0 retain the interpretations of β, γ, and δ in Equations (5a) or (5b), and β_1, γ_1, and δ_1 are "fixed" effect variables. Models (7a) and (5a) (or (7b) and (5b)) should then be compared using a log-likelihood test [2].

The same approach can be taken if model (6a) is used. A log-likelihood test should be used to compare that model with the following:

$$g\left(\mu_{ij}\right) = \alpha_i + \beta_0 x_{ij} + \beta_{1i} x_{ij} + \gamma_0 z_{ij} + \gamma_{1i} z_{ij} + \delta_{W0} x_{ij}\left(z_{ij} - \overline{z}_i\right) + \delta_{W1i} x_{ij}\left(z_{ij} - \overline{z}_i\right) + \delta_{A0} x_{ij} \overline{z}_i + \delta_{A1i} x_{ij} \overline{z}_i \tag{8a}$$

Similarly, model (6b) should be compared with the following:

$$\psi^{-1}\left[h_j(t)/h_0(t)\right] = \alpha_i + \beta_0 x_{ij} + \beta_{1i} x_{ij} + \gamma_0 z_{ij} + \gamma_{1i} z_{ij} + \delta_{W0} x_{ij}\left(z_{ij} - \overline{z}_i\right) + \delta_{W1i} x_{ij}\left(z_{ij} - \overline{z}_i\right) + \delta_{A0} x_{ij} \overline{z}_i + \delta_{A1i} x_{ij} \overline{z}_i \tag{8b}$$

Testing for treatment-effect differences across covariate subgroups (TDCS)

A two-stage approach, in which inverse-variance pooled treatment effects $\widehat{\beta}_k$ are first estimated within each patient subgroup defined by $z = k$. In the second stage, a trend line or ANOVA is fitted to these $\widehat{\beta}_k$. We assume that a linear interaction term is appropriate. In GLM form, this is:

$$\widehat{\beta}_k = \beta_0 + \delta k + \epsilon_k, \epsilon_k \sim N\left(0, se\left(\widehat{\beta}_k\right)^2\right) \tag{9}$$

where δ is the interaction effect of interest.

Both the initial estimates $\widehat{\beta}_k$ and the interaction effect δ can be estimated using either fixed effects or random effects, although fixed effects are almost exclusively used in practice.

Because δ is defined across subgroups, a direct assessment of trial heterogeneity in the treatment-covariate interaction effect cannot be performed. However, the trial heterogeneity of each of the $\widehat{\beta}_k$ is straightforward to assess using the Peto method [3].

"Manually" combining separately calculated within- and across-trial effects (CWA)

The PWT method involves calculating an interaction effect within each trial independently and therefore estimates only the within-trial effect, while meta-regression estimates only the *across*-trial effect. The CWA method combines these two interaction effect estimates using a further weighted average.

The meta-regression model to estimate across-trial interaction δ_A is

$$\widehat{\beta}_i = \beta_0 + \delta_A \overline{z}_i + \epsilon_i,\ \epsilon_i \sim N\left(0, \widehat{\sigma}_i^2 + \tau^2\right) \tag{10}$$

where $\widehat{\beta}_i$ is the estimate of treatment effect for trial i with variance $\widehat{\sigma}_i^2$, $\overline{z}_i$ is the mean value of the covariate, and $\tau^2 > 0$ specifies a random-effect meta-regression.

We then combine Equations (10) with (2) as follows:

$$\widehat{\delta} = \frac{w_W \widehat{\delta}_W + w_A \widehat{\delta}_A}{w_W + w_A};\ se\left(\widehat{\delta}\right) = \frac{1}{\sqrt{w_W + w_A}} \tag{11}$$

where $w_W = 1/se\left(\widehat{\delta}_W\right)^2$ and $w_A = 1/se\left(\widehat{\delta}_A\right)^2$.

Assessment of treatment-covariate interaction effect heterogeneity cannot be done for the combined estimate. Conclusions must be based on the heterogeneity assessment from PWT.

References for Appendix B

1. Simmonds MC. Statistical methodology of individual patient data meta-analysis [PhD thesis], University of Cambridge, 2005.
2. Smith CT, Williamson PR, Marson AG. Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. Stat Med 2005;24:1307–19.
3. Deeks J, Altman D, Bradburn M. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Davey Smith G, Altman DG, editors, Systematic reviews in health care: meta-analysis in context. 2nd edition. London: BMJ Books; 2001.

References

[1] DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177–88.

[2] Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. Stat Med 1999;18:2693–708.

[3] Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HIAnti-Lymphocyte Antibody Induction Therapy Study Group. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. Stat Med 2002;21:371–87.

[4] Simmonds MC, Higgins JP, Stewart LA, Tierney JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. Clin Trials 2005;2:209–17.

[5] Simmonds MC. Statistical methodology of individual patient data meta-analysis. [PhD thesis]. Manuscripts Room, Cambridge University Library, West Road, Cambridge CB3 9DR, UK. Classmark: PhD.28903. 2005

[6] Thompson S, Kaptoge S, White I, Wood A, Perry P, Danesh J. The Emerging Risk Factors Collaboration. Statistical methods for the analysis of individual participant data from multiple epidemiological studies. Int J Epidemiol 2010;39:1345–59.

[7] Thompson SG, Higgins JP. Treating individuals 4: can meta-analysis help target interventions at individuals most likely to benefit? Lancet 2005;365:341–6.

[8] Altman DG. Practical statistics for medical research. London, UK: Chapman and Hall; 1991. p. 379.

[9] Higgins JPT, Whitehead A, Turner RM. Meta-analysis of continuous outcome data from individual patients. Stat Med 2001;20:2733–49.

[10] Sarcoma Meta-analysis Collaboration. Adjuvant chemotherapy for localised resectable soft-tissue sarcoma of adults: meta-analysis of individual data. Lancet 1997;350:1647–54.

[11] Chemoradiotherapy for Cervical Cancer Meta-Analysis Collaboration. Reducing uncertainties about the effects of chemoradiotherapy for cervical cancer: a systematic review and meta-analysis of individual patient data from 18 randomised trials. J Clin. Oncol 2008;26:5802–12.

[12] PORT Meta-analysis Trialists Group. Postoperative radiotherapy in non-small-cell lung cancer: systematic review and meta-analysis of individual patient data from nine randomised controlled trials. Lancet 1998;352:257–63.

[13] Advanced Bladder Cancer (ABC) Meta-analysis Collaboration. Adjuvant chemotherapy in invasive bladder cancer: a systematic review and meta-analysis of individual patient data. Eur Urol 2005;48:189–99.

[14] 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.

[15] Deeks J, Altman D, Bradburn M. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Davey Smith G, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. 2nd edition. London, UK: BMJ Books; 2001:285–312.

[16] Riley RD, Lambert PC, Staessen JA, Wang J, Gueyffier F, Thijs L, et al. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. Stat Med 2008;27:1870–93.

[17] Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? Stat Med 2002;21:1559–73.

[18] Koopman L, van der Heijden GJMG, Glasziou PP, Grobbee DE, Rovers MM. A systematic review of analytical methods used to study subgroups in (individual patient data) meta-analyses. J Clin Epidemiol 2007;60:1002–9.

[19] Simmonds MC, Higgins JP. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. Stat Med 2007;26:2982–99.

[20] Riley RD, Dodd SR, Craig JV, Thompson JR, Williamson PR. Meta-analysis of diagnostic test studies using individual patient data and aggregate data. Stat Med 2008;27:6111–36.

[21] Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. J Clin Epidemiol 2002;55:86–94.

[22] Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. Prog Cardiovasc Dis 1985;27:335–71.

[23] Greenland S, Salvan A. Bias in the one-step method for pooling study results. Stat Med 1990;9:247–52.

[24] Deeks J, Bradburn M, Bilker W, Localio R, Berlin J. Much ado about nothing: statistical methods for meta-analysis with rare events. Syst Rev Evid Action Int Cochrane Colloq 6th 1998 Baltim Md 1998;6:50.

[25] Whitehead A. Meta-analysis of controlled clinical trials. Chichester, UK: Wiley; 2002. p. 5.

[26] Royston P, Sauerbrei W. Multivariable model-building: a pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables. Chichester, UK: Wiley; 2008.

[27] Pignon JP, le Maître A, Maillard E, Bourhis J, MACH-NC Collaborative Group. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): an update on 93 randomised trials and 17,346 patients. Radiother Oncol 2009;92:4−14.

[28] Aupérin A, le Pechoux C, Rolland E, Curran WJ, Furuse K, Fournel P, et al. Meta-analysis of concomitant versus sequential radiochemotherapy in locally advanced non-small-cell lung cancer. J Clin Oncol 2010;28:2181−90.

[29] Baigent C, Blackwell L, Collins R, Emberson J, Godwin J, Peto R, et al, Antithrombotic Trialists' (ATT) Collaboration. Aspirin in the primary and secondary prevention of vascular disease: collaborative meta-analysis of individual participant data from randomised trials. Lancet 2009;373:1849−60.

[30] Kearney PM, Blackwell L, Collins R, Keech A, Simes J, Peto R, et al, Cholesterol Treatment Trialists' (CTT) Collaborators. Efficacy of cholesterol-lowering therapy in 18 686 people with diabetes in 14 randomised trials of statins: a meta-analysis. Lancet 2008;371:117−25.

[31] Berkey CS, Hoaglin DC, Antczak-Bouckoms A, Mosteller F, Colditz GA. Meta-analysis of multiple outcomes by regression with random effects. Stat Med 1998;17:2537−50.

[32] White IR. Multivariate random-effects meta-analysis. Stata J 2009;9: 40−56.

[33] Sauerbrei W, Royston P. A new strategy for meta-analysis of continuous covariates in observational studies. Stat Med [in press].

[34] Hahn S, Williamson PR, Hutton JL, Garner P, Flynn EV. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. Stat Med 2000;19:3325−36.

[35] Ades AE. A chain of evidence with mixed comparisons: models for multi-parameter evidence synthesis and consistency of evidence. Stat Med 2003;22:2995−3016.