CLINICAL TRIALS ARTICLE Clinical Trials 2012; 9: 610–620

Meta-analysis of effect sizes reported at multiple time points: A multivariate approach

Thomas A Trikalinos[a] and Ingram Olkin[b]

Background Many comparative studies report results at multiple time points. Such data are correlated because they pertain to the same patients, but are typically meta-analyzed as separate quantitative syntheses at each time point, ignoring the correlations between time points.

Purpose To develop a meta-analytic approach that estimates treatment effects at successive time points and takes account of the stochastic dependencies of those effects.

Methods We present both fixed and random effects methods for multivariate meta-analysis of effect sizes reported at multiple time points. We provide formulas for calculating the covariance (and correlations) of the effect sizes at successive time points for four common metrics (log odds ratio, log risk ratio, risk difference, and arcsine difference) based on data reported in the primary studies. We work through an example of a meta-analysis of 17 randomized trials of radiotherapy and chemotherapy versus radiotherapy alone for the postoperative treatment of patients with malignant gliomas, where in each trial survival is assessed at 6, 12, 18, and 24 months post randomization. We also provide software code for the main analyses described in the article.

Results We discuss the estimation of fixed and random effects models and explore five options for the structure of the covariance matrix of the random effects. In the example, we compare separate (univariate) meta-analyses at each of the four time points with joint analyses across all four time points using the proposed methods. Although results of univariate and multivariate analyses are generally similar in the example, there are small differences in the magnitude of the effect sizes and the corresponding standard errors. We also discuss conditional multivariate analyses where one compares treatment effects at later time points given observed data at earlier time points.

Limitations Simulation and empirical studies are needed to clarify the gains of multivariate analyses compared with separate meta-analyses under a variety of conditions.

Conclusions Data reported at multiple time points are multivariate in nature and are efficiently analyzed using multivariate methods. The latter are an attractive alternative or complement to performing separate meta-analyses. *Clinical Trials* 2012; **9:** 610–620. http://ctj.sagepub.com

[a]Center for Evidence-based Medicine, Brown University, Providence, RI, USA, [b]Department of Statistics, Stanford University, Stanford, CA, USA

Author for correspondence: Thomas A Trikalinos, Center for Evidence-based Medicine, Brown University, P.O. Box G-S121-7, 121 South Main Street, Providence, RI 02912, USA.
Email: thomas_trikalinos@brown.edu

10.1177/1740774512453218

Introduction

Many comparative studies report outcomes at multiple time points. For example, studies may report survival among patients receiving the experimental and comparator interventions at periods of 6 months, 1 year, and 4 years. By their very nature, effect sizes calculated at multiple time points are stochastically dependent (correlated), because they pertain to the same group of patients. That is, patients alive at 1 year include all patients alive at 4 years. The dependency becomes more evident when the time points are close to one another. A typical meta-analytic approach for such data is to perform several separate quantitative syntheses (one at each time point), thereby ignoring the correlations between the time-specific effect sizes. Several authors have commented on the drawbacks of such restricted and separate analyses, noting that it is generally preferable to analyze the data simultaneously in a multivariate framework, so as to provide unified conclusions [1–5]. In particular, simultaneous confidence intervals (CIs) for each time point and for comparisons across time points (showing changes in the effects as time progresses) are developed. In contrast to the CIs obtained from separate meta-analyses, the methodology of simultaneous CIs provides correct coverage probabilities.

The meta-analysis of outcomes reported at distinct time points has been discussed in the context of synthesizing summary survival curves [1,6–8]. Dear [3] shows how to calculate the correlation between consecutive survival probabilities per study arm and describes a fixed effects iterative generalized least squares approach to analyze them jointly. Arends [1,9] extends this method to account for random effects across different studies and demonstrates how to fit such models with standard software. The common pattern of all aforementioned methodologies is to model the survival curve of each treatment arm in each study [6] and then calculate a treatment effect based on the survival curve of each arm.

We propose a meta-analytic approach that estimates treatment effects within studies and then averages them at each time point taking into account the stochastic dependencies of the treatment effects. The methodology does not depend on the spacing of the time points. A meta-analysis that compares survival for two interventions at four time points illustrates the alternative multivariate models for fixed and random effects.

A motivating example

Our example consists of 17 trials comparing postoperative radiation therapy with and without adjuvant chemotherapy in patients with malignant gliomas (Table 1). The outcome is survival at 6, 12, 18, and 24 months. The key feature of these data is that patients who have survived for 24 months have also survived for 6, 12, and 18 months, thereby generating inclusions that need to be taken into account. The data are reconstructed from Table 2 in Fine *et al*. [10] and Table 3 in Dear [3], which show survival proportions and the corresponding standard errors. There is little or no censoring in these

Table 1. Number of survivors at four time points from randomized trials on the treatment of malignant gliomas using radiotherapy plus adjuvant chemotherapy (*E*) versus radiotherapy alone (*C*)

Study	Sample size, *E* (*C*)	Number of survivors, *E* (*C*)			
		6 months	12 months	18 months	24 months
1	19 (22)	16 (20)	11 (12)	4 (8)	4 (3)
2	34 (35)	22 (22)	18 (12)	15 (8)	15 (6)
3	72 (68)	44 (40)	21 (15)	10 (3)	3 (0)
4	22 (20)	19 (12)	14 (5)	5 (4)	2 (3)
5	70 (32)	62 (27)	42 (13)	26 (6)	15 (5)
6	183 (94)	130 (65)	80 (33)	47 (14)	30 (11)
7	26 (50)	24 (30)	13 (18)	5 (10)	3 (9)
8	61 (55)	51 (44)	37 (30)	19 (19)	11 (15)
9	36 (25)	30 (17)	23 (12)	13 (4)	10 (4)
10	45 (35)	43 (35)	19 (14)	8 (4)	6 (0)
11	246 (208)	169 (139)	106 (76)	67 (42)	51 (35)
12	386 (141)	279 (97)	170 (46)	97 (21)	73 (8)
13	59 (32)	56 (30)	34 (17)	21 (9)	20 (7)
14	45 (15)	42 (10)	18 (3)	9 (1)[a]	9 (1)[a]
15	14 (18)	14 (18)	13 (14)	12 (13)	9 (12)
16	26 (19)	21 (15)	12 (10)	6 (4)	5 (1)
17	74 (75)	–	42 (40)	–	23 (30)

Counts are reconstructed from proportions and corresponding standard errors. The reconstructed counts may differ slightly from the ones used in the actual analyses by Fine *et al.* because of rounding.

[a]Note that in study 14, counts at 18 and 24 months are identical. This will result in a singular covariance matrix (see text).

Table 2. Notation for *K* studies comparing an experimental and comparator arm over *m* time points

Study	Experimental arm: $E_1, \ldots, E_m$	Comparator arm: $C_1, \ldots, C_m$	Effect sizes: $d_1, \ldots, d_m$
1	$p^E_{11}, \ldots, p^E_{1m}$	$p^C_{11}, \ldots, p^C_{1m}$	$d_{11}, \ldots, d_{1m}$
2	$p^E_{21}, \ldots, p^E_{2m}$	$p^C_{21}, \ldots, p^C_{2m}$	$d_{21}, \ldots, d_{2m}$
⋮	⋮	⋮	⋮
K	$p^E_{K1}, \ldots, p^E_{Km}$	$p^C_{K1}, \ldots, p^C_{Km}$	$d_{K1}, \ldots, d_{Km}$

Here, $p^{E\,\text{or}\,C}_{kt}$ denote the survival proportion for the experimental or control arm (superscript *E* or *C*), for study *k* at time point *t*. The effect sizes can be the difference in the survival proportions, the log ratio of the survival proportions, or the corresponding log odds ratio (see text and supplementary material Appendix A).

Table 3. Variants for random effect covariance matrices

Variant *A*: common variance and common correlation (two parameters)	Variant *B*: Different variances and common correlation ($m + 1 = 5$ parameters)
$\mathbf{T}_A = \tau^2 \begin{bmatrix} 1 & & & \\ \rho & 1 & & \\ \rho & \rho & 1 & \\ \rho & \rho & \rho & 1 \end{bmatrix}$	$\mathbf{T}_B = \begin{bmatrix} \tau_1^2 & & & \\ \rho\tau_2\tau_1 & \tau_2^2 & & \\ \rho\tau_3\tau_1 & \rho\tau_3\tau_2 & \tau_3^2 & \\ \rho\tau_4\tau_1 & \rho\tau_4\tau_2 & \rho\tau_4\tau_3 & \tau_4^2 \end{bmatrix}$
Variant *C*: common variance and autoregressive correlations (two parameters)	Variant *D*: different variances and autoregressive correlations ($m + 1 = 5$ parameters)
$\mathbf{T}_C = \tau^2 \begin{bmatrix} 1 & & & \\ \rho & 1 & & \\ \rho^2 & \rho & 1 & \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$	$\mathbf{T}_D = \begin{bmatrix} \tau_1^2 & & & \\ \rho\tau_2\tau_1 & \tau_2^2 & & \\ \rho^2\tau_3\tau_1 & \rho\tau_3\tau_2 & \tau_3^2 & \\ \rho^3\tau_4\tau_1 & \rho^2\tau_4\tau_2 & \rho\tau_4\tau_3 & \tau_4^2 \end{bmatrix}$
	Variant *E*: unstructured matrix ($m(m + 1)/2 = 10$ parameters)
	$\mathbf{T}_E = \begin{bmatrix} \tau_{11} & & & \\ \tau_{21} & \tau_{22} & & \\ \tau_{31} & \tau_{32} & \tau_{33} & \\ \tau_{41} & \tau_{42} & \tau_{43} & \tau_{44} \end{bmatrix}$

Examples for $m = 4$ time points.

survival data. We chose this familiar example because it was also used in a classic article on the meta-analysis of survival curves [3]. It is a motivating example, but it is not current, and for that reason not clinically relevant. Nowadays, after initial surgery, most patients with glioblastomas and many patients with anaplastic astrocytomas receive adjuvant chemotherapy with temozolomide, an oral alkylating agent, which was not studied in the 17 trials of the example. Survival proportions at specific time points, such as those in Table 1, can be extracted from survival curves as discussed in Tierney *et al*. [11] (see Fine *et al*. [10] for details on extracting the data in the current example).

Methods

Suppose a meta-analysis of K studies, in which each study reports a proportion for the outcome of interest (e.g., mortality or survival) for the treatment and comparator at each of m time points. Table 2 may help to fix notations. Here, p_{kt}^E and p_{kt}^C denote the proportions from study $k (k = 1, \ldots, K)$ at time point $t (t = 1, \ldots, m)$ in the experimental (E) and comparator arms (C), respectively. The proportions at each successive time point are based on the same sets of patients. For example, p_{12}^E and p_{13}^E represent the proportions from the experimental arm of study 1 at time points 2 and 3, and all patients who are alive at time point 3 were alive at time point 2, so that $p_{12}^E \geq p_{13}^E$. More generally, $p_{kt_1}^{\{E,C\}} \geq p_{kt_2}^{\{E,C\}}$ for arbitrary time points $t_1 < t_2$ in the experimental or control arms of study k. Because of the temporal ordering, the proportions (and the effects) have constraints that make them not interchangeable. In Table 2, d_{kt} are estimates of the respective unobserved population effect size δ_{kt} that compare these proportions. The d_{kt} can be any effect size, including but not limited to (log) odds ratio (OR), (log) relative risk (RR), risk difference (RD), and arcsine differences (ASDs). The ASDs are particularly important in studies with rare outcomes (e.g., safety studies), which may lead to zero cells [12]. The formulas for the variances and covariances of the effect sizes are listed in Appendix A supplementary material. The formulas are not applicable to all multivariate proportions – they require the ordering in our model. At this point, no study-level covariates are assumed.

Meta-analysis models – fixed effects

Under the fixed effects model, at a given time point t, the population effect sizes δ_{kt} are all equal across studies (i.e., $\delta_{kt} = \delta_t$ for all K studies). The expectation of the sample effect size d_{kt} of study k at time point t is the corresponding population effect size δ_t or $E(d_{kt}) = \delta_t$. This model has the general form $d_{kt} = \delta_t + e_{kt}$, where the errors e_{kt} are distributed as discussed below. Following Gleser and Olkin [4,5], we use a general least squares regression framework to relate the sample effect sizes d_{kt} to each other or to study-level covariates. In standard matrix notation for regression models

$$\mathbf{d} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \tag{1}$$

where the Km effect sizes are arranged in the column vector $\mathbf{d}$ first by study and then within study by time, and where $\mathbf{e}$ is a conformably arranged column vector of Km random errors. If there are no covariates, the design matrix $\mathbf{X}$ has Km rows and m columns, and $\boldsymbol{\beta}$ is the column vector of the m time-point-specific summary effect sizes. More explicitly

$$\mathbf{d} = \underset{\text{study } 1}{(d_{11}, \ldots, d_{1m},} \quad \ldots \quad \underset{\text{study } K}{, d_{K1}, \ldots, d_{Km})'},$$

$$\boldsymbol{\beta} = (\delta_1, \ldots, \delta_m)',$$

$$\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_K)' = (\mathbf{I}_m, \ldots, \mathbf{I}_m)',$$

where for $k = 1, \ldots, K$, $\mathbf{X}_k = \mathbf{I}_m$, the identity matrix of order m. Both the univariate (separate) and

multivariate (joint) meta-analysis cases can be represented this way.

The $Km \times Km$ covariance matrix of $\mathbf{d}$ is a block diagonal matrix

$$\boldsymbol{\Sigma} = \mathrm{diag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_K), \tag{2}$$

where $\boldsymbol{\Sigma}_k$ is the $m \times m$ covariance matrix between effect sizes at the different time points within study k. In our example, $\boldsymbol{\Sigma}_k$ is a 4×4 matrix representing the four time points at 6, 12, 18, and 24 months. When one treats the effect sizes at m time points as independent, the off-diagonal elements of $\boldsymbol{\Sigma}_k$ are 0 (which yield multiple univariate analyses).

Meta-analysis models – random effects

The random effects model has the general form $d_{kt} = \delta_t^* + \xi_{kt} + e_{kt}$, where δ_t^* is the true overall mean at time point t, and ξ_{kt} represents the deviation of the true mean in study k from the true overall mean, and Equation (1) is replaced by

$$\mathbf{d} = \mathbf{X}\boldsymbol{\beta} + \Xi + \mathbf{e}, \tag{3}$$

where $\Xi = (\boldsymbol{\xi}_1', \ldots, \boldsymbol{\xi}_k')'$ and each $\boldsymbol{\xi}_k$ represents the $m \times 1$ vector of random effects associated with study k. The random effects $\boldsymbol{\xi}_k$ follow a multivariate (m variate) normal distribution with zero mean vector and $m \times m$ covariance matrix $\boldsymbol{\tau}$, where each element represents the random effects covariance between two time points. When $m = 1$, the model reduces to the classical univariate random effects model.

Estimation

Fixed effects model

The least squares solution to the model (1) provides estimates $\mathbf{b}$ of $\boldsymbol{\beta}$ given by

$$\mathbf{b} = \left(\mathbf{X}'\mathbf{S}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{S}^{-1}\mathbf{d}, \tag{4}$$

where the block diagonal matrix $\mathbf{S} = \mathrm{diag}(\mathbf{S}_1, \ldots, \mathbf{S}_K)$ is a fixed effects consistent estimate (usually the sample covariance matrix) of the covariance matrix $\boldsymbol{\Sigma}$ of $\mathbf{d}$. The m regression coefficients $(b_1, \ldots, b_m)$ are correlated, and their covariance matrix is estimated by

$$\mathbf{C} = \left(\mathbf{X}'\mathbf{S}^{-1}\mathbf{X}\right)^{-1} = (c_{ij}). \tag{5}$$

The goodness of fit of the model is assessed by the statistic

$$Q = (\mathbf{d} - \mathbf{X}\mathbf{b})'\mathbf{S}^{-1}(\mathbf{d} - \mathbf{X}\mathbf{b}), \tag{6}$$

which, based on large sample theory, has a chi-squared distribution with $m(K-1)$ degrees of freedom under the null hypothesis. The Q statistic of Equation (6) is an omnibus test statistic, and the chi-squared distribution may not be a good approximation to the distribution of Q when the number K of studies in a meta-analysis is small, when the sample sizes in the studies are small, or when there is a departure from the normality assumption for the effect sizes. Analysts should therefore be cautious and avoid placing undue emphasis on the statistical significance of the Q statistic. Statistics that quantify the extent of between-study heterogeneity (analogous to I^2 in univariate meta-analysis [13]) may have a more intuitive interpretation than Q and may be more helpful to meta-analysts. However, developing such metrics for multivariate meta-analysis is beyond the scope of this study.

Random effects model

In the random effects model, Equation (3), the covariance matrices $\boldsymbol{\Sigma}_k$ in Equation (2) are replaced by $\boldsymbol{\Sigma}_k + \boldsymbol{\tau}$, where $\boldsymbol{\tau}$ is the random effects covariance matrix, and the covariance matrices $\mathbf{S}_k$ are replaced by $\mathbf{S}_k + \mathbf{T}$, where $\mathbf{T}$ is a consistent estimator of $\boldsymbol{\tau}$. Although $\mathbf{S}_k$ is computed directly from the data, there is no direct method to estimate $\boldsymbol{\tau}$. Furthermore, we can define structural variants for the specification of $\boldsymbol{\tau}$ that represent assumptions on the interdependence of the true effects across time points. In Table 3, we list five alternative patterns for $m = 4$ outcomes.

Using restricted maximum likelihood (REML) for the random effects covariances, one optimizes the log likelihood ($LogL$)

$$LogL = \frac{1}{2}\sum_{k=1}^{K}\left(-m\log(2\pi) + \log(|\mathbf{W}_k|) - \mathbf{D}'_k\mathbf{W}_k\mathbf{D}_k\right) + \frac{1}{2}\log\left(\left|\sum_{k=1}^{K}\mathbf{W}_k\right|\right), \tag{7}$$

where $\pi = 3.14159\ldots$, $\mathbf{W}_k = (\mathbf{S}_k + \mathbf{T})^{-1}$ and $\mathbf{D}_k = \mathbf{d}_k - \mathbf{X}_k\mathbf{b}$, and $|\mathbf{W}_k|$ denotes the determinant of $\mathbf{W}_k$. The optimization is on the elements of $\mathbf{b}$ and the parameters of $\mathbf{T}$ in a joint fashion.

We can compare the fit of alternative REML models to the data using the Akaike Information Criterion (AIC)

$$AIC = 2n^* - 2(LogL^*), \tag{8}$$

which penalizes the log likelihood of the fitted model ($LogL^*$) for the number n^* of estimated parameters in each model. The model with the smaller value of AIC is the preferable.

Inferences

CIs in the univariate case

In the univariate case, a CI for the effect size at time point t is given by

$$\left(b_t - z_{\alpha/2}\sqrt{c_{tt}}, \quad b_t + z_{\alpha/2}\sqrt{c_{tt}}\right), \tag{9}$$

where $z_{\alpha/2}$ is the corresponding percentile of the normal distribution, and c_{tt} is an element of the covariance matrix $\mathbf{C}$ of Equation (5) for the fixed effects model or its counterpart for the random effects model.

As mentioned before, because the univariate case ignores the correlations between the m time points, the multiple CIs obtained from Equation (9) do not provide correct coverage. The Boole–Bonferroni inequalities provide a simple adjustment to control the type I error: $100(1-\alpha)$ percent CIs are obtained by substituting $z_{\alpha/2m}$ for $z_{\alpha/2}$ in Equation (9). These CIs are conservative, especially for moderate to large values of m. In practice, adjustments for multiple comparisons are not often performed in published meta-analyses nor are they incorporated in software. Adjustments in the context of false discovery rates (FDRs) are gaining popularity. In section 'Example', we show results for the univariate case with and without adjustments for multiple comparisons, to allow the reader to appreciate the differences in the worked example.

CIs in the multivariate case

Simultaneous CIs for individual elements β_t, $t = 1, \ldots, m$ or linear combinations of the βs are obtained as follows. For vectors $\mathbf{a} = (a_1, \ldots, a_m)'$ and $\beta = (\beta_1, \ldots, \beta_m)'$, let $L(\mathbf{a}, \boldsymbol{\beta}) = \mathbf{a}'\boldsymbol{\beta} = (a_1\beta_1 + \cdots + a_m\beta_m)$ be a linear combination of the βs and $\hat{L}(\mathbf{a}, \mathbf{b}) = (a_1 b_1 + \cdots + a_m b_m)$ be its estimate. In the most usual cases, each element of $\mathbf{a}$ is ± 1 or 0. Then, a $100(1-\alpha)$ percent (usually 95%) *simultaneous* CIs is given by

$$\left(\hat{L}(\mathbf{a}, \mathbf{b}) - q_\alpha\sqrt{\mathbf{a}'\mathbf{C}\mathbf{a}}, \quad \hat{L}(\mathbf{a}, \mathbf{b}) + q_\alpha\sqrt{\mathbf{a}'\mathbf{C}\mathbf{a}}\right), \tag{3}$$

where q_α is the square root of the $100(1-\alpha)$th percentile of the chi-square distribution with m degrees of freedom. These simultaneous CIs are a special case of Scheffé's F-projections for multiple comparisons [14].

In particular, a CI for the t-th effect size β_t obtained from Equation (10) is

$$\left(b_t - q_\alpha\sqrt{c_{tt}}, \quad b_t + q_\alpha\sqrt{c_{tt}}\right) \tag{11}$$

The CI for a difference $\beta_1 - \beta_2$, for example, obtained from Equation (10) with $\mathbf{a} = (1, -1, \ldots, 0)'$ is

$$\left(b_1 - b_2 - q_\alpha\sqrt{c_{11} + c_{22} - 2c_{12}}, \right.$$
$$\left. b_1 - b_2 + q_\alpha\sqrt{c_{11} + c_{22} - 2c_{12}}\right) \tag{12}$$

The simultaneous CIs are preferable to the Bonferroni corrections when there are many multiple comparisons. For example, with $m = 4$, $q_{0.05} = 3.08$, which corresponds to the Bonferroni corrections for approximately 25 multiple comparisons ($z_{0.05/(2\cdot25)} = 3.09$).

Conditional analyses

Sometimes, it may be desirable to distinguish outcomes experienced at earlier versus later time points. For example, assume a trial in patients with ischemic heart disease comparing revascularization using open coronary artery bypass graft surgery (a very invasive procedure) versus hybrid coronary revascularization (combination of minimally invasive surgery and percutaneous catheter-based coronary intervention). It is probably reasonable to give different interpretations to survival in the immediate postintervention period and in the longer term. Periprocedural harmful events, such as deaths in the first 30 days post surgery, can be thought of as providing information on both the safety and effectiveness of the procedure, and longer term events as providing information primarily on its effectiveness. Of course such interpretations are context-specific and subjective, but provide a motivation for summarizing intervention effects at later time points, conditional of surviving up to an earlier time point.

In our example, we do not have a good reason to estimate conditional treatment effects. For the sake of exposition though, assume that we are interested in the summary treatment effect of adjuvant chemotherapy at 12, 18, and 24 months, *given that patients have survived until 6 months*. These *conditional treatment effects* are different from the summary effects of radiation therapy at 12, 18, and 24 months obtained by Equation (4) for fixed effects or by optimizing the log likelihood function of Equation (7) for random effects.

Furthermore, the conditional treatment effects do not have the same causal interpretation as the unconditional summary treatment effects obtained from Equations (4) and (7): An obvious selection bias is at work, because those who survive the first 6 months are not a random subset of those who entered the trial. Yet, conditional analyses can be useful for specific purposes. A practical application

is in decision, cost-effectiveness, or value of information modeling that is using Markov processes. If treatment effects are time dependent (are not constant over time), then the simulation should be parameterized using conditional treatment effects.

To calculate the effect of therapy at 12, 18, and 24 months given that patients have survived until 6 months, we first obtain the conditional survival proportions. It is easy to show that these are estimated by simply excluding people who died between 0 and 6 months from the calculations. For example, the conditional proportion of surviving to time point 2 (12 months), given survival by time point 1 (6 months) in the experimental arm of any given study, $\pi^E_{2\bullet 1}$, is estimated by dividing the number of those alive at 12 months by the number of those at risk (for simplicity, we have suppressed the study index): $p^E_{2\bullet 1}$ = (*number alive at time t* = 2)/(*number alive at time t* = 1). After some algebra the estimated proportion in the experimental arm is $p^E_{2\bullet 1} = p^E_2/(1 - p^E_1)$ and its variance is $p^E_2(1 - p^E_1 - p^E_2)/(N^E\,(1 - p^E_1)^2)$, where N^E is the number of patients in the experimental arm; and similarly for the comparator arm.

Then, one can use the analyses described in sections 'Estimation' and 'Inferences'. In general, when proportion numerators and denominators become small, a multivariate normal distribution may not be a satisfactory approximation. In that case, it is preferable to base analyses directly on the multinomial likelihood, rather than its approximation by the multivariate normal.

Adding covariates and dealing with missing values

It is relatively straightforward to add covariates in the current framework, in a manner directly analogous to the methods proposed in Hedges and Olkin (Chapter 6) [15] and Berkey *et al*. [16]. As is the case in the worked example, the methods we describe here do not strictly require that all studies have data at all time points. The methodology remains valid even when there are missing data in some studies. In the extreme where each study has data at a single time point, the method reduces to separate standard meta-analyses for each time point.

Implementation

These approaches can be carried out in many programming environments such as Stata, SAS, Octave, or R/S-plus. We implemented analyses in Stata/SE 11 (StataCorp, College Station, TX) and provide Stata code for all analyses at http://research.brown.edu/myresearch/thomas_trikalinos and http://github.com/ttrikalin/multiple_time_points (both websites last accessed on 5 April 2012).

Example

The log OR (or its equivalent OR) metric is used in the example; formulas for the log OR and other metrics are given in Appendix A. Figure 1 shows forest plots of the 17 trials at the four time points. The covariance matrix of study 14 (Table 1) was singular, as no patient died between 18 and 24 months in either arm. In the main analyses, we opted to correct $\mathbf{S}_{14}$ using the ridge regression [17] approach with $\varepsilon = 0.08$. The choice of ε is explored in Appendix A, which also includes a comparison of the ridge regression approach with other options for dealing with singular covariance matrices (supplementary material Figures A1–A3 and Table A1).

Fixed effects meta-analyses

Figure 2 presents summary effects per time point with fixed effects analyses. Overall, the point estimates of the summary ORs are in the direction that favors the combination intervention. OR estimates range from 1.25 to 1.63 in univariate analyses and from 1.21 to 1.53 in multivariate analyses. Although the results of univariate and multivariate analyses are generally similar, there are several observations to make. The small differences in point estimates and their standard errors between univariate and multivariate analyses can affect borderline inferences of statistical significance. For example, the uncorrected CI of the OR at 6 months barely excludes or includes zero in the univariate or multivariate analyses, respectively (Figure 2 and Table A1). Such differences emerge because the multivariate analyses account for the correlations between the four time points, whereas the univariate analyses assume no correlation.

The Q statistic (6) does not suggest lack of fit of the multivariate fixed effects model to the data in the main analysis (P = 0.46). If instead of a ridge regression one excludes the data of study 14 at 18 or 24 months, the corresponding P value becomes 0.06 (Table A1). In univariate fixed effects analyses, the Q statistic suggests lack of fit to the 24 months data (i.e., larger variability in the effect of each study; Figure 1, last forest plot; P = 0.05), but not to the 6, 12, or 18 months data (P = 0.35, 0.88, or 0.66, respectively, Table A1).

Random effects meta-analyses

As previously noted, one can use various structures for the covariance matrix of the random effects.

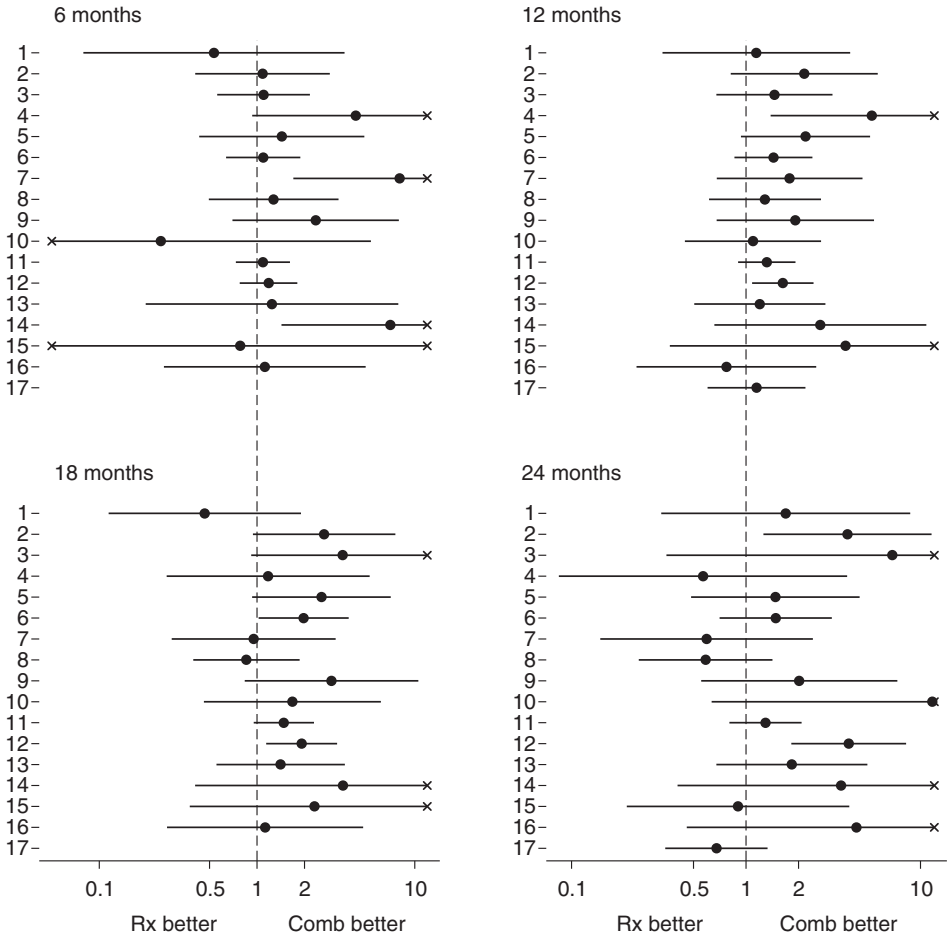

Figure 1. Forest plots of odds ratios for survival at 6, 12, 18, and 24 months across 17 trials comparing combination of chemotherapy and radiotherapy with radiotherapy. Study IDs correspond to those in Table 1. Small 'x' signs mark truncated confidence intervals. Comb: combination of chemotherapy and radiotherapy; Rx: radiotherapy.

Here, we list five variants for $\mathbf{T}$ denoted as $\mathbf{T}_A$, $\mathbf{T}_B$, $\mathbf{T}_C$, $\mathbf{T}_D$, and $\mathbf{T}_E$. $\mathbf{T}_A$ is the most constrained and $\mathbf{T}_E$ is the most general. In variants $\mathbf{T}_A$ and $\mathbf{T}_B$ in Table 3, the correlation between time points is common; the variances are common in $\mathbf{T}_A$ and different in $\mathbf{T}_B$. Variants $\mathbf{T}_C$ and $\mathbf{T}_D$ each have an autoregressive correlation structure between time points; the variances are common in $\mathbf{T}_C$ and different in $\mathbf{T}_D$. No structural assumptions are made in variant $\mathbf{T}_E$. Each of these variants has different connotations. For example, variants $\mathbf{T}_C$ and $\mathbf{T}_D$, which assume a lag-1 autoregressive correlation structure, may be more meaningful when the time points are equally spaced. The number of parameters ranges from 2 in $\mathbf{T}_A$ and $\mathbf{T}_C$, to $m+1$ in $\mathbf{T}_B$ and $\mathbf{T}_D$, and $m(m+1)/2$ in $\mathbf{T}_E$. Appendix A discusses technical details when fitting the random effects models using the five variants.

Results from univariate random effects analyses (estimated with REML) are listed in Table A2. The summary treatment effects are very similar to those from univariate fixed effects analyses in magnitude and statistical significance, and are otherwise unremarkable.

Figure 3 shows summary treatment effects with the five variants of the multivariate random effects model. There seem to be two patterns in the results. The results of variant $\mathbf{T}_A$ are most similar to those of variant $\mathbf{T}_C$, in that the CIs of all four summary effects have the same length. The results of variant $\mathbf{T}_B$ are most similar to those of variants $\mathbf{T}_D$ and $\mathbf{T}_E$. CIs are narrowest at 12 months and become wider at 18 months and more so at 24 months. The two patterns are a direct consequence of the specification of the random effects covariance matrices. Matrices $\mathbf{T}_A$ and $\mathbf{T}_C$ assume a common random effects variance for all four time points. In contrast, matrices $\mathbf{T}_B$, $\mathbf{T}_D$, and $\mathbf{T}_E$ all assume different random effects variances for the four time points.

In Table A2, when the variant assumes a common variance ($\mathbf{T}_A$ and $\mathbf{T}_C$), all the entries in the

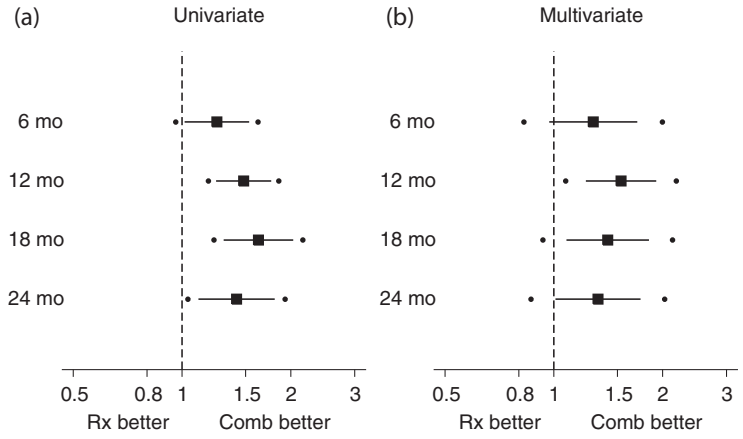

Figure 2. Comparison of summary odds ratios for survival in univariate and multivariate fixed effects meta-analyses. Shown are summary effects and 95% confidence intervals unadjusted for multiple comparisons. Small dots mark the endpoints of the Bonferroni-corrected confidence intervals for four comparisons ((a) univariate analyses) or simultaneous confidence intervals ((b) multivariate analyses) (see Table A1 for numerical data).
Comb: combination of chemotherapy and radiotherapy; mo: months; Rx: radiotherapy.

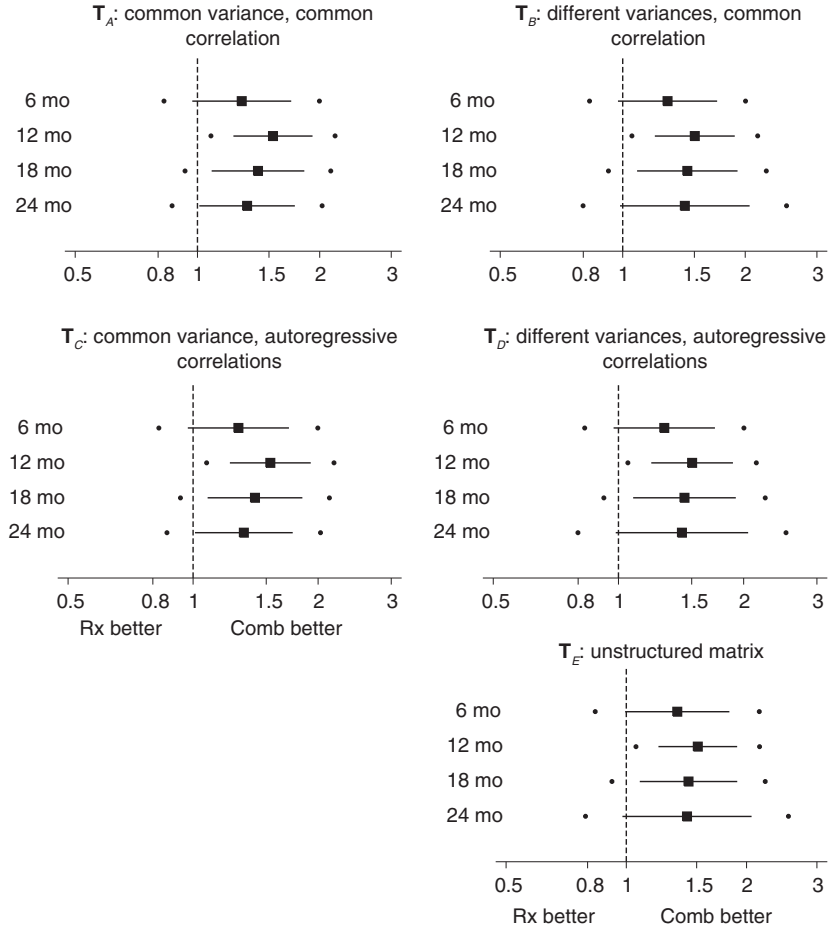

Figure 3. Comparison of summary odds ratios for survival in multivariate random effects meta-analyses with five variants of the covariance structure. Shown are summary effects and 95% confidence intervals unadjusted for multiple comparisons. Small dots mark the endpoints of the simultaneous confidence intervals (see Table A2 for numerical data).
Comb: combination of chemotherapy and radiotherapy; mo: months; Rx: radiotherapy.

Table 4. Fit of multivariate fixed and random effects models to data in the worked example

Model	Log likelihood	Number of parameters	Akaike Information Criterion
Fixed (multivariate)	−60.62	4	129.24
Random, variant $\mathbf{T}_A$	−60.62	6	133.23
Random, variant $\mathbf{T}_B$	−59.09	9	136.18
Random, variant $\mathbf{T}_C$	−60.62	6	133.24
Random, variant $\mathbf{T}_D$	−59.09	9	136.18
Random, variant $\mathbf{T}_E$	−58.63	14	145.25

covariance matrix are small, but when all the variances are allowed to differ ($\mathbf{T}_B$,$\mathbf{T}_D$, and $\mathbf{T}_E$) entries corresponding to the variances and covariances between 12, 18, and 24 month data are larger. The estimated value for the common variance in $\mathbf{T}_A$ and $\mathbf{T}_C$ is an 'average' for all time points. Because data at 6, 12, and 18 months are not statistically heterogeneous and only data at 24 months have evidence for heterogeneity (Table A1), the common estimate obtains a small value when maximizing the likelihood.

The point estimates of the five random effects variants as well as those of the multivariate fixed effects model are quite similar. Based on the AIC, we would favor the fixed effects model among the multivariate fixed and five random effects models (Table 4).

Linear combinations

Table 5 depicts examples of simultaneous linear combinations with different models. Although there may be many linear combinations that can be examined, in practice, only those that compare effects at earlier versus later time points are of interest. For example, to compare summary effects at time points 1 (6 months) and 4 (24 months) use the contrast vector $\mathbf{a} = (1, 0, 0, -1)'$ in Equation (10). As shown in Table 5, in all analyses the treatment effect at 6 months is a bit smaller than that at 24 months, but not significantly so. To calculate the average of the summary treatment effects, we form the contrast vector $\mathbf{a} = (0.25, 0.25, 0.25, 0.25)'$, and so on. Note that the exponentiated average of log OR is the geometric mean of OR.

Table 5. Corrected simultaneous CIs of linear combinations based on the multivariate analyses

Model	Relative OR (95% CI) between 6 and 24 months data[a]	Geometric mean of summary ORs (95% CI)[b]
Fixed	0.97 (0.55, 1.70)	1.39 (1.04, 1.85)
Random, variant $\mathbf{T}_A$	0.97 (0.55, 1.70)	1.39 (1.04, 1.85)
Random, variant $\mathbf{T}_B$	0.91 (0.46, 1.80)	1.41 (1.03, 1.94)
Random, variant $\mathbf{T}_C$	0.97 (0.55, 1.70)	1.39 (1.04, 1.85)
Random, variant $\mathbf{T}_D$	0.91 (0.46, 1.80)	1.41 (1.03, 1.94)
Random, variant $\mathbf{T}_E$	0.95 (0.43, 2.08)	1.42 (1.05, 1.93)

CI: confidence interval; OR: odds ratio.
[a]The first column is the exponentiated difference between the summary log OR at time points 1 (6 months) and 4 (24 months).
[b]The second column is the exponentiated mean of the summary log ORs across all four time points.

Table 6. Results from conditional analyses

Time point	Conditional analyses, OR (95% CI)
Conditional on surviving up to 6 months	
12 months given 6 months	1.33 (0.90, 1.98)
18 months given 6 months	1.54 (1.11, 2.14)
24 months given 6 months	1.56 (1.06, 2.28)
Conditional on surviving up to 12 months	
18 months given 12 months	1.31 (0.93, 1.84)
24 months given 12 months	1.55 (1.16, 2.06)
Conditional on surviving up to 18 months	
24 months given 18 months	1.34 (1.03, 1.75)

CI: confidence interval; OR: odds ratio.
Shown are 95% simultaneous CIs. We exclude study 17 that has missing data at 6 and 18 months from the analyses described in this table.

Conditional analyses

The conditional analysis is presented in Table 6. For example, the relative OR at 24 months given 6 months data is 1.56 (95% CI: 1.06, 2.28), with ORs above 1 favoring the addition of chemotherapy to postoperative radiation therapy. The corresponding effects given 12 and 18 months data are 1.55 (95% CI: 1.16, 2.06) and 1.34 (95% CI: 1.03, 1.75), respectively.

In the example, sample sizes in most studies are small, and thus the multivariate normal distribution may be a poor approximation to the distribution of the transformed survival proportions, which is important to the conditional analyses. This would of course affect the validity of the results in Table 6, which are nevertheless presented for completeness.

Discussion

Many systematic reviews use quantitative analyses to compare multiple treatments with respect to multiple outcomes assessed at multiple time points. Because these are multivariate problems, multivariate meta-analysis methods should be used in the

analysis. A series of methodological publications has described multivariate methodologies to address increasingly complex data in a joint, efficient fashion [1–5,9,16,18–20]. In this study, we describe multivariate fixed and random effects meta-analysis approaches that take into account the correlations between effects sizes at different time points. An example shows the comparison of the multivariate model with the separate univariate meta-analyses for each time point. The methodology for calculating a conditional analysis is also provided.

In our application, the payoff of multivariate analyses was modest at best, in that the differences in the summary estimates from univariate analyses were not that pronounced. It can be shown that, for the models presented here, the summary point estimates will be the same in multivariate and univariate analyses if the within-study covariances are zero, or if the within-study covariance matrices are all equal ($\mathbf{S}_k = \mathbf{S}$ for all k) [21]. The standard errors of the summary estimates will generally be different between univariate and multivariate meta-analyses. A large-scale empirical evaluation of real examples would be particularly informative.

So should one use univariate or multivariate meta-analysis for analyzing data of the type described here? This study introduces a methodology and is not sufficient for making methodological recommendations. The decision for using a univariate or multivariate analysis depends on the underlying assumptions that the researcher believes to be characteristic of the data. This decision is not straightforward because there are many factors that come into play: (1) should we assume that the data are normally distributed or not, (2) should we assume independent binomial outcomes or assume a multinomial distribution, and (3) should we impute missing values, and many others. These decisions should be made early on in the analysis and not after an examination of the data. In some instances alternative analyses will yield similar results, in other instances they may not. In the present case, we propose a multivariate analysis that is designed to provide correct coverage of the CIs, whereas univariate analyses may or may not yield correct coverage. The fact that there is little difference in the analyzed example is not a rationale for its general use.

More generally, methodological recommendations address the problem of choosing between alternative methodologies, and developing them should be approached as a decisional problem. In brief, one must define the *decisional context* that includes specifying (1) the *perspective* from which the problem is approached (e.g., recommendations for specialists who perform publicly funded research may have to set a much higher bar compared to recommendations for the meta-analysis community, which may want to set pragmatic minimum standards); (2) all reasonable *alternative choices*; (3) the *type of problems* to which the recommendation applies; (4) the *characteristics or quantities (utilities)* that will be used for making a decision, and their relative weights if more than one exist. The latter represent not only the scientific rigor of each alternative but also issues such as the feasibility of performing an analysis without access to statistical expertise or specialized software, and other concerns. Subsequently, one must examine theoretical arguments, results from simulation analyses, or empirical data, as applicable, and be transparent on how the recommendation is reached. Pending an adequate exploration of this decisional problem in their setting, our opinion is that meta-analysts facing problems such the one analyzed here should consider performing a multivariate meta-analysis, even as a sensitivity analysis.

Our methodology is not a meta-analysis of survival curves [1,9], but an alternative analysis for studies that report multiple time points, and in which censoring is not an issue. This type of model occurs frequently in the social sciences, when examining the effects of interventions on dropouts at different educational attainment levels. In studies with long duration, often people are lost to follow-up (right censoring), and therefore, it cannot be assumed that they have the same number of patients as baseline at the follow-up time points, an assumption we make to derive the correlation between time points. If the percentage lost to follow-up is substantial and if survival data are reported in the primary study in adequate detail, it may be preferable to perform a meta-analysis of survival curves [1,9].

Our method assumes that studies report results at 'similar' time points. In reality, the exact times may be slightly different across studies, and some judgment will be required to group 'similar' time points into the same category. The method does not assume specific spacing for the time points. For example, in the surgical literature, deaths during the early postintervention period (typically within 30 days post intervention) are often considered procedure-related, and potentially of different etiology than subsequent deaths in the mid- or long-term. Studies reporting deaths at time points very near 30 days (e.g., between 27 and 33 days) could be considered together for the '30 days' time point. Consultation with a content expert may be necessary to decide whether a study reporting follow-up at, for example, 50 days can be reasonably assigned to the same category. Greater tolerance may be reasonable for longer follow-ups. In the previous example, it may be acceptable to group

together data reported at 20–28 months in the same time point category (at approximately '2 years'). In all cases, the decision is clinical and epidemiological, rather than statistical.

Finally, this method may be relevant when one wants to distinguish treatment effects in earlier versus later time points. In fact, the described conditional type analyses are particularly amenable to such explorations. Furthermore, our method affords us the opportunity to test whether treatment effects are constant throughout the follow-up period. This is conceptually related to the assumption of proportional hazards in survival-type analyses. We did not use the log hazard ratio, that is, the difference in the complementary log–log-transformed proportions, as the metric of interest in this article, but one could derive the formulas for this metric as well. In any case, how to perform such explorations and whether they should or should not be routinely performed is a question for further research.

Funding

This study was supported in part by a grant from the Agency for Health Care Research and Quality (grant number R01HS018574). Ingram Olkin was also supported by the National Science Foundation Grant 063413.

References

1. **Arends LR**. *Multivariate Meta-Analysis: Modelling the Heterogeneity*. Doctoral Thesis, Erasmus University, The Netherlands, 2006.
2. **Berkey CS, Anderson JJ, Hoaglin DC**. Multiple-outcome meta-analysis of clinical trials. *Stat Med* 1996; **15**(5): 537–57.
3. **Dear KB**. Iterative generalized least squares for meta-analysis of survival data at multiple times. *Biometrics* 1994; **50**(4): 989–1002.
4. **Gleser LJ, Olkin I**. Stochastically dependent effect sizes. In Cooper H, Hedges LV. (eds). *The Handbook of Research Synthesis* (1st edn). Russell Sage Foundation, New York, 1994, pp. 339–55.
5. **Gleser LJ, Olkin I**. Stochastically dependent effect sizes. In Cooper H, Hedges LV, Valentine JC. (eds). *The Handbook of Research Synthesis and Meta-Analysis* (2nd edn). Russell Sage Foundation, New York, 2009, pp. 357–76.
6. **Earle CC, Pham B, Wells GA**. An assessment of methods to combine published survival curves. *Med Decis Making* 2000; **20**(1): 104–11.
7. **Hunink MG, Wong JB**. Meta-analysis of failure-time data with adjustment for covariates. *Med Decis Making* 1994; **14**(1): 59–70.
8. **Voest EE, van Houwelingen JC, Neijt JP**. A meta-analysis of prognostic factors in advanced ovarian cancer with median survival and overall survival (measured with the log (relative risk)) as main objectives. *Eur J Cancer Clin Oncol* 1989; **25**(4): 711–20.
9. **Arends LR, Hunink MGM, Stijnen T**. Meta-analysis of summary survival curve data. *Stat Med* 2008; **27**(22): 4381–96.
10. **Fine HA, Dear KB, Loeffler JS, Black PM, Canellos GP**. Meta-analysis of radiation therapy with and without adjuvant chemotherapy for malignant gliomas in adults. *Cancer* 1993; **71**(8): 2585–97.
11. **Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR**. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials* 2007; **8**: 16.
12. **Rücker G, Schwarzer G, Carpenter J, Olkin I**. Why add anything to nothing? The arcsine difference as a measure of treatment effect in meta-analysis with zero cells. *Stat Med* 2009; **28**(5): 721–38.
13. **Higgins JP, Thompson SG**. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; **21**(11): 1539–58.
14. **Miller RG**. *Simultaneous Statistical Inference* (2nd edn). Springer-Verlag, New York, 1981.
15. **Hedges LV, Olkin I**. *Statistical Methods for Meta-Analysis*. Academic Press Inc., San Diego, CA, 1985.
16. **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**(22): 2537–50.
17. **Bishop C**. *Pattern Recognition and Machine Learning* (1st edn). Springer, Ithaca, NY, 2006.
18. **Riley RD, Abrams KR, Sutton AJ, Lambert PC, Thompson JR**. Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Med Res Methodol* 2007; **7**: 3.
19. **Trikalinos TA, Olkin I**. A method for the meta-analysis of mutually exclusive binary outcomes. *Stat Med* 2008; **27**(21): 4279–300.
20. **van Houwelingen HC, Arends LR, Stijnen T**. Advanced methods in meta-analysis: Multivariate approach and meta-regression. *Stat Med* 2002; **21**(4): 589–624.
21. **Riley RD, Abrams KR, Lambert PC, Sutton AJ, Thompson JR**. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Stat Med* 2007; **26**(1): 78–97.