STATISTICS IN MEDICINE
Statist. Med. 2008; **27**:4279–4300
Published online 17 April 2008 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3299

A method for the meta-analysis of mutually exclusive binary outcomes

Thomas A. Trikalinos[1,*,†] and Ingram Olkin[2]

[1]*Center for Clinical Evidence Synthesis, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, 800 Washington St, Box #63, Boston, MA 02111, U.S.A.*
[2]*Department of Statistics, Stanford University, Stanford, CA, U.S.A.*

SUMMARY

Meta-analyses of multiple outcomes need to take into account the within-study correlation across the different outcomes. Here we focus on the meta-analysis of dichotomous outcomes that are mutually exclusive and exhaustive. Correlations between effect sizes for mutually exclusive outcomes are negative and can be obtained from data already available. We present both fixed-effects and random-effects methods that account for the negative correlations and yield correct simultaneous confidence intervals for both the marginal outcome-specific effect sizes and the relative effect sizes between outcomes. Formulae for the odds ratio, risk ratio, risk difference, and the differences in the arcsin-transformed risks are provided. An example of a meta-analysis of randomized trials of radiotherapy and mastectomy with axillary lymph node clearance *versus* only mastectomy with axillary clearance for early breast cancer is presented. The mutually exclusive outcomes of breast cancer deaths and deaths secondary to other causes are examined in separate meta-analyses, and also by taking the between-outcome correlation into account. We argue that mutually exclusive outcomes in the meta-analyses of binary data are optimally analyzed in a multinomial setting. This may also be applicable when a meta-analysis examines only one out of several mutually exclusive outcomes. For large sample sizes and/or low event counts, the covariances between outcome-specific effect sizes are small, and either ignoring them or accounting for them would result in similar estimates for any practical purpose. However, meta-analysts should explore the robustness of the findings from individual meta-analyses when mutually exclusive outcomes are assessed. Copyright © 2008 John Wiley & Sons, Ltd.

KEY WORDS: multinomial outcomes; multivariate regression; mutually exclusive outcomes; meta-analysis

*Correspondence to: Thomas A. Trikalinos, Center for Clinical Evidence Synthesis, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, 800 Washington St, Box #63, Boston, MA 02111, U.S.A.
†E-mail: ttrikalin@mac.com

Contract/grant sponsor: Evidence-based Center of Excellence, Pfizer Inc.
Contract/grant sponsor: National Science Foundation; contract/grant number: 063413

Received 9 March 2007
Accepted 10 March 2008

1. INTRODUCTION

Many comparative studies in the biomedical domain evaluate more than one outcome measure for each subject. In their simplest form, multiple-outcome (multiple-endpoint) studies compare two interventions with respect to two or more endpoints. Effect sizes can be calculated for each outcome, and because they are assessed in the same patients, they are stochastically dependent. The usual practice in meta-analyses is to consider such outcomes in separate syntheses, ignoring the within-study correlations of the evaluated effect sizes. Various frameworks that account for correlations have been proposed for binary [1] and continuous (both standardized [1] and unstandardized [2]) endpoints. These approaches are essentially generalized least-squares (GLS) regressions that incorporate the between-outcome (within-study) correlations as provided by the primary studies, or more often, as known from previous research [1]. Random-effects approaches for stochastically dependent outcomes have also been explored from both a frequentist perspective [3, 4] and a Bayesian perspective [5]. In this study we present a similar framework for the meta-analysis of mutually exclusive and dichotomous outcomes. In contrast to other types of stochastically dependent outcomes, correlations between effect sizes for mutually exclusive dichotomous outcomes can be obtained from data already available.

A case of mutually exclusive outcomes is mortality secondary to different causes. For example, consider three mutually exclusive health states: 'dead secondary to breast cancer', 'dead secondary to any other cause', and 'alive'. These are adequately described by analyzing two mutually exclusive outcomes (i.e. the two cause-specific mortalities).

By their very nature, dichotomous mutually exclusive endpoints are negatively correlated: For every patient who is lost to breast cancer, one less patient is at risk for other outcomes. Ignoring these (negative) correlations (as happens in separate meta-analyses) can yield incorrect confidence intervals for simultaneous inferences. This seems to be the usual practice: In an electronic search of the full text of the Cochrane Library of Systematic Reviews (3rd issue 2007) using the terms 'multivariate' or 'multinomial', none of the identified meta-analyses of aggregate data addressed correlations between mutually exclusive outcomes. The same was true among abstracts identified in a focused PubMed search.

The organization of this paper is as follows: In Section 2 an illustrative example is presented. In Section 3, we present fixed-effects and random-effects approaches for the meta-analysis of competing endpoints for the commonly used effect sizes (odds ratio (OR), relative risk (RR), and risk difference (RD)); as well as a less commonly used effect size, namely the difference in the arcsin-transformed risks (ASD). Section 4 contains analyses of the example, and Section 5 provides a discussion of the limitations together with conclusions.

2. META-ANALYSIS EXAMPLE

2.1. Radiotherapy for early breast cancer

In women with early breast cancer (stages I–III of the disease) addition of radiotherapy to an intervention (mastectomy with axillary lymph node clearance, mastectomy with axillary node sampling, mastectomy alone or lumpectomy with axillary node clearance) has been compared with the intervention without radiotherapy in randomized trials. The mutually exclusive outcomes of 'death secondary to causes other than breast cancer' and 'death secondary to breast cancer' were

DOI: 10.1002/sim

Table I. Data from a meta-analysis on radiotherapy in early breast cancer.

	Radiotherapy and mastectomy with axillary node clearance				Mastectomy with axillary node clearance			
(ID) Trial	N	n_{BC}	n_{noBC}	n_{alive}	N	n_{BC}	n_{noBC}	n_{alive}
(A) 16 smaller trials*	1252	591	150	511	1257	615	97	545
(B) BCCA Vancouver	164	59	1	104	154	70	2	82
(C) ECOG EST3181	171	85	8	78	161	75	12	74
(D) Oslo Co-60	278	85	100	93	285	100	68	117
(E) Oslo X-ray	285	112	90	83	267	110	72	85
(F) Swedish BCG	386	140	36	210	382	146	27	209
(G) SASIB	186	79	9	98	191	71	3	117
(H) Stockholm	639	258	115	266	642	308	90	244

Note: The counts corresponding to the three health states sum to the total number of women randomized to the pertinent arm.
ID, identification code; N, total number of randomized women; n_{alive}, number of women who are alive; n_{BC}, number of breast cancer deaths; n_{noBC}, number of non-breast cancer deaths.
*The 16 smaller trials were treated as a single mega-study in the Cochrane systematic review.

examined among 20 175 women in 40 trials [6, 7]. Here, for simplicity of exposition we use the subgroup of these studies that used mastectomy with axillary lymph node clearance (total of 6700 randomized women) as an illustrative example. The data have been extracted from the corresponding Cochrane meta-analyses (Table I). We caution the reader that this re-analysis is only illustrative and does not aim to re-interpret the findings of the Cochrane reviewers, who used individual participant data and different analyses to draw inferences.

3. METHODS

Suppose that there are k studies and m mutually exclusive outcomes (describing $m+1$ mutually exclusive and exhaustive health states), and assume that all outcomes across the k studies have been assessed at comparable follow-up times, and that their definitions allow a clinically meaningful quantitative synthesis.

Table II may help to fix notation. Here, p_{ij}^{E} and p_{ij}^{C} denote the proportions from study (trial) i $(i=1,\ldots,k)$ on the jth outcome measure $(j=1,\ldots,m)$ for the experimental and comparator groups, respectively. In Table II, the d_{ij} estimate the corresponding population effect sizes δ_{ij} that compare the proportions in the experimental and comparator arms, and could be the (log) OR, (log) RR, the RD, or the ASD. For example, the log odds ratio (log OR) would be $\log(p_{ij}^{\text{E}}/(1-p_{ij}^{\text{E}}))-\log(p_{ij}^{\text{C}}/(1-p_{ij}^{\text{C}}))$; the formulae for the other effect sizes are well known, and for completeness they are presented in Appendix A. For a discussion of these see, e.g. Normand [8]. For simplicity, we assume that there are no study-level covariates of interest.

In order to clarify the ultimate multivariate random-effects model, we first present the univariate extension of fixed to random effects and then develop the multivariate versions.

DOI: 10.1002/sim

Table II. Notation.

Study or trial	Experimental arm $E_1, \ldots, E_m$	Comparator arm $C_1, \ldots, C_m$	Effect sizes $d_1, \ldots, d_m$
1	$p^{\mathrm{E}}_{11}, \ldots, p^{\mathrm{E}}_{1m}$	$p^{\mathrm{C}}_{11}, \ldots, p^{\mathrm{C}}_{1m}$	$d_{11}, \ldots, d_{1m}$
2	$p^{\mathrm{E}}_{21}, \ldots, p^{\mathrm{E}}_{2m}$	$p^{\mathrm{C}}_{21}, \ldots, p^{\mathrm{C}}_{2m}$	$d_{21}, \ldots, d_{2m}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	$p^{\mathrm{E}}_{k1}, \ldots, p^{\mathrm{E}}_{km}$	$p^{\mathrm{C}}_{k1}, \ldots, p^{\mathrm{C}}_{km}$	$d_{k1}, \ldots, d_{km}$

3.1. The univariate case

In the univariate case, each of the m mutually exclusive outcomes is analyzed separately. For clarity, we retain the outcome index in the notation using a capital J to emphasize that a single outcome is considered. Because the outcomes are binary, their analysis is based on the binomial model.

3.1.1. Fixed effects. In a fixed model the true effect sizes (the population effect sizes), δ_{iJ}, are equal to a fixed value δ_{+J}. The effect size in the population is estimated by a weighted linear combination (convex combination) of the effect sizes d_{iJ}, where the non-negative weights W_{iJ} sum to unity. The weights that minimize the variance of the vector $\mathbf{d}_J = (d_{1J}, \ldots, d_{kJ})$ are inversely proportional to the variance of each study i, and the summary effect d_{+J} is

$$d_{+J} = W_{1J}d_{1J} + \cdots + W_{kJ}d_{kJ} = \frac{w_{1J}d_{1J} + \cdots + w_{kJ}d_{kJ}}{w_{1J} + \cdots + w_{kJ}} = \frac{d_{1J}/\upsilon_{1J} + \cdots + d_{kJ}/\upsilon_{kJ}}{1/\upsilon_{1J} + \cdots + 1/\upsilon_{kJ}} \tag{1}$$

where $W_{iJ} = w_{iJ}/(w_{1J} + \cdots + w_{kJ})$ and $w_{iJ} = \upsilon_{iJ}^{-1} = (\mathrm{var}(d_{iJ}))^{-1}$, $i = 1, \ldots, k$. Here, υ_{iJ} is the (conditional) variance of the study-level effect sizes, assumed to be known and calculated from the data. However, note that for any study, e.g. Study 5, the effect sizes $d_{51}, \ldots, d_{5m}$ for the m mutually exclusive outcomes are correlated.

Using normal approximations for d_{iJ} and assuming known variances (or consistent estimates for the variances) the weighted linear estimator (convex combination) of equation (1) is the maximum likelihood estimator of d_{+J} [9]. The same d_{+J} can also be obtained within a least squares regression framework [1, 9], which implies that d_{+J} is the minimum variance unbiased estimator.

3.1.2. Random effects. In the random-effects model, the true study-level effect sizes δ_{iJ} differ from their mean value δ^*_{+J} by ζ_{iJ}. The quantities ζ_{iJ} represent the random part of the treatment effects and follow a normal distribution with mean 0 and variance τ^2_J, which expresses between-study heterogeneity and is calculated from the data. The summary effect size for outcome J is obtained by an equation analogous to equation (1)

$$d^*_{+J} = \frac{w^*_{1J}d_{1J} + \cdots + w^*_{kJ}d_{kJ}}{w^*_{kJ} + \cdots + w^*_{kJ}} \tag{2}$$

where the weights become $w^*_{iJ} = 1/(\upsilon_{iJ} + \tau^2_J)$.

The computation of τ^2_J and d^*_{+J} is discussed in Section 3.3.2.

DOI: 10.1002/sim

3.2. The multivariate case

Because the different outcomes are mutually exclusive, their analysis follows a multinomial model (for the $m+1$ mutually exclusive health states).

3.2.1. Fixed effects. The estimates of the effect sizes in the multinomial model mimic equation (1) except that each $\mathbf{d}_i$ is now a row vector of effect sizes, and each variance becomes a covariance matrix $\mathbf{S}_i = \hat{\boldsymbol{\Sigma}}_i$, where $\boldsymbol{\Sigma}_i$ denote the true covariance matrices and $\mathbf{S}_i$ (or $\hat{\boldsymbol{\Sigma}}_i$) its estimate. For notational simplicity, let $\boldsymbol{\Lambda}_i = \boldsymbol{\Sigma}_i^{-1}$ denote the inverse of $\boldsymbol{\Sigma}_i$, and $\hat{\boldsymbol{\Lambda}}_i = \mathbf{S}_i^{-1}$ be the corresponding estimator. We define $\hat{\boldsymbol{\Lambda}}_+ = \hat{\boldsymbol{\Lambda}}_1 + \cdots + \hat{\boldsymbol{\Lambda}}_k$ to be the sum of the inverse covariance matrices. Then, the counterpart of equation (1) becomes

$$\mathbf{d}_+ = (\mathbf{d}_1 \hat{\boldsymbol{\Lambda}}_1 + \cdots + \mathbf{d}_k \hat{\boldsymbol{\Lambda}}_k) \hat{\boldsymbol{\Lambda}}_+^{-1} \tag{3}$$

with $\hat{\text{cov}}(\mathbf{d}_+) = (\hat{\boldsymbol{\Lambda}}_+)^{-1}$.

Appendix A shows the variances and covariances for the log RR, the log OR, the RD, and the ASD.

3.2.2. Random effects. To extend the aforementioned multivariate fixed-effects method to allow for random treatment effects, let $\boldsymbol{\zeta}_i = (\zeta_{i1}, \ldots, \zeta_{im})$ be a row vector of m random effects that follow a multivariate (m-variate) normal distribution with zero mean vector, $\mathbf{0}$, and covariance matrix $\mathbf{T}$. The covariance matrix $\mathbf{T}$ of the random effects is assumed to be known (or is estimated by a consistent estimator) and is calculated from the data as described in Section 3.3.2. Proceeding as in the univariate case, define $\boldsymbol{\Lambda}_i^* = (\boldsymbol{\Sigma}_i + \mathbf{T})^{-1}$ and $\boldsymbol{\Lambda}_+^* = \boldsymbol{\Lambda}_1^* + \cdots + \boldsymbol{\Lambda}_k^*$, with the corresponding estimators being $\hat{\boldsymbol{\Lambda}}_i^* = (\mathbf{S}_i + \hat{\mathbf{T}})^{-1}$ and $\hat{\boldsymbol{\Lambda}}_+^* = \hat{\boldsymbol{\Lambda}}_1^* + \cdots + \hat{\boldsymbol{\Lambda}}_k^*$. The counterpart of equation (3) becomes

$$\mathbf{d}_+^* = (\mathbf{d}_1 \hat{\boldsymbol{\Lambda}}_1^* + \cdots + \mathbf{d}_k \hat{\boldsymbol{\Lambda}}_k^*)(\hat{\boldsymbol{\Lambda}}_+^*)^{-1} \tag{4}$$

with $\hat{\text{cov}}(\mathbf{d}_+^*) = (\hat{\boldsymbol{\Lambda}}_+^*)^{-1}$.

3.3. Computations

3.3.1. Fixed effects—GLS estimation. Equations (1) and (3) are explicit. To obtain their solution by GLS we proceed as described by Gleser and Olkin [1]. Briefly, we need to fit the regression

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{5}$$

where the effect sizes for all outcomes are rearranged (stacked) in a $km \times 1$ vector $\mathbf{y}$, $\mathbf{X}$ is the design matrix of the regression with km rows and m columns, and $\mathbf{e}$ is a column vector of random errors. In equation (5), $\boldsymbol{\beta}$ is an $m \times 1$ vector of the summary effect sizes (note that no covariates have been specified). More explicitly

$$\mathbf{y}^{\mathrm{T}} = (\mathbf{y}_1^{\mathrm{T}}, \ldots, \mathbf{y}_k^{\mathrm{T}}) = \left(\underbrace{d_{11}, \ldots, d_{1m}}_{\text{study } 1}, \ldots, \underbrace{d_{k1}, \ldots, d_{km}}_{\text{study } k} \right)$$

DOI: 10.1002/sim

and

$$\mathbf{X}=\begin{bmatrix}\mathbf{X}_1\\ \hline \vdots \\ \hline \mathbf{X}_k\end{bmatrix}=\left[\begin{array}{ccc}1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & 1\\ \hline \vdots & \ddots & \vdots\\ \hline 1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & 1\end{array}\right]$$

The notation $\mathbf{y}^{\mathrm{T}}$ denotes the transpose of $\mathbf{y}$ and is a row vector. The $km\times km$ block diagonal matrix $\mathbf{S}$

$$\mathbf{S}=\mathrm{diag}(\mathbf{S}_1,\ldots,\mathbf{S}_k) \tag{6}$$

is the fixed-effects estimate of the covariance matrix of $\mathbf{y}$. In the univariate case the off-diagonal elements of each $\mathbf{S}_i$ are equal to 0, whereas in the multivariate case the full $\mathbf{S}_i$ matrix is used.

The solution to equation (5) is given by

$$\mathbf{d}_+=\hat{\boldsymbol{\beta}}=(\mathbf{X}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{y} \tag{7}$$

whereas

$$\hat{\boldsymbol{\Psi}}=(\mathbf{X}^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{X})^{-1} \tag{8}$$

estimates the covariance matrix $\boldsymbol{\Psi}$ of $\hat{\boldsymbol{\beta}}$.

A heterogeneity (goodness-of-fit) statistic can be estimated in a standard manner by

$$Q=(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})^{\mathrm{T}}\mathbf{S}^{-1}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}) \tag{9}$$

where Q has a chi-square distribution with c_1-c_2 degrees of freedom under the null hypothesis; here $c_1=km$ is the dimension of $\mathbf{y}$ and $c_2=m$ is the dimension of $\hat{\boldsymbol{\beta}}$. It is well appreciated that heterogeneity testing should not dictate decisions on statistical model choice [10–12]. Clinical and methodological heterogeneity are always present and they may or may not result in a statistically significant heterogeneity statistic, Q [13–15].

3.3.2. Random effects. The random-effects model fits the equation $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\zeta}+\mathbf{e}$, where $\boldsymbol{\zeta}$ is now a column vector of random effects. We fit this regression model using two methods proposed by Berkey *et al.* [3].

Iteratively reweighted least squares: Here we alternate in estimating $\mathbf{T}$ (the covariance matrix for the random effects) and $\boldsymbol{\beta}$. We begin the first iteration by setting $\hat{\boldsymbol{\beta}}_{(1)}$ (i.e. the estimate of $\boldsymbol{\beta}$ at the first iteration) equal to the estimate from the fixed-effects model in equation (6). The corresponding $\hat{\mathbf{T}}_{(1)}$ is an $m\times m$ matrix with all elements 0.

DOI: 10.1002/sim

At each subsequent iteration n, we compute the following steps:
Step 1:

$$\mathbf{S}^*_{(n)} = \text{diag}(\mathbf{S}_1 + \hat{\mathbf{T}}_{(n-1)}, \ldots, \mathbf{S}_k + \hat{\mathbf{T}}_{(n-1)})$$

Step 2:

$$\hat{\boldsymbol{\beta}}_{(n)} = (\mathbf{X}^{\mathrm{T}} \mathbf{S}^{*-1}_{(n)} \mathbf{X})^{-1} \mathbf{X}^{\mathrm{T}} \mathbf{S}^{*-1}_{(n)} \mathbf{y}$$

Step 3:

$$\hat{\mathbf{T}}_{(n)} = \underbrace{\frac{1}{k-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{(n)})^{\mathrm{T}} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{(n)})}_{\text{total variance}} - \underbrace{\frac{1}{k} \sum_{i=1}^{k} (\mathbf{S}_i)}_{\text{conditional variance}}$$

Note that we have rearranged the $km \times 1$ vector of the residuals $(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$ as a $k \times m$ matrix, so that residuals from the same study are in the same row and from the same outcome in the same column.

Step 4: Convergence is achieved when all elements of $\hat{\mathbf{T}}_{(n)}$ have changed less than a small value (e.g. 10^{-6}) compared with $\hat{\mathbf{T}}_{(n-1)}$. If convergence has been achieved, the iterations stop. If $\hat{\mathbf{T}}_{(n)}$ is not positive definite, $\hat{\mathbf{T}} \equiv \mathbf{0}$; otherwise set $\hat{\mathbf{T}} = \hat{\mathbf{T}}_{(n)}$.

Hedges and Olkin method: Hedges and Olkin [9] described the corresponding method in the univariate case. For the calculations in the univariate case we keep the off-diagonal elements of $\mathbf{S}_i$ and $\hat{\mathbf{T}}_{(n)}$ equal to 0 in the aforementioned algorithm. If any diagonal element of $\hat{\mathbf{T}}_{(n)}$ becomes negative during the process, it is set equal to 0. We stop the iterations when all diagonal elements of $\hat{\mathbf{T}}_{(n)}$ have changed less than a small value (e.g. 10^{-6}) since the previous iteration.

Maximum likelihood: Berkey *et al.* [3] propose a maximum likelihood estimate for $\mathbf{T}$ in the multivariate normal case. The algorithm is similar to the steps described above, only that the 3rd step becomes

$$\hat{\mathbf{T}}_{(n)} = \frac{1}{k} \sum_{i=1}^{k} [\hat{\mathbf{T}}_{(n-1)} (\mathbf{S}_i + \hat{\mathbf{T}}_{(n-1)})^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}) (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})^{\mathrm{T}} (\mathbf{S}_i + \hat{\mathbf{T}}_{(n-1)})^{-1} \hat{\mathbf{T}}_{(n-1)}$$

$$+ \hat{\mathbf{T}}_{(n-1)} - \hat{\mathbf{T}}_{(n-1)} (\mathbf{S}_i + \hat{\mathbf{T}}_{(n-1)})^{-1} \hat{\mathbf{T}}_{(n-1)}]$$

where $\mathbf{y}_i$ is a column vector of the m outcome responses in the ith study, and $\mathbf{X}_i$ is the corresponding design matrix, as shown above in the description of the fixed-effects GLS approach. We stop the iterations when convergence has been achieved.

For the corresponding estimation in the univariate case we would again set the off-diagonal elements of $\mathbf{S}_i$ and $\hat{\mathbf{T}}_{(n)}$ equal to 0. We apply the same stopping rules as in the univariate case for iteratively reweighted least squares.

DerSimonian and Laird method: For comparison, and only in the univariate case, we also used the DerSimonian and Laird method [16]. In the notation we retain the outcome index, but we use a capital J to emphasize that we deal with a single outcome. Here, we calculate for each

DOI: 10.1002/sim

outcome the quantity $Q_J = \sum_{i=1}^{k} w_{iJ}(d_{iJ} - d_{+J})^2$ (quantities as described in Section 3.1.1). The between-study heterogeneity τ_J^2 is calculated as

$$\tau_J^2 = \max\left(0, (Q_J - k + 1) \Big/ \left(\sum_{i=1}^{k} w_{iJ} - \sum_{i=1}^{k} (w_{iJ}^2) \Big/ \sum_{i=1}^{k} w_{iJ}\right)\right)$$

The summary effect size is calculated as described in Section 3.2.1.

3.3.3. Confidence intervals in the univariate case. In the univariate case the confidence intervals for the effect size of outcome J are

$$\left(\hat{\beta}_J - z_{\alpha/2}\sqrt{\hat{\mathrm{var}}(\hat{\beta}_J)},\ \hat{\beta}_J + z_{\alpha/2}\sqrt{\hat{\mathrm{var}}(\hat{\beta}_J)}\right) \tag{10}$$

where $z_{\alpha/2}$ is the corresponding percentile of the normal distribution. The estimated variance of $\hat{\beta}_J$ for the fixed- and the random-effects models is the inverse of the sum of the corresponding weights described in Sections 3.1.1 and 3.2.1.

As mentioned before, because the univariate case ignores the negative correlations between the m outcomes, confidence intervals obtained from equation (10) are not simultaneously correct. One can use the Boole–Bonferroni inequality as a simple (but conservative) adjustment to control type I error: $100(1-\alpha)$ per cent confidence intervals are obtained by substituting $z_{\alpha/2m}$ for $z_{\alpha/2}$ in equation (10). In practice, in published meta-analyses adjustments for multiple comparisons are not often performed. Here we report results for the univariate case with and without adjustments for multiple comparisons to allow the reader to appreciate the differences in the worked example.

3.3.4. Simultaneous confidence intervals for the multivariate case. To obtain simultaneous confidence intervals for the elements $\beta_1, \ldots, \beta_m$ or linear combinations $L(\mathbf{a}) = (a_1\beta_1 + \cdots + a_m\beta_m)$ thereof, let $\mathbf{a} = (a_1, \ldots, a_m)$ and $\hat{L}(\mathbf{a}) = (a_1\hat{\beta}_1 + \cdots + a_m\hat{\beta}_m)$. Then the $100(1-\alpha)$ per cent simultaneous confidence intervals for $L(\mathbf{a})$ are given by

$$\left(\hat{L}(\mathbf{a}) - c_\alpha\sqrt{\mathbf{a}^{\mathrm{T}}\hat{\boldsymbol{\Psi}}\mathbf{a}},\ \hat{L}(\mathbf{a}) + c_\alpha\sqrt{\mathbf{a}^{\mathrm{T}}\hat{\boldsymbol{\Psi}}\mathbf{a}}\right) \tag{11}$$

where c_α is the square root of the $100(1-\alpha)$ percentile of the chi-square distribution with m degrees of freedom.

In particular, a confidence interval for β_i is obtained from (11):

$$\left(\hat{\beta}_i - c_\alpha\sqrt{\hat{\Psi}_{ii}},\ \hat{\beta}_i + c_\alpha\sqrt{\hat{\Psi}_{ii}}\right) \tag{12}$$

and a confidence interval for a difference $\beta_i - \beta_j$ is obtained from (11):

$$\left(\hat{\beta}_i - \hat{\beta}_j - c_\alpha\sqrt{\hat{\Psi}_{ii} + \hat{\Psi}_{jj} - 2\hat{\Psi}_{ij}},\ \hat{\beta}_i - \hat{\beta}_j + c_\alpha\sqrt{\hat{\Psi}_{ii} + \hat{\Psi}_{jj} - 2\hat{\Psi}_{ij}}\right) \tag{13}$$

DOI: 10.1002/sim

A note on confidence intervals in the univariate case: In the univariate case, in equations (11)–(13), the off-diagonal elements of $\hat{\mathbf{\Psi}}$ (i.e. the covariances $\hat{\Psi}_{ij}$) are 0, and c_α is equal to the $z_{\alpha/2}$ percentile from a normal distribution. Therefore, equation (12) is equivalent to equation (10) in the univariate setting.

We also note that the simultaneous confidence intervals in Section 3.3.4 are based on normal approximations. Exact simultaneous confidence intervals would require numerical computations (e.g. Monte Carlo-based permutation testing).

3.3.5. Adding covariates and dealing with studies reporting only some outcomes. It is relatively straightforward to add covariates in the current framework, in a manner that is directly analogous to the methods proposed in Hedges and Olkin [9, Chapter 6] and Berkey *et al.* [3].

The aforementioned methods (except for the maximum likelihood approach in Section 3.3.2) do not require that all studies have data on all outcomes. Gleser and Olkin [1] describe how one specifies $\mathbf{y}, \mathbf{X}$, and $\mathbf{S}$ when some studies do not have data on all outcomes. In addition, as noted by Berkey, the algorithm for the calculation of the iteratively reweighted least squares for the multivariate random-effects approach and the corresponding maximum likelihood approach become complicated.

3.3.6. Implementation. These approaches can be carried out in many programming environments such as Stata, SAS, Matlab (or Octave, its open-source equivalent), and R/S-plus. We developed routines in Octave [17] (ver. 2.9.9) and in R (ver. 2.4.1) [18] to handle the calculations.

3.4. Interpretation of effect sizes with the multivariate approach

Allowing for some caveats, the summary effect sizes from the multivariate approach are interpreted similarly to the corresponding summary effect sizes of the separate meta-analyses. An important point is that, in contrast to separate meta-analyses with binary outcomes, in the multinomial case one cannot reverse the definition of the outcomes, even for the OR metric. This is because after reversing the outcome definitions, the outcomes will always have common parts and are no longer mutually exclusive. For example, in the breast cancer meta-analysis, the mutually exclusive outcomes are 'death not related to breast cancer' and 'breast cancer deaths'. The corresponding reversely defined outcomes would be 'alive or lost to breast cancer' and 'alive or lost to causes other than breast cancer', which are clearly not disjoint.

It may be superfluous to clarify that the multinomial case of the multivariate framework is not similar to a multinomial logistic regression, even when one refers to multiplicative metrics. A multinomial (nominal, polytomous) logistic regression would contrast the odds of the health state of interest (outcome of interest) *versus* a reference/baseline category. In contrast, the multivariate approach always contrasts the presence of the outcome of interest *versus* its absence.

4. EXAMPLE: BREAST CANCER META-ANALYSIS

The worked example is presented in the following paragraphs. We perform analyses for the four metrics (OR, RR, RD, and ASD) with fixed- and random-effects models for both the univariate and the multivariate cases.

DOI: 10.1002/sim

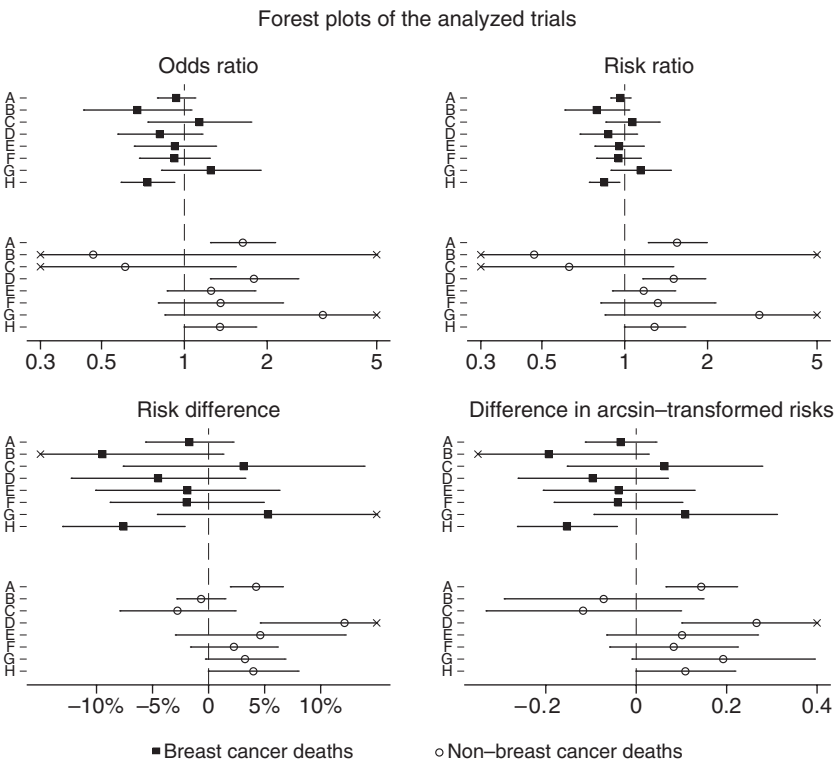

Figure 1. Forest plots of the analyzed trials. Shown are forest plots for the four metrics (the four panels) and for both mutually exclusive outcomes. Filled squares or open circles stand for the point estimates; the horizontal lines represent the corresponding 95 per cent confidence intervals. Some confidence intervals extend beyond the boundaries of the graph (small 'x' signs mark them). The dashed vertical line is the line of no effect. Trial IDs as described in Table I. Summary effect sizes are not depicted in this figure.

We also calculate the summary relative effects between the mutually exclusive outcomes, i.e. a linear combination of the summary effect sizes for the two outcomes. These are defined as follows for the four metrics: the ratio of the outcome-specific ORs (relative OR, ROR); the ratio of the outcome-specific risk ratios (relative RR, RRR); the difference between the outcome-specific RDs (difference of RDs, DRDs); and the differences between the outcome-specific arcsin-transformed risks (difference of ASD, DASD). Relative effects are greater than 1 (for ROR and RRR) or 0 (for DRD and DASD) when the risk of non-breast cancer deaths is greater than the risk of breast cancer deaths in the experimental arm compared with the controls.

4.1. Radiotherapy added to mastectomy with axillary node clearance

As shown in Table I, the Cochrane review merged 16 small trials into a single entry in the meta-analysis. For the purposes of this analysis we also consider these trials as a single mega-study. Thus, the eight 'trial entries' in the meta-analysis had very different sample sizes ranging from 318 to 1309 (for the aggregate of the '16 small studies'). The proportion of breast cancer deaths ranged from 1 to 36 per cent and 1 to 37 per cent in the arms with and without radiotherapy, respectively.

DOI: 10.1002/sim

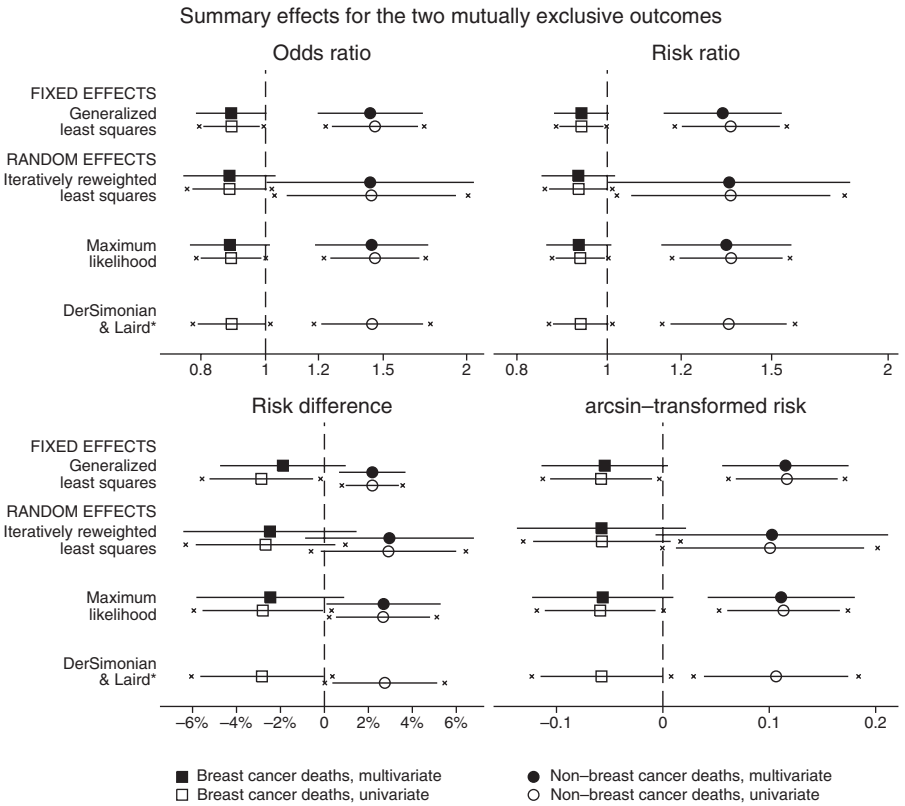

Figure 2. Summary effects for the two mutually exclusive outcomes. This figure is composed of four panels corresponding to the four metrics, and juxtaposes summary effect sizes from multivariate and univariate analyses with fixed- and random-effects models. The horizontal lines represent the calculated 95 per cent confidence intervals. In all multivariate analyses simultaneous confidence intervals have been calculated. In univariate analyses the horizontal lines denote confidence intervals uncorrected for multiple comparisons. The breadth of the corresponding Bonferroni-corrected confidence intervals is marked with small 'x' markers. The dashed line is the line of no effect. The iteratively reweighted least squares method is the Hedges and Olkin model in the univariate case, and the Berkey model in the multivariate case. We did not calculate the multivariate extension of the DerSimonian and Laird random-effects model.

The corresponding proportions for non-breast cancer deaths ranged from 31 to 50 per cent and 30 to 53 per cent. Figure 1 shows the forest plots of the individual studies for both mutually exclusive outcomes and all four metrics.

4.1.1. Effect sizes for breast cancer and non-breast cancer deaths. Overall, addition of radiotherapy has a protective effect from breast cancer deaths, but at the same time predisposed towards non-breast cancer deaths (Figure 2). For all four metrics, the point estimates for the outcome-specific effect sizes were largely similar across all analyses (fixed- and different random-effects calculations, univariate and multivariate). For the RR and OR metrics, the point estimate for non-breast cancer mortality became somehow smaller in the fixed-effects multivariate analyses compared with the respective univariate analyses. In contrast, the corresponding point estimates for

DOI: 10.1002/sim

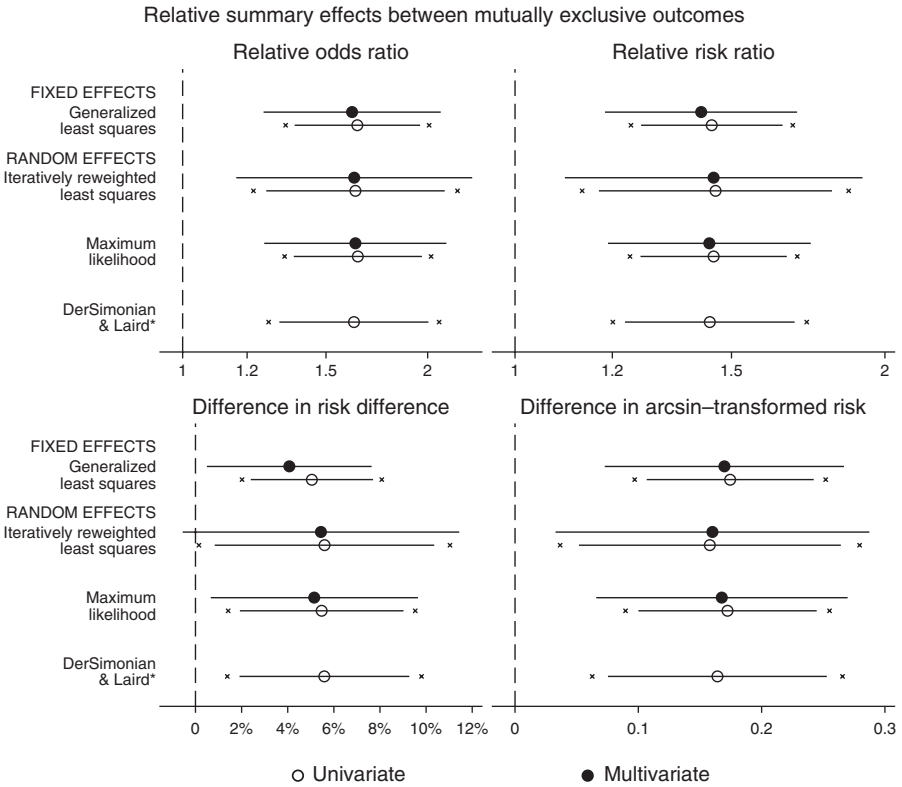

Figure 3. Relative summary effects between mutually exclusive outcomes. The layout is similar to Figure 2. Relative odds ratio and relative risk ratio more than 1, and difference in risk difference and difference in the arcsin-transformed risks more than 0 imply that the risk of non-breast cancer deaths is greater than the risk of breast cancer deaths in the experimental arm compared with the controls. The iteratively reweighted least squares method is the Hedges and Olkin model in the univariate case, and the Berkey model in the multivariate case. We did not calculate the multivariate extension of the DerSimonian and Laird random-effects model.

breast cancer deaths remained very similar in univariate and multivariate analyses. This was not as evident in random-effects analyses. For the RD metric, multivariate *versus* univariate fixed-effects analyses yielded smaller point estimates for breast cancer deaths (with virtually no changes for non-breast cancer deaths).

As expected, multivariate analyses (with simultaneous confidence intervals) and random-effects models generally resulted in wider confidence intervals (Figure 2). Overall, between-study heterogeneity testing was significant for the RD in univariate analyses for the non-breast cancer death outcome, and for the RD and ASD in the multivariate approach (Appendix B, Tables BI–BIII).

For all four metrics, univariate fixed-effects analyses resulted in statistically significant inferences on both mutually exclusive outcomes, even after Bonferroni corrections for multiple (two) comparisons. For breast cancer deaths, the simultaneous confidence intervals from fixed-effects multivariate analyses crossed the line of no effect. This was more evident for the RD metric, and secondarily for the ASD (these metrics had evidence for statistically significant between-study heterogeneity in some analyses).

DOI: 10.1002/sim

Random-effects analyses yielded wider confidence intervals compared with the respective fixed-effects models. For all four metrics and in both univariate and multivariate analyses, iteratively reweighted least squares resulted in the broadest confidence intervals. It is interesting to note that multivariate random effects with iteratively reweighted least squares resulted in loss of formal statistical significance for non-breast cancer deaths for the RD and ASD metrics, and marginal inferences for OR and RR (Figure 2, Appendix B, Tables BI–BIII).

4.1.2. Relative summary effect sizes between breast cancer and non-breast cancer deaths. Figure 3 and Tables BI–BIII show the relative effect sizes (ROR, RRR, DRD, DASD) between the two outcomes. The direction of the relative effects implied that the risk for breast cancer deaths is smaller than the risk for non-breast cancer deaths in treated individuals compared with controls. Point estimates for summary relative effect sizes were somehow smaller in multivariate analyses compared with the corresponding univariate analyses (more evident for the DRD in Figure 3). Multivariate analyses resulted in wider confidence intervals compared with the respective univariate analyses, even when Bonferroni corrections were used for the latter. As was observed for the outcome-specific summary effect sizes, random-effects analyses with iteratively reweighted least squares resulted in wider confidence intervals.

In this particular example, multivariate analyses yielded more conservative confidence intervals compared with the corresponding univariate analyses. The loss of statistical significance in the absolute reduction of breast-cancer-related deaths (RD and ASD) when the negative correlations were taken into account would be viewed with skepticism in an actual analysis, because of the lack of fit of the pertinent models. However, it is interesting to note as an exhibit that inferential changes might very well occur. Because clinical and methodological heterogeneities are abundant, in an actual meta-analysis, random-effects approaches would be preferable.

5. DISCUSSION AND CONCLUSION

We have described approaches that take into account the negative correlations between mutually exclusive outcomes. Using data from a typical example, we compared univariate meta-analyses (separate meta-analyses for each outcome) with the multivariate approach for both fixed and random treatment effects. In the example, changes in the point estimates of the outcome-specific risks with specific metrics (RD and ASD) were evident, albeit of dubious clinical significance. However, it is certainly likely that clinically significant differences compared with univariate analyses might occur in other instances. Moreover, the multivariate approach estimates asymptotically correct confidence intervals for linear combinations (relative effects) of the analyzed mutually exclusive outcomes.

Multivariate meta-analysis is being increasingly used. Gleser and Olkin [1] describe a fixed effects GLS model for stochastically dependent endpoints, which was extended to a linear mixed effects regression by Berkey *et al.* [2]. Van Houwelingen *et al.* [4] in their tutorial for advanced meta-analysis and Arends in a thesis dissertation [19] explain how to fit such a mixed model with (restricted) maximum likelihood in SAS Proc MIXED. Riley *et al.* [20] also argue in favor of a multivariate random-effects meta-analysis rather than separate univariate analyses. However, in none of the aforementioned papers did the authors address mutually exclusive outcomes. The uniqueness of this setting is that the covariances of the synthesized effect sizes can be calculated from the data. This is generally not the case for other types of stochastically dependent outcomes [1, 20, 21]. Finally, Dear [22] described a fixed-effects multivariate linear model in which

DOI: 10.1002/sim

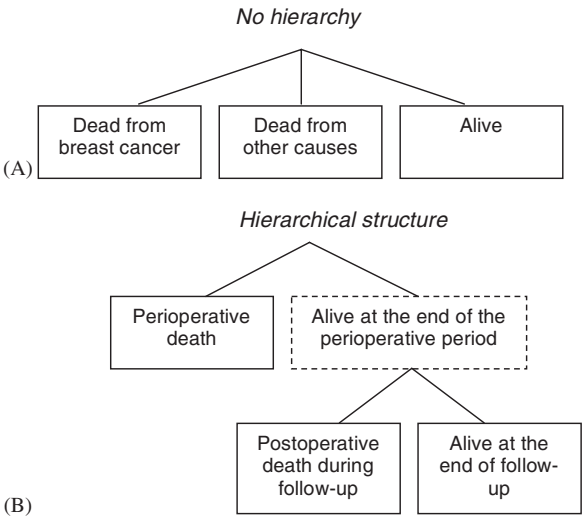

Figure 4. Distinction between hierarchical structures at the same level and at different levels. Both panels of this figure represent health states that are mutually exclusive and exhaustive, and pertain to corresponding outcomes. A hierarchical structure is implied. In panel A all health states/outcomes of interest belong to the same hierarchical level. In panel B health states/outcomes in the second level are conditioned on an implied health state in the upper level of hierarchy, and they cannot affect the outcomes in their parent hierarchical level.

correlations in the survival proportions at different follow-up times are estimated iteratively from the data. (However, Dear's approach [22] does not account for the negative correlations between mutually exclusive outcomes.)

5.1. A note on different categories of mutually exclusive outcomes

Depending on the type of end-point (and study definitions), one can distinguish two broad categories of disjoint outcomes: those that are disjoint at the same hierarchical level and those that belong to different hierarchical levels (Figure 4). In this study we explore the analysis of outcomes that belong to the first category. The meta-analysis of outcomes that belong to the second category constitutes a different problem and is not pursued here.

5.1.1. Mutually exclusive outcomes at the same hierarchical level. The worked example was a case with two outcomes that are mutually exclusive at the same hierarchical level. Indeed many instances of cause-specific mortality would fall under this category. Changes in the rate of any of the mutually exclusive outcomes affect the maximum possible counts of all other outcomes. The prospective likelihood for these outcomes is appropriately modeled by a multinomial distribution.

5.1.2. Disjoint outcomes that belong to different hierarchical levels. Consider a study that evaluated deaths among patients who are followed for a year after a major surgery and—to make the example very obvious—suppose that the study counts how many people do something trivial

DOI: 10.1002/sim

during the last month of the follow-up (such as go out for a beer more than two times per week). The outcomes of 'perioperative death', 'being alive and going for a beer', and 'being alive and not going for a beer' are disjoint, but they belong to two different hierarchical levels. Fluctuations in the number of people performing the trivial task (outcome in the lower hierarchical level) cannot impact on the number of people who die perioperatively (outcome in the higher hierarchical level). In this case, the prospective likelihood of the disjoint outcomes is more appropriately modeled as (a series of) conditional distributions.

5.2. Usefulness of meta-analyses of mutually exclusive outcomes

Most meta-analyses of mutually exclusive outcomes generally pertain to cause-specific mortality [7, 23–25]. Meta-analysis of cause-specific mortality is also challenging for non-statistical (meta-epidemiological) reasons that undermine its validity and applicability. First of all, there is a wide margin for misclassifications, because the exact cause of death is not easy to deduce [26]. Second, definitions of cause-specific mortality might be quite different across trials, rendering a summary estimate difficult to interpret. Third, cause-specific deaths may be selectively reported when a difference was found, and selectively non-reported in the opposite case (outcome reporting bias) [27–29].

The above is not only a theoretical concern: Early meta-analyses found increased risk for non-illness mortality (suicide and deaths from accidents and violence) and decreased risk for illness-related mortality among patients receiving lipid-lowering interventions (diet, medication, or surgery) *versus* controls [30–34]. Later analyses that incorporated data from subsequent trials in which cause-specific mortality was better reported concluded that the aforementioned findings were misleading [33].

5.3. When is the multivariate approach preferable to separate meta-analyses?

When the covariances of the synthesized effect sizes are very small, accounting for between-outcome correlations or ignoring them would result in estimates that are similar for any practical purpose. However, although the estimates are similar, the confidence intervals differ. Depending on the metric used this would likely be true for large sample sizes or event rates that are closer to 0 (or 1). Especially for small sample sizes the covariances could be substantial. In any case, we argue that it is good practice to account for the between-outcome correlations in all cases, even as a secondary analysis.

An interesting caveat pertains to meta-analyses that evaluate *only one* of several mutually exclusive outcomes (e.g. evaluation of arrhythmic deaths in a meta-analysis on the use or not of implantable cardioverter defibrillators [23, 24], without assessing non-arrhythmic deaths explicitly). It is definitely possible that the summary estimate of interest can be quite different in magnitude or change statistical significance status in a univariate approach *versus* a multivariate approach (that accounts for the remaining mutually exclusive outcomes). The interpretation of the meta-analysis findings would probably be more conservative in the presence of discrepant univariate and multivariate analyses.

5.4. Conclusion

We consider it good practice to analyze mutually exclusive outcomes in the multinomial setting. We caution that this pertains to meta-analyses that focus on the marginal effect of only one out

DOI: 10.1002/sim

of several mutually exclusive outcomes. Although for large sample sizes or extreme proportion of events the results might practically be the same with those obtained from separate meta-analyses, we believe that this should at least be ascertained in secondary analyses.

APPENDIX A: EFFECT SIZES AND CORRESPONDING COVARIANCE MATRICES FOR MUTUALLY EXCLUSIVE OUTCOMES

We provide formulae for the variances and the covariances of the effect sizes in the multinomial case. Note that the variances have the usual formulae for the binomial case. To fix notation suppose that there are two outcome measures for each study; the extension to more than two outcomes will be clear after the example. As stated in the text, let the superscripts E and C denote the experimental and the comparator arms, respectively, and the indices 1 and 2 the two mutually exclusive outcomes. The vectors that hold the proportions of events in the two arms are $(p_1^{\mathrm{E}}, p_2^{\mathrm{E}})$ and $(p_1^{\mathrm{C}}, p_2^{\mathrm{C}})$, respectively. Because we focus on a single study, the study index i is omitted.

Users who wish to program the analysis should take due diligence to minimize rounding errors by avoiding the numerical evaluation of expressions when deriving intermediate quantities and by using suitable numerical techniques. For this reason, the following formulae for the effect sizes and the covariance matrices should be re-expressed as functions of the counts in the pertinent 2-by-2 tables.

A.1. The risk difference

The vector of the estimators of the effect sizes for the two outcomes is

$$\mathbf{d} = (p_1^{\mathrm{E}} - p_1^{\mathrm{C}},\ \ p_2^{\mathrm{E}} - p_2^{\mathrm{C}})$$

and the corresponding estimator of the covariance matrix is

$$\mathbf{S} = \begin{bmatrix} \dfrac{p_1^{\mathrm{E}}(1-p_1^{\mathrm{E}})}{N^{\mathrm{E}}} + \dfrac{p_1^{\mathrm{C}}(1-p_1^{\mathrm{C}})}{N^{\mathrm{C}}} & -\dfrac{p_1^{\mathrm{E}} p_2^{\mathrm{E}}}{N^{\mathrm{E}}} - \dfrac{p_1^{\mathrm{C}} p_2^{\mathrm{C}}}{N^{\mathrm{C}}} \\ & \dfrac{p_2^{\mathrm{E}}(1-p_2^{\mathrm{E}})}{N^{\mathrm{E}}} + \dfrac{p_2^{\mathrm{C}}(1-p_2^{\mathrm{C}})}{N^{\mathrm{C}}} \end{bmatrix}$$

The element $s_{12} = s_{21}$.

A.2. The log risk ratio

The vector of the estimators of the effect sizes for the two outcomes is

$$\mathbf{d} = (\log p_1^{\mathrm{E}} - \log p_1^{\mathrm{C}},\ \ \log p_2^{\mathrm{E}} - \log p_2^{\mathrm{C}})$$

DOI: 10.1002/sim

and the corresponding estimator of the covariance matrix is

$$\mathbf{S}=\begin{bmatrix} \dfrac{(1-p_1^{\mathrm{E}})}{p_1^{\mathrm{E}}N^{\mathrm{E}}}+\dfrac{(1-p_1^{\mathrm{C}})}{p_1^{\mathrm{C}}N^{\mathrm{C}}} & -\dfrac{1}{N^{\mathrm{E}}}-\dfrac{1}{N^{\mathrm{C}}} \\ & \dfrac{(1-p_2^{\mathrm{E}})}{p_2^{\mathrm{E}}N^{\mathrm{E}}}+\dfrac{(1-p_2^{\mathrm{C}})}{p_2^{\mathrm{C}}N^{\mathrm{C}}} \end{bmatrix}$$

The element $s_{12}=s_{21}$.

A.3. The log odds ratio

The vector of the estimators of the effect sizes for the two outcomes is

$$\mathbf{d}=(\log(p_1^{\mathrm{E}}/(1-p_1^{\mathrm{E}}))-\log(p_1^{\mathrm{C}}/(1-p_1^{\mathrm{C}})),\ \ \log(p_2^{\mathrm{E}}/(1-p_2^{\mathrm{E}}))-\log(p_2^{\mathrm{C}}/(1-p_2^{\mathrm{C}})))$$

and the corresponding estimator for the covariance matrix is

$$\mathbf{S}=\begin{bmatrix} \dfrac{1}{N^{\mathrm{E}}p_1^{\mathrm{E}}(1-p_1^{\mathrm{E}})}+\dfrac{1}{N^{\mathrm{C}}p_1^{\mathrm{C}}(1-p_1^{\mathrm{C}})} & -\dfrac{1}{N^{\mathrm{E}}(1-p_1^{\mathrm{E}})(1-p_2^{\mathrm{E}})}-\dfrac{1}{N^{\mathrm{C}}(1-p_1^{\mathrm{C}})(1-p_2^{\mathrm{C}})} \\ & \dfrac{1}{N^{\mathrm{E}}p_2^{\mathrm{E}}(1-p_2^{\mathrm{E}})}+\dfrac{1}{N^{\mathrm{C}}p_2^{\mathrm{C}}(1-p_2^{\mathrm{C}})} \end{bmatrix}$$

The element $s_{12}=s_{21}$.

A.4. The difference in the arcsin-transformed risks

The vector of the estimators of the effect sizes for the two outcomes is

$$\mathbf{d}=(2\arcsin\sqrt{p_1^{\mathrm{E}}}-2\arcsin\sqrt{p_1^{\mathrm{C}}},\ \ 2\arcsin\sqrt{p_2^{\mathrm{E}}}-2\arcsin\sqrt{p_2^{\mathrm{C}}})$$

and the corresponding estimator of the covariance matrix is

$$\mathbf{S}=\begin{bmatrix} \dfrac{1}{N^{\mathrm{E}}}+\dfrac{1}{N^{\mathrm{C}}} & -\dfrac{\sqrt{p_1^{\mathrm{E}}p_2^{\mathrm{E}}}}{N^{\mathrm{E}}\sqrt{(1-p_1^{\mathrm{E}})(1-p_2^{\mathrm{E}})}}-\dfrac{\sqrt{p_1^{\mathrm{C}}p_2^{\mathrm{C}}}}{N^{\mathrm{C}}\sqrt{(1-p_1^{\mathrm{C}})(1-p_2^{\mathrm{C}})}} \\ & \dfrac{1}{N^{\mathrm{E}}}+\dfrac{1}{N^{\mathrm{C}}} \end{bmatrix}$$

The element $s_{12}=s_{21}$.

APPENDIX B: WORKED EXAMPLE RESULTS

Univariate analyses without adjustment and with Bonferroni adjustments for multiple comparisons are summarized in Tables BI and BII, respectively. Multivariate analyses with simultaneous confidence intervals are summarized in Table BIII.

DOI: 10.1002/sim

Table BI. Univariate analyses without adjustment for multiple comparisons.

Model and metric	Breast cancer deaths		Non-breast cancer deaths		
	Effect (95 per cent CI)	P_{fit}	Effect (95 per cent CI)	P_{fit}	Relative effects* (95 per cent CI)
Fixed effects, generalized least squares					
log OR	−0.117 (−0.215, −0.020)	0.265	0.377 (0.229, 0.525)	0.288	0.494 (0.317, 0.672)
log RR	−0.064 (−0.118, −0.010)	0.249	0.305 (0.184, 0.426)	0.285	0.369 (0.236, 0.501)
RD (per cent)	−2.871 (−5.229, −0.514)	0.265	2.178 (0.967, 3.388)	0.003	5.049 (2.399, 7.699)
ASD	−0.058 (−0.106, −0.010)	0.265	0.116 (0.069, 0.164)	0.109	0.175 (0.107, 0.242)
Random effects, iteratively reweighted least squares (Hedges and Olkin method)					
log OR	−0.124 (−0.252, 0.003)	—	0.364 (0.073, 0.656)	—	0.489 (0.236, 0.741)
log RR	−0.071 (−0.144, 0.002)	—	0.305 (0.059, 0.550)	—	0.375 (0.157, 0.594)
RD (per cent)	−2.678 (−5.855, 0.498)	—	2.915 (−0.166, 5.996)	—	5.593 (0.833, 10.353)
ASD	−0.057 (−0.122, 0.007)	—	0.101 (0.012, 0.189)	—	0.158 (0.052, 0.264)
Random effects, maximum likelihood					
log OR	−0.119 (−0.224, −0.014)	—	0.377 (0.223, 0.530)	—	0.496 (0.314, 0.677)
log RR	−0.067 (−0.127, −0.006)	—	0.305 (0.178, 0.433)	—	0.372 (0.235, 0.509)
RD (per cent)	−2.806 (−5.547, −0.066)	—	2.667 (0.525, 4.808)	—	5.473 (1.926, 9.020)
ASD	−0.059 (−0.111, −0.007)	—	0.113 (0.060, 0.166)	—	0.172 (0.100, 0.245)
Random effects, DerSimonian and Laird					
log OR	−0.117 (−0.234, −0.000)	—	0.367 (0.192, 0.543)	—	0.484 (0.274, 0.695)
log RR	−0.066 (−0.134, 0.003)	—	0.299 (0.156, 0.443)	—	0.365 (0.206, 0.524)
RD (per cent)	−2.844 (−5.649, −0.039)	—	2.747 (0.362, 5.133)	—	5.591 (1.909, 9.273)
ASD	−0.058 (−0.115, −0.000)	—	0.106 (0.039, 0.174)	—	0.164 (0.075, 0.253)

Note: The spurious accuracy of three decimal points is retained for the convenience of anyone wishing to repeat the examples.

ASD, difference in arcsin-transformed risks; CI, confidence interval; OR, odds ratio; P_{fit}, p-value from goodness-of-fit test for the fixed-effects model (see text); RD, risk difference; and RR, risk ratio.

*Log relative odds ratios, log relative risk ratios, difference of risk differences, or differences of differences in arcsin-transformed risk between the two outcomes. P_{fit} is not applicable here.

DOI: 10.1002/sim

Table BII. Univariate analyses with Bonferroni adjustments for multiple (two) comparisons.

	Breast cancer deaths		Non-breast cancer deaths		
Model and metric	Effect (95 per cent CI)	P_{fit}	Effect (95 per cent CI)	P_{fit}	Relative effects* (95 per cent CI)
Fixed effects, generalized least squares					
log OR	−0.117 (−0.229, −0.006)	0.265	0.377 (0.207, 0.547)	0.288	0.494 (0.291, 0.697)
log RR	−0.064 (−0.126, −0.002)	0.249	0.305 (0.167, 0.443)	0.285	0.369 (0.217, 0.520)
RD (per cent)	−2.871 (−5.567, −0.176)	0.265	2.178 (0.793, 3.562)	0.003	5.049 (2.019, 8.080)
ASD	−0.058 (−0.113, −0.003)	0.265	0.116 (0.062, 0.171)	0.109	0.175 (0.097, 0.252)
Random effects, iteratively reweighted least squares (Hedges and Olkin method)					
log OR	−0.124 (−0.271, 0.022)	—	0.364 (0.031, 0.698)	—	0.489 (0.200, 0.777)
log RR	−0.071 (−0.154, 0.013)	—	0.305 (0.024, 0.586)	—	0.375 (0.126, 0.625)
RD (per cent)	−2.678 (−6.311, 0.955)	—	2.915 (−0.609, 6.439)	—	5.593 (0.150, 11.037)
ASD	−0.057 (−0.131, 0.017)	—	0.101 (−0.000, 0.202)	—	0.158 (0.037, 0.279)
Random effects, maximum likelihood					
log OR	−0.119 (−0.239, 0.001)	—	0.377 (0.201, 0.552)	—	0.496 (0.288, 0.703)
log RR	−0.067 (−0.136, 0.003)	—	0.305 (0.160, 0.451)	—	0.372 (0.215, 0.529)
RD (per cent)	−2.806 (−5.941, 0.328)	—	2.667 (0.218, 5.116)	—	5.473 (1.417, 9.529)
ASD	−0.059 (−0.118, 0.001)	—	0.113 (0.053, 0.174)	—	0.172 (0.090, 0.255)
Random effects, DerSimonian and Laird					
log OR	−0.117 (−0.251, 0.017)	—	0.367 (0.167, 0.568)	—	0.484 (0.243, 0.726)
log RR	−0.066 (−0.144, 0.013)	—	0.299 (0.135, 0.463)	—	0.365 (0.183, 0.547)
RD (per cent)	−2.844 (−6.052, 0.364)	—	2.747 (0.019, 5.475)	—	5.591 (1.380, 9.802)
ASD	−0.058 (−0.123, 0.008)	—	0.106 (0.029, 0.184)	—	0.164 (0.063, 0.266)

Note: The spurious accuracy of three decimal points is retained for the convenience of anyone wishing to repeat the examples.

ASD, difference in arcsin-transformed risks; CI, confidence interval; OR, odds ratio; P_{fit}, p-value from goodness-of-fit test for the fixed-effects model (see text); RD, risk difference; and RR, risk ratio.

*Log relative odds ratios, log relative risk ratios, difference of risk differences, or differences of differences in arcsin-transformed risk between the two outcomes. P_{fit} is not applicable here.

DOI: 10.1002/sim

Table BIII. Multivariate analyses with simultaneous confidence intervals.

Model and metric	Breast cancer deaths, effect (95 per cent CI)	Non-breast cancer deaths, effect (95 per cent CI)	Relative effects* (95 per cent CI)	P_{fit}
Fixed effects, generalized least squares				
log OR	−0.119 (−0.240, 0.003)	0.361 (0.180, 0.542)	0.479 (0.229, 0.730)	0.171
log RR	−0.064 (−0.131, 0.004)	0.285 (0.139, 0.431)	0.349 (0.169, 0.528)	0.154
RD (per cent)	−1.888 (−4.747, 0.970)	2.177 (0.666, 3.689)	4.066 (0.494, 7.637)	0.005
ASD	−0.055 (−0.114, 0.005)	0.115 (0.056, 0.175)	0.170 (0.073, 0.267)	0.089
Random effects, iteratively reweighted least squares (Berkey method)				
log OR	−0.124 (−0.283, 0.035)	0.361 (0.004, 0.718)	0.485 (0.152, 0.819)	—
log RR	−0.072 (−0.162, 0.019)	0.301 (0.002, 0.599)	0.372 (0.093, 0.651)	—
RD (per cent)	−2.479 (−6.421, 1.463)	2.962 (−0.879, 6.803)	5.441 (−0.556, 11.438)	—
ASD	−0.058 (−0.137, 0.022)	0.102 (−0.007, 0.212)	0.160 (0.033, 0.287)	—
Random effects, maximum likelihood				
log OR	−0.123 (−0.261, 0.015)	0.366 (0.170, 0.561)	0.488 (0.231, 0.746)	—
log RR	−0.070 (−0.151, 0.010)	0.294 (0.133, 0.454)	0.364 (0.174, 0.554)	—
RD (per cent)	−2.463 (−5.827, 0.900)	2.691 (0.091, 5.290)	5.154 (0.663, 9.646)	—
ASD	−0.057 (−0.123, 0.010)	0.111 (0.042, 0.180)	0.168 (0.066, 0.270)	—

Note: The spurious accuracy of three decimal points is retained for the convenience of anyone wishing to repeat the examples. The DerSimonian and Laird model was not used in the multivariate case.

ASD, difference in arcsin-transformed risks; CI, confidence interval; OR, odds ratio; P_{fit}, p-value from goodness-of-fit test for the fixed-effects model (see text); RD, risk difference; and RR, risk ratio.

*Log relative odds ratios, log relative risk ratios, difference of risk differences, or differences of differences in the arcsin-transformed risks between the two outcomes.

DOI: 10.1002/sim

ACKNOWLEDGEMENTS

This work was supported by the Evidence-based Center of Excellence, Pfizer Inc. under a contract for methodology development. Author IO was also supported by the National Science Foundation Grant 063413.

REFERENCES

1. Gleser LJ, Olkin I. Stochastically dependent effect sizes. In *The Handbook of Research Synthesis*, Cooper H, Hedges LV (eds). Russell Sage Foundation: New York, 1994; 339–355.
2. Berkey CS, Anderson JJ, Hoaglin DC. Multiple-outcome meta-analysis of clinical trials. *Statistics in Medicine* 1996; **15**(5):537–557.
3. Berkey CS, Hoaglin DC, Antczak-Bouckoms A, Mosteller F, Colditz GA. Meta-analysis of multiple outcomes by regression with random effects. *Statistics in Medicine* 1998; **17**(22):2537–2550.
4. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine* 2002; **21**(4):589–624.
5. Lu G, Ades AE, Sutton AJ, Cooper NJ, Briggs AH, Caldwell DM. Meta-analysis of mixed treatment comparisons at multiple follow-up times. *Statistics in Medicine* 2007; **26**(20):3681–3699.
6. Early Breast Cancer Trialists' Collaborative Group. Favourable and unfavourable effects on long-term survival of radiotherapy for early breast cancer: an overview of the randomised trials. *The Lancet* 2000; **355**(9217):1757–1770.
7. Early Breast Cancer Trialists' Collaborative Group. Radiotherapy for early breast cancer. *Cochrane Database of Systematic Reviews* 2002; **2**:CD003647.
8. Normand SL. Meta-analysis: formulating, evaluating, combining, and reporting. *Statistics in Medicine* 1999; **18**(3):321–359.
9. Hedges LV, Olkin I. *Statistical Methods for Meta-analysis.* Academic Press: San Diego, 1985.
10. Higgins J, Thompson S, Deeks J, Altman D. Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. *Journal of Health Services Research and Policy* 2002; **7**(1):51–61.
11. Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. *Annals of Internal Medicine* 1997; **127**(3):820–826.
12. Petitti DB. Approaches to heterogeneity in meta-analysis. *Statistics in Medicine* 2001; **20**(23):3625–3633.
13. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 2002; **21**(11): 1539–1558.
14. Lau J, Ioannidis JP, Schmid CH. Summing up evidence: one answer is not always enough. *The Lancet* 1998; **351**(9096):123–127.
15. Engels EA, Schmid CH, Terrin N, Olkin I, Lau J. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. *Statistics in Medicine* 2000; **19**(13):1707–1728.
16. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**(3):177–188.
17. Eaton JW. Octave (Computer Program). 2006. Available from: http://www.octave.org (accessed 11 February 2007).
18. R Development Core Team. *R*: *A Language and Environment for Statistical Computing* (*Computer Program*). R Foundation for Statistical Computing: Vienna, Austria, 2006.
19. Arends LR. Multivariate meta-analysis: modelling the heterogeneity. *Thesis/Dissertation*, 2006.
20. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Thompson JR. Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Medical Research Methodology* 2007; **7**:3.
21. Arends LR, Voko Z, Stijnen T. Combining multiple outcome measures in a meta-analysis: an application. *Statistics in Medicine* 2003; **22**(8):1335–1353.
22. Dear KB. Iterative generalized least squares for meta-analysis of survival data at multiple times. *Biometrics* 1994; **50**(4):989–1002.
23. Ezekowitz JA, Armstrong PW, McAlister FA. Implantable cardioverter defibrillators in primary and secondary prevention: a systematic review of randomized, controlled trials. *Annals of Internal Medicine* 2003; **138**(6): 445–452.
24. Lee DS, Green LD, Liu PP, Dorian P, Newman DM, Grant FC, Tu JV, Alter DA. Effectiveness of implantable defibrillators for preventing arrhythmic events and death: a meta-analysis. *Journal of the American College of Cardiology* 2003; **41**(3):1573–1582.

DOI: 10.1002/sim

25. Mielke O, Wardlaw J, Liu M. Thrombolysis (different doses, routes of administration and agents) for acute ischaemic stroke. *Cochrane Database of Systematic Reviews* 2004; **40**(4):CD000514.
26. Roulson J, Benbow EW, Hasleton PS. Discrepancies between clinical and autopsy diagnosis and the value of post mortem histology: a meta-analysis and review. *Histopathology* 2005; **47**(6):551–559.
27. Chan AW, Krleza-Jeric K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *Canadian Medical Association Journal* 2004; **171**(7):735–740.
28. Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *The Journal of the American Medical Association* 2004; **291**(20):2457–2465.
29. Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *British Medical Journal* 2005; **330**(7494):753.
30. Davey Smith G, Pekkanen J. Should there be a moratorium on the use of cholesterol lowering drugs? *British Medical Journal* 1992; **304**(6824):431–434.
31. Muldoon MF, Manuck SB, Matthews KA. Lowering cholesterol concentrations and mortality: a quantitative review of primary prevention trials. *British Medical Journal* 1990; **301**(6747):309–314.
32. Muldoon MF, Rossouw JE, Manuck SB, Glueck CJ, Kaplan JR, Kaufmann PG. Low or lowered cholesterol and risk of death from suicide and trauma. *Metabolism* 1993; **42**(9 Suppl 1):45–56.
33. Muldoon MF, Manuck SB, Mendelsohn AB, Kaplan JR, Belle SH. Cholesterol reduction and non-illness mortality: meta-analysis of randomised clinical trials. *British Medical Journal* 2001; **322**(7277):11–15.
34. Davey Smith G, Song F, Sheldon TA. Cholesterol lowering and mortality: the importance of considering initial level of risk. *British Medical Journal* 1993; **306**(6889):1367–1373.

DOI: 10.1002/sim