

Research article

Open Access

## Bivariate random-effects meta-analysis and the estimation of between-study correlation

Richard D Riley\*<sup>1</sup>, Keith R Abrams<sup>2</sup>, Alexander J Sutton<sup>2</sup>, Paul C Lambert<sup>2</sup> and John R Thompson<sup>2</sup>

Address: <sup>1</sup>Centre for Medical Statistics and Health Evaluation, School of Health Sciences, University of Liverpool, Shelley's Cottage, Brownlow Street, Liverpool, L69 3GS, UK and <sup>2</sup>Centre for Biostatistics and Genetic Epidemiology, Department of Health Sciences, University of Leicester, 2nd Floor, Adrian Building, University Road, Leicester, LE1 7RH, UK

Email: Richard D Riley\* - richard.riley@liv.ac.uk; Keith R Abrams - kra1@le.ac.uk; Alexander J Sutton - ajs22@le.ac.uk; Paul C Lambert - pl4@le.ac.uk; John R Thompson - trj@le.ac.uk

\* Corresponding author

Published: 12 January 2007

Received: 4 September 2006

BMC Medical Research Methodology 2007, 7:3 doi:10.1186/1471-2288-7-3

Accepted: 12 January 2007

This article is available from: <http://www.biomedcentral.com/1471-2288/7/3>

© 2007 Riley et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** When multiple endpoints are of interest in evidence synthesis, a multivariate meta-analysis can jointly synthesise those endpoints and utilise their correlation. A multivariate random-effects meta-analysis must incorporate and estimate the between-study correlation ( $\rho_B$ ).

**Methods:** In this paper we assess maximum likelihood estimation of a general normal model and a generalised model for bivariate random-effects meta-analysis (BRMA). We consider two applied examples, one involving a diagnostic marker and the other a surrogate outcome. These motivate a simulation study where estimation properties from BRMA are compared with those from two separate univariate random-effects meta-analyses (URMAs), the traditional approach.

**Results:** The normal BRMA model estimates  $\rho_B$  as -1 in both applied examples. Analytically we show this is due to the maximum likelihood estimator sensibly truncating the between-study covariance matrix on the boundary of its parameter space. Our simulations reveal this commonly occurs when the number of studies is small or the within-study variation is relatively large; it also causes upwardly biased between-study variance estimates, which are inflated to compensate for the restriction on  $\hat{\rho}_B$ . Importantly, this does not induce any systematic bias in the pooled estimates and produces conservative standard errors and mean-square errors. Furthermore, the normal BRMA is preferable to two normal URMAs; the mean-square error and standard error of pooled estimates is generally smaller in the BRMA, especially given data missing at random. For meta-analysis of proportions we then show that a generalised BRMA model is better still. This correctly uses a binomial rather than normal distribution, and produces better estimates than the normal BRMA and also two generalised URMAs; however the model may sometimes not converge due to difficulties estimating  $\rho_B$ .

**Conclusion:** A BRMA model offers numerous advantages over separate univariate syntheses; this paper highlights some of these benefits in both a normal and generalised modelling framework, and examines the estimation of between-study correlation to aid practitioners.

## Background

Traditionally, meta-analysis models combine summary measures of a single quantitative endpoint, taken from different studies, to produce a single pooled result. However, multiple pooled results are required whenever there are multiple outcomes [1] or multiple treatment groups [2]. A multivariate meta-analysis model uses the correlation between the endpoints and obtains the multiple pooled results collectively [3,4]. For example, Reitsma et al. [5] have suggested a bivariate random-effects meta-analysis (BRMA) to jointly synthesise logit-sensitivity and logit-specificity values from diagnostic studies. Multivariate meta-analysis has been quite widely used, including application to genetic associations [6], surrogate endpoints [7,8], psychological findings [9] and prognostic markers [10].

Often the advantage of a multivariate random-effects meta-analysis lies in its ability to use the between-study correlation of the multiple endpoints of interest. For example, in diagnostic studies the sensitivity and specificity are usually negatively correlated across studies due to the use of different thresholds [5]. Van Houwelingen et al. [4] use the between-study correlation to describe the shape of the bivariate relationship between the true log-odds in a treatment group and the true log-odds in a control group (baseline risk). Riley et al. [10] algebraically assess BRMA and show that the correlation allows a 'borrowing of strength' across endpoints. This leads to pooled estimates that have a smaller standard error than those from corresponding univariate random-effects meta-analyses (URMAs), especially when some endpoints are missing at random across studies.

Some recent articles have indicated the between-study correlation may often be estimated at the end of its parameter space, as +1 or -1. Thompson et al. [6] apply a normal BRMA model to genetic studies of coronary heart disease and report that the between-study correlation was 'poorly estimated' with the likelihood peaking at +1, and an estimate of +1 has also been reported in other applications [4,11]. To aid practitioners, in this paper we analytically consider why this occurs, and then explore the impact, if any, this has on the pooled estimates and between-study variance estimates. This investigation also allows us to examine the general benefits of BRMA over URMA, to build on a number of other recent articles [5,10], and encourage a greater use and appreciation of BRMA in practice.

The outline of the paper is as follows. We begin by introducing the general normal model for BRMA and discuss analytically why the between-study correlation can be estimated at the edge of its parameter space. We then apply the model to two real examples from the literature,

one involving a diagnostic marker and one involving a surrogate outcome, and these both give a between-study correlation estimate of -1. This then motivates a realistic simulation study to examine how the estimate of between-study correlation affects the statistical properties of the pooled and between-study variance estimates. It also allows us to compare the performance of the normal BRMA model to two separate URMAs, the more common approach in practice [10]. We then extend our work to consider meta-analysis of proportions, and highlight why a generalised model for BRMA is preferred to the general normal BRMA model [12], and also two separate generalised URMA models. We conclude by summarising the broad benefits of BRMA for practitioners, and discuss future research priorities.

## Methods

In this section we introduce a hierarchical normal framework for BRMA and URMA. We describe how the BRMA normal model is estimated, and analytically consider the estimation of between-study correlation. We then describe the rationale and methodology for our simulation study of BRMA versus URMA. To ensure a real world context, this section also includes two motivating examples from the medical literature where a BRMA is potentially important.

### Motivating example 1 – the telomerase data

Glas et al. [13] systematically review the sensitivity and specificity of tumour markers used for diagnosing primary bladder cancer. One of these markers was telomerase, a ribonucleoprotein enzyme, evaluated in 10 studies as shown in Table 1. Rather than applying a URMA independently for sensitivity and specificity, the authors jointly synthesise the logit-transformed sensitivity and the logit-transformed specificity in a normal BRMA model as described below and recently proposed elsewhere [5].

### A general normal model for bivariate random-effects meta-analysis (BRMA)

Suppose that  $i = 1$  to  $n$  studies are identified by a systematic review, and that two endpoints ( $j = 1$  or  $2$ ) are available from each study. Each study supplies summary measures,  $Y_{ij}$ , and associated standard errors,  $s_{ij}$ , for each endpoint. For instance, for diagnostic studies Reitsma et al. [5] suggest the logit-sensitivity is  $Y_{i1}$  and the logit-specificity is  $Y_{i2}$ . Each summary statistic ( $Y_{ij}$ ) is assumed to be an estimate of a true value ( $\theta_{ij}$ ) in each study, and in a hierarchical structure each  $\theta_{ij}$  is assumed to be drawn from a distribution with mean (or 'pooled') value  $\beta_j$  and between-study variance  $\tau_j^2$ . If  $Y_{ij}/\theta_{ij}$  and  $\theta_{ij}$  are both

**Table 1: The telomerase data taken from the bladder cancer review of Glas et al. [13]**

<b>Study</b>	<b>Number of patients with bladder cancer</b>	<b>Number of patients with bladder cancer and a positive test result</b>	<b>Sensitivity</b>	<b>Logit-sensitivity <math>Y_{i1}</math></b>	<b>Standard error of logit-sensitivity <math>S_{i1}</math></b>	<b>Number of patients without bladder cancer</b>	<b>Number of patients without bladder cancer and a negative test result</b>	<b>Specificity</b>	<b>Logit-specificity <math>Y_{i2}</math></b>	<b>Standard error of logit-specificity <math>S_{i2}</math></b>	<b>Within-study correlation <math>\rho_{wi}</math></b>
1	33	25	0.758	1.139	0.406	26	25	0.962	3.219	1.020	0
2	21	17	0.810	1.447	0.556	14	11	0.786	1.299	0.651	0
3	104	88	0.846	1.705	0.272	47	31	0.660	0.661	0.308	0
4	26	16	0.615	0.470	0.403	83	80	0.964	3.283	0.588	0
5	57	40	0.702	0.856	0.290	138	137	0.993	4.920	1.004	0
6	47	38	0.809	1.440	0.371	30	24	0.800	1.386	0.456	0
7*	43	23.5	0.547	0.187	0.306	13	12.5	0.962	3.219	1.442	0
8	33	27	0.818	1.504	0.451	20	18	0.900	2.197	0.745	0
9	17	14	0.824	1.540	0.636	32	29	0.906	2.269	0.606	0
10	44	37	0.841	1.665	0.412	29	7	0.241	-1.145	0.434	0

\* 0.5 was added to each cell of the 2 by 2 table for this study, as a continuity correction was needed given there were zero false negatives

assumed normally distributed then the BRMA can be specified as:

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix}, \delta_i\right) \quad \delta_i = \begin{pmatrix} s_{i1}^2 & s_{i1}s_{i2}\rho_{wi} \\ s_{i1}s_{i2}\rho_{wi} & s_{i2}^2 \end{pmatrix} \quad (1)$$

$$\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \Omega\right) \quad \Omega = \begin{pmatrix} \tau_1^2 & \tau_1\tau_2\rho_B \\ \tau_1\tau_2\rho_B & \tau_2^2 \end{pmatrix}$$

This model is the general normal model for BRMA [4], where  $\delta_i$  and  $\Omega$  are the within-study and the between-study covariance matrices respectively. Usually of key interest from the analysis are the pooled estimates of  $\beta_1$  and  $\beta_2$ , although sometimes an estimated function of these may be desired; for instance in the telomerase example an estimate of the log of the diagnostic odds ratio is given by  $\hat{\beta}_1 + \hat{\beta}_2$ . The BRMA differs from two independent URMA's by the inclusion of the within-study correlations (i.e. the  $\rho_{wi}$  s) and the between-study correlation ( $\rho_B$ ). Equation (2) is equivalent to two independent URMA's when  $\rho_{wi} = \rho_B = 0$  for all  $i$ , i.e. there is zero correlation:

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix}, \delta_i\right) \quad \delta_i = \begin{pmatrix} s_{i1}^2 & 0 \\ 0 & s_{i2}^2 \end{pmatrix} \quad (2)$$

$$\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \Omega\right) \quad \Omega = \begin{pmatrix} \tau_1^2 & 0 \\ 0 & \tau_2^2 \end{pmatrix}$$

In equation (1) it is common to assume the  $s_{ij}^2$  s and  $\rho_{wi}$  s are known [1,4]. Including the uncertainty of the  $s_{ij}^2$  s is unnecessary for URMA [14], but whether the uncertainty of the  $s_{ij}^2$  s and  $\rho_{wi}$  s should be incorporated in BRMA is yet to be examined. This issue is outside the scope of this paper but we note that a Bayesian framework is particularly flexible for incorporating such uncertainty [15].

**Within- and between-study correlation**

The *within-study correlation*,  $\rho_{wi}$  indicates whether  $Y_{i1}$  and  $Y_{i2}$  are correlated within a study, and these  $\rho_{wi}$  s are usually assumed known. For the telomerase data the  $\rho_{wi}$  s might be assumed to be zero because sensitivity and specificity values are calculated independently in a study using different sets of patients. In other BRMA applications the  $\rho_{wi}$  s can be non-zero, for example where the two endpoints are overall and disease-free survival [10]. In practice it may be difficult to obtain the value of non-zero

$\rho_{wi}$  s, although it can be done as evident in Berkey et al. [1] and the 'motivating example 2' below [7]. Suggestions for limiting the problem of unavailable  $\rho_{wi}$  s have been proposed [15-17], and this issue is considered further in the discussion.

The *between-study correlation*,  $\rho_B$ , is not generally known and has to be estimated when fitting the BRMA. It indicates how the underlying *true* values, i.e. the  $\theta_{i1}$  s and the  $\theta_{i2}$  s, are related across studies. It may be caused by differences across studies in patient-level characteristics, such as age and stage of disease, or changes in study-level characteristics, such as the threshold level in diagnostic studies

**Motivating example 2 – the CD4 data**

Daniels and Hughes [7] assess whether the change in CD4 cell count is a surrogate for time to either development of AIDS or death in drug trials of patients with HIV. They consider between-treatment-arm log-hazard ratios of time to onset of AIDS or death ( $Y_{i1}$ ), and between-treatment-arm differences in mean changes in CD4 count ( $Y_{i2}$ ) from pre-treatment baseline to about six months. Fifteen relevant trials were identified. Some of the trials involved three or four treatment arms, but to enable application to BRMA here we only consider outcome differences between the control arm and the first treatment arm in the reported dataset [7]. All fifteen trials provided complete data, including the within-study correlations which were quite small, varying between -0.22 and 0.17 with a mean of -0.08.

**Estimation**

In our analyses of equation (1) in this paper the between-study parameters (i.e.  $\tau_1^2$ ,  $\tau_2^2$  and  $\rho_B$ ) and the two pooled values ( $\beta_1$  and  $\beta_2$ ) are estimated iteratively using restricted maximum likelihood (REML) in SAS Proc Mixed, as described elsewhere [4]. Unless otherwise stated, we also use Cholesky decomposition [18] of  $\Omega$  to ensure that this matrix is estimated to be positive semi-definite and therefore that the between-study correlation estimate,  $\hat{\rho}_B$ , is in the range [-1,1]. Cholesky decomposition of  $\Omega$  also helps ensure convergence when  $\hat{\rho}_B$  is very close to 1 or -1.

**Analytic consideration of the between-study covariance parameters**

Estimation and inference in classical linear mixed models are based on the marginal model, which for equation (1) is the bivariate normal model with variance-covariance  $\delta_i + \Omega$ . Assume for the sake of simplicity that the within-study covariance matrix  $\delta_i = \delta$  for all studies. Then the covariance matrix of the observed  $Y_{i1}$  s and  $Y_{i2}$  s is given by  $V = \delta + \Omega$ , where  $\delta$  is known. Now, this puts (severe) restric-

tions on  $V$ , namely that  $V - \delta$  is a covariance matrix, that is non-negative definite. If the estimated  $V$  does not satisfy this restriction, the maximum likelihood estimate of  $\Omega$  will be truncated on the boundary of its parameter space. This means that if  $\Omega$  is diagonal (as for URMA) the maximum likelihood estimate on the boundary will have one or both of the between-study variances equal to zero; else if  $\Omega$  is non-diagonal (as for BRMA) then either one or both of the between-study variances equals zero or else the between-study correlation equals -1 or +1. In meta-analysis a between-study variance estimate of zero is well-understood, but a between-study correlation estimate of -1 or +1 is likely to be less familiar to practitioners. Thompson et al. [6] refer to this issue as 'poor estimation' but it perhaps should rather be considered a natural consequence of the sensible restrictions imposed on  $V$ , and one that prevents a variance estimate  $< 0$  or a correlation estimate  $> +1$  or  $< -1$ , as might otherwise be obtained.

**Rationale for the research in this paper**

In this paper, to aid practitioners we will further assess the normal BRMA model and the role of between-study correlation. In particular, we aim to identify the situations when  $\hat{\rho}_B$  is likely to be +1 or -1 and examine the impact this has on the pooled and between-study variance estimates. We will also evaluate the benefits of BRMA over two separate URMA, the more common approach in

practice, and explore extensions to a generalised BRMA model for meta-analysis of proportions. To achieve these goals we firstly apply the normal BRMA of equation (1) to the two motivating examples. We then perform a simulation study of the normal BRMA model, as described below. The generalised BRMA model is then introduced and assessed in relation to the normal BRMA model and two separate generalised URMA (see Results).

**A simulation study to assess BRMA and the between-study correlation**

We carried out a simulation study of the general normal model for BRMA in 11 scenarios, labelled (i) to (xi) (Table 2). Each of scenarios (i) to (xi) relates to a different but realistic specification of equation (1). Scenarios (i) to (vi) consider complete data, as in the telomerase example, whereas scenarios (vii) to (vi) consider when some data are missing at random across studies, as assumed in the BRMA of Thompson et al. [6]. The scenarios also vary in the relative sizes of the within- and between-study correlations, and also the within- and between-study variation. For example, scenarios (i) to (iv) involve within-study variances similar in size to the between-study variances, as observed in prognostic studies [10], and in scenario (vi) there is one relatively low and one relatively high between-study variance, as for the CD4 dataset. The sizes of the meta-analysis were either  $n = 5$  or  $n = 50$  studies for complete data, and either  $n = 10$  or  $n = 50$  for missing

**Table 2: Scenarios used in the simulations based on equation (1)**

Scenario	Pooled values		Between-study variances		Within- and between-study correlation		Within-study variation		Description
	$\beta_1$	$\beta_2$	$\tau_1^2$	$\tau_2^2$	$\rho_{wi}$	$\rho_B$	Median value of the $S_{ij}^2$ s:		
<i>Complete data</i>							$n = 50$	$n = 5$	
(i)	0	2	0.25	0.25	0	0	0.254	0.147	Zero correlation; within-study variation similar to between-study variation
(ii)	0	2	0.25	0.25	0	0.8	0.254	0.147	No within-study correlation but high between-study correlation; within-study variation similar to between-study variation
(iii)	0	2	0.25	0.25	0.8	0	0.254	0.147	High within-study correlation but no between-study correlation; within-study variation similar to between-study variation
(iv)	0	2	0.25	0.25	0.8	0.8	0.254	0.147	High within- and between-study correlation; within-study variation similar to between-study variation
(v)	0	2	0.0025	0.0025	0.8	0.8	0.254	0.147	High within- and between-study correlation; within-study variation large relative to between-study variation
(vi)	0	2	0.0025	1.5	0.8	0.8	0.254	0.147	High within- and between-study correlation; within-study variation large (for endpoint 1) and small (for endpoint 2) relative to between-study variation
(vii)	0	2	1.5	1.5	0.8	0.8	0.254	0.147	High within- and between-study correlation; within-study variation small relative to between-study variation
<i>Missing data</i>							$n = 50$	$n = 10$	
(viii)	0	2	1.5	1.5	0	0.8	0.244	0.183	No within-study correlation but high between-study correlation; within-study variance small relative to between-study variance
(ix)	0	2	0.25	0.25	0	0.8	0.244	0.183	No within-study correlation but high between-study correlation; within-study variance similar to between-study variance
(x)	0	2	1.5	1.5	0.8	0.8	0.244	0.183	High within- and between-study correlation; within-study variance small relative to between-study variance
(xi)	0	2	0.25	0.25	0.8	0.8	0.244	0.183	High within- and between-study correlation; within-study variance similar to between-study variance

data. Our method of simulation was deliberately chosen to be similar to that previously used by Berkey et al. [1] and Sohn [19]. As an example, we now describe the simulation procedure for scenario (i) with  $n = 50$ .

#### **Description of the simulation procedure for scenario (i) with $n = 50$**

##### *Generation of a dataset of 1000 meta-analyses*

We chose  $\beta_1 = 0$  in order to reflect little clinical benefit (e.g. a sensitivity of 50%) and in contrast  $\beta_2 = 2$  (e.g. a specificity of 88%). The within study variances required, i.e. the 50  $s_{i1}^2$  s and 50  $s_{i2}^2$  s, were each found by sampling from a  $N(0.25, 0.50)$  distribution and squaring the value obtained. This produced a median  $s_{ij}^2$  of 0.25 and an interquartile range of 0.7, values similar to those for the telomerase data (median  $s_{ij}^2 = 0.21$ ) and for the BRMA of Thompson et al. [6] (median  $s_{ij}^2 = 0.28$ , interquartile range = 0.76). The between-study variances,  $\tau_1^2$  and  $\tau_2^2$ , were chosen to be 0.25 in scenario (i), which meant that they were similar in size to the median within-study variances. The within and between-study correlations were both set to zero for this scenario. All these choices were substituted into equation (1) and 1000 meta-analyses each of 50 studies were generated. Calculations were performed in S-Plus using the 'rmvnorm' function for generating bivariate normal values (code available upon request).

##### *Estimation using the dataset of 1000 meta-analyses*

Each of the 1000 meta-analyses in scenario (i) were analysed separately by:

- fitting two separate URMAAs as in equation (2) (where  $\rho_B = 0$ ) using REML to estimate  $\beta_1$ ,  $\beta_2$ ,  $\tau_1^2$  and  $\tau_2^2$
- fitting a BRMA as in equation (1) using REML to estimate  $\beta_1$ ,  $\beta_2$ ,  $\tau_1^2$ ,  $\tau_2^2$ , and  $\rho_B$

The 1000 BRMA estimates and the corresponding 1000 URMA estimates from scenario (i) were then compared by calculating:

- average parameter estimates across all the simulations (to check for bias)
- coverage of the 95% confidence intervals for  $\beta_1$  and  $\beta_2$

- average standard error and mean-square error (MSE) of  $\beta_1$  and  $\beta_2$

- the number of occasions  $\hat{\rho}_B$  was equal to +1 or -1 in the BRMA

To assess coverage, the 95% confidence intervals for  $\beta_j$  were calculated using:

$$\hat{\beta}_j \pm \left( t_{n_j-1}(0.05) * \sqrt{\text{var}(\hat{\beta}_j)} \right),$$

with  $n_j$  the number of studies providing endpoint  $j$ . This  $t$ -distribution is commonly used in the meta-analysis literature, although it is only an approximation [20].

#### **Description of the simulation procedure for scenarios (ii) to (xi)**

Simulations in the other scenarios followed in the same manner as described above but with the data generated from the parameter values specific to each scenario as given in Table 2. For those missing data simulations of scenarios (viii) to (xi) we simulated data as described for complete data, except that for each generated meta-analysis we removed the data for the second endpoint in a randomly selected 50% of studies. So, for example, with  $n = 50$  in scenario (viii) each of the 1000 simulated meta-analyses contained 25 studies with complete data and 25 studies with data for the first endpoint only.

## **Results**

### **Application to the telomerase and CD4 data**

The normal model for BRMA (equation (1)) and then two separate URMAAs (equation (2)) were applied to the telomerase data. Both approaches gave a pooled sensitivity of about 76% and a pooled specificity of about 88% (Table 3). The BRMA gave a between-study correlation estimate of -1 but the profile likelihood reveals that there is little information regarding  $\rho_B$  with the log-likelihood gradually increasing as  $\hat{\rho}_B$  approaches -1 (Figure 1), the end of its parameter space. Interestingly, the between-study variances were estimated to be somewhat larger in the BRMA than the URMA, and the standard errors of the pooled estimates were also slightly larger in the BRMA; just the opposite of what one might expect from a bivariate analysis utilising large correlation [10]. A similar finding was observed upon application of BRMA and URMA to the CD4 data. The BRMA again gave a between-study correlation estimate of -1 and both between-study variances were estimated somewhat larger in the BRMA than the URMA, as were the standard errors of the pooled estimates.

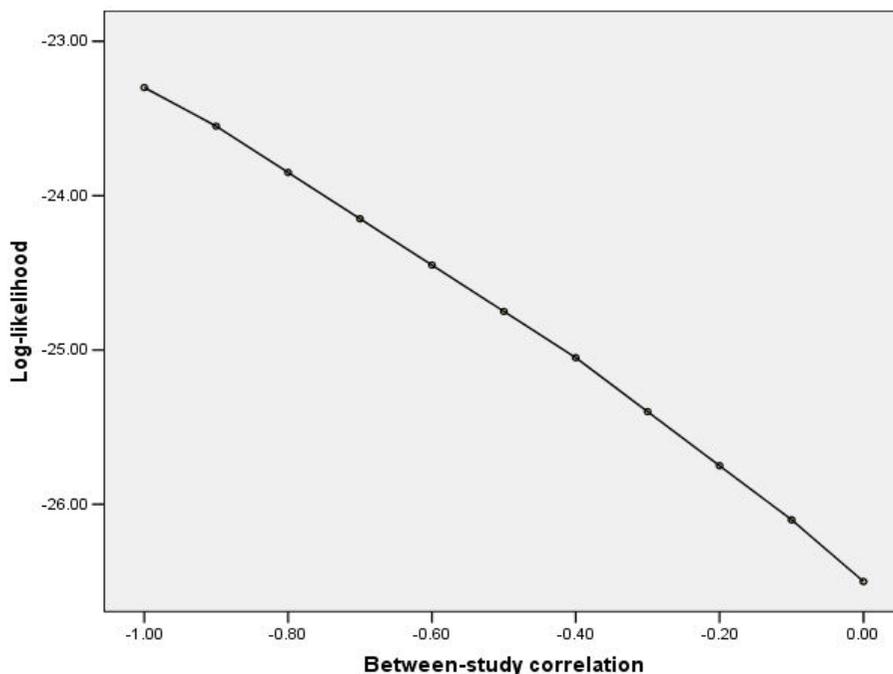
**Table 3: URMA and BRMA results for the telomerase and CD4 datasets**

Dataset	Model	Pooled value endpoint 1 $\hat{\beta}_1$ (s.e.)	Between-study variance endpoint $2 \hat{\tau}_1^2$	Pooled value endpoint 2 $\hat{\beta}_2$ (s.e.)	Between-study variance endpoint $2 \hat{\tau}_2^2$	Between-study correlation $\hat{\rho}_B$
Telomerase	Normal URMA	1.155 (0.186)	0.186	1.964 (0.541)	2.386	NA
	Normal BRMA	1.166 (0.186)	0.202	2.058 (0.554)	2.584	-1.0
	Generalised URMA	1.182 (0.176)	0.155	2.215 (0.578)	2.680	NA
CD4	Normal URMA	-0.049 (0.0695)	0.025	17.300 (5.561)	379.73	NA
	Normal BRMA	-0.109 (0.0748)	0.048	18.314 (5.740)	412.96	-1.0

s.e. = standard error; NA = not applicable; 95% confidence interval calculated using t-distribution with 9 degrees of freedom. Restricted maximum likelihood estimation was used for the normal models, whereas maximum likelihood estimation was used for the generalised model.

The question is thus posed: is the estimation of  $\rho_B$  at the boundary of its parameter space adversely influencing the other BRMA parameter estimates and, if so, how (e.g. does it introduce bias)? Also, in terms of the individual pooled estimates, the telomerase and CD4 examples indicate little benefit of BRMA over two separate URMAs, despite the utilisation of correlation; but is this generally true and in what situations should a BRMA be preferred? To understand the answers to these important questions, it is helpful to now consider the results from our simulation study.

We note at this point that, for both telomerase and CD4, we also tried estimation of BRMA using an unstructured form of  $\Omega$ , rather than using Cholesky decomposition of  $\Omega$  as previously. Interestingly, this approach produced non-sensical between-study correlation estimates of -1.12 and -1.074 for the telomerase and CD4 datasets respectively. This emphasises the importance of a boundary constraint on  $\rho_B$  as imposed by the Cholesky decomposition.



**Figure 1**  
Profile log-likelihood for the between-study correlation from the general normal BRMA of the telomerase data.

### Simulation results

Table 4 includes the simulation results for scenarios (i), (ii), (viii) and (ix). The results for all other scenarios are provided in Appendix 1 (see Additional file 1). We now discuss the key findings.

#### Between-study correlation

One can see from Table 4 that the normal model for BRMA often estimates the between-study correlation,  $\rho_B$ , as either +1 or -1, especially when the number of studies in the meta-analysis is small. For example, with  $n = 5$  in scenario (ii), where the true  $\rho_B$  was 0.8, 605 of the 1000 simulations (60.5%) gave  $\hat{\rho}_B$  equal to +1 and 103 simulations (10.3%) gave  $\hat{\rho}_B$  equal to -1. This led to a mean value of  $\hat{\rho}_B$  equal to 0.639, a downward bias of about 20%. However, this downward bias is not in itself a concern because it is simply caused by the maximum likelihood estimator sensibly truncating  $\hat{\rho}_B$  at +1 and -1, which improves the mean-square error of  $\hat{\rho}_B$ . Furthermore,  $\hat{\rho}_B$  is clearly asymptotically unbiased, with the occurrence of  $\hat{\rho}_B$  equal to -1 or +1 and thus the bias in mean  $\hat{\rho}_B$  becoming increasingly less as the number of studies in the meta-analysis increases (Table 4). Interestingly though, the number of studies required to reduce the occurrence of  $\hat{\rho}_B$  equal to -1 or +1 was far greater when the within-study variation was large relative to the between-study variation. For example, even with a large  $n = 50$  studies in scenario (v), where the within-study variation was relatively large, 58% of the simulations gave a between-study correlation of -1 or +1.

These findings indicate why  $\hat{\rho}_B$  equals -1 in the BRMAs of the telomerase and CD4 datasets. For the telomerase data there are 10 studies; the simulations show that this magnitude of studies will often provide little information about  $\rho_B$ , causing  $\hat{\rho}_B$  to often be constrained at -1 or +1 so that the restrictions imposed on  $\mathbf{V}$  are met. For the CD4 data, even though there are five more studies than telomerase, the mean within-study variance for endpoint  $j = 1$  is 0.15 and this is large relative to the between-study variation ( $\hat{\tau}_1^2 = 0.048$ ). In such situations where the within-study variation dominates, the simulations again show that  $\hat{\rho}_B$  will often require truncation at the end of its parameter space to ensure  $\mathbf{V}$  is non-negative definite.

#### Between-study variance estimates

Our simulation results show that the between-study variance estimates were less frequently truncated at zero in the BRMA than the URMA (Table 4). For example, in scenario (ii) with  $n = 5$   $\hat{\tau}_2^2$  was zero for 104 of the URMA simulations and none of the BRMA simulations. Furthermore, in those scenarios where the between-study correlation was often +1 or -1 (e.g. scenario (ii) with  $n = 5$ ), the normal BRMA model produces a noticeable upward bias in the between-study variance estimates. For example, in scenario (ii) with  $n = 5$  there was an upward bias of 0.024 in  $\hat{\tau}_1^2$  and  $\hat{\tau}_2^2$ , about 10% above their true value. To understand analytically why this occurs, we need to consider that the between-study covariance ( $\tau_{12}$ ) is formulated by  $\tau_{12} = \rho_B \tau_1 \tau_2$ . Now, if  $\hat{\rho}_B$  is constrained at -1 or +1, then to obtain the necessary solution for  $\hat{\tau}_{12}$  the maximum likelihood estimator can only increase the  $\hat{\tau}_j$ s, which do not have an upper bound constraint. Thus the between-study variance estimates are inflated to compensate for the constraint on  $\hat{\rho}_B$ . This explains why the BRMAs of the telomerase and CD4 data, where  $\hat{\rho}_B$  was truncated at -1, give  $\hat{\tau}_j$ s that are noticeably larger than those from two separate URMAs. Practitioners need to be aware of this issue; however, we do not consider it a major concern as the maximum likelihood estimator for  $\tau_j^2$  is still asymptotically unbiased (the bias decreases as the number of studies increases) and the inflation is simply caused by the sensible and necessary constraint on  $\hat{\rho}_B$ . Furthermore, the inflation is essentially conservative, leading to a larger standard error and mean-square error of pooled estimates as now discussed.

#### Pooled estimates

For all complete and missing data scenarios, the pooled estimates were approximately unbiased for both BRMA and URMA. Even in those scenarios where  $\hat{\rho}_B$  was often +1 or -1 it is encouraging that, despite the upward bias in between-study variances, there was no systematic bias in the pooled estimates from the BRMA (Table 4 and Appendix 1 – see Additional file 1). The main affect on the pooled estimates was an inflated standard error and mean-square error. This is a conservative property, but meant that in some complete data scenarios the BRMA performed slightly worse than URMA, despite the utilisation of correlation in the BRMA that might be expected to improve efficiency [10]. For example, in the  $n = 5$  results

**Table 4: Simulation results of the normal BRMA and URMA models for scenarios (i), (ii), (viii), and (ix)**

Meta-analysis model	<i>n</i>	No. of the 1000 simulations that converged	Bias of mean $\hat{\beta}_1$	Mean s.e. of $\hat{\beta}_1$	MSE of $\hat{\beta}_1$	Coverage of the 95% CIs for $\hat{\beta}_1$	Bias of mean $\hat{\beta}_2$	Mean s.e. of $\hat{\beta}_2$	MSE of $\hat{\beta}_2$	Coverage of the 95% CIs for $\hat{\beta}_2$	Bias of mean $\hat{\tau}_1^2$ (no. of $\hat{\tau}_1^2 = 0$ )	Bias of mean $\hat{\tau}_2^2$ (no. of $\hat{\tau}_2^2 = 0$ )	Bias of mean $\hat{\rho}_B$	% of $\hat{\rho}_B = -1$	% of $\hat{\rho}_B = 1$
<b>Scenario (i): Complete data – zero correlation; within-study variance similar to between-study variance</b>															
URMA	50	1000	-0.005	0.102	0.010	94.8%	0.001	0.107	0.0108	95.4%	-0.003 (0)	0.005 (1)	-	-	-
BRMA	50	1000	-0.005	0.101	0.010	94.7%	0.001	0.106	0.0108	95.5%	-0.003 (0)	0.005 (0)	-0.001	0.2%	0.4%
URMA	5	1000	-0.002	0.267	0.081	96.0%	-0.006	0.267	0.0887	94.0%	-0.006 (89)	0.015 (81)	-	-	-
BRMA	5	998	-0.002	0.274	0.081	96.7%	-0.006	0.269	0.0894	95.3%	0.008 (10)	0.024 (0)	-0.027	29.6%	29.0%
<b>Scenario (ii): Complete data – no within-study correlation, high between-study correlation; within-study variance similar to between-study variance</b>															
URMA	50	1000	-0.004	0.102	0.010	95.6%	0.001	0.106	0.0114	94.4%	0 (0)	-0.004 (1)	-	-	-
BRMA	50	1000	-0.004	0.100	0.010	95.3%	0	0.104	0.0107	95.4%	0.001 (0)	-0.001 (0)	-0.005	0%	25.2%
URMA	5	999	-0.002	0.271	0.077	97.3%	-0.005	0.263	0.0826	94.2%	0.004 (80)	0.005 (104)	-	-	-
BRMA	5	1000	-0.002	0.279	0.077	98.1%	-0.008	0.268	0.0819	95.7%	0.024 (15)	0.024 (0)	-0.161	10.3%	60.5%
<b>Scenario (viii): Missing data – no within-study correlation, high between-study correlation; within-study variance smaller than between-study variance</b>															
URMA	50	1000	-0.004	0.071	0.005	94.9%	0	0.099	0.0101	95.0%	-0.006 (0)	-0.005 (0)	-	-	-
BRMA	50	1000	-0.004	0.071	0.005	94.9%	0	0.082	0.0068	95.2%	-0.006 (0)	-0.007 (0)	-0.001	0%	0%
URMA	10	1000	-0.002	0.154	0.028	94.1%	-0.003	0.209	0.0576	93.7%	-0.006 (0)	-0.006 (0)	-	-	-
BRMA	10	1000	-0.002	0.154	0.028	94.1%	-0.001	0.174	0.0427	93.3%	-0.006 (0)	0.006 (0)	-0.040	0%	3.9%
<b>Scenario (ix): Missing data – no within-study correlation, high between-study correlation; within-study variance similar to between-study variance</b>															
URMA	50	1000	-0.004	0.102	0.010	95.6%	-0.001	0.145	0.0228	94.2%	0 (0)	-0.003 (6)	-	-	-
BRMA	50	1000	-0.004	0.101	0.010	95.8%	-0.003	0.137	0.0203	94.7%	0.001 (0)	0.003 (0)	-0.012	0.1%	35.6%
URMA	10	1000	-0.001	0.218	0.045	93.9%	-0.005	0.263	0.0825	94.2%	0.006 (45)	0.005 (84)	-	-	-
BRMA	10	997	-0.001	0.222	0.045	96.5%	-0.007	0.255	0.0797	95.6%	0.020 (0)	0.025 (0)	-0.164	10.1%	60.2%

MSE = mean-square-error, *n* = number of studies in each meta-analysis, CIs = confidence intervals, s.e. = standard error;

for scenario (iii), where about 56% of simulations gave  $\hat{\rho}_B$  equal to +1 or -1 and there was an upward bias in between-study variances, the standard error/mean-square error of  $\hat{\beta}_1$  was larger in the BRMA (0.272/0.083) than the URMA (0.268/0.081). This explains the BRMA results for telomerase and CD4, where the standard errors of pooled estimates were larger in the BRMA than the URMA due to the inflated between-study variances.

The coverage of  $\beta_1$  and  $\beta_2$  was between 93% and 98% in most scenarios assessed, and was often similar in URMA and BRMA. It is hard, though, to make general conclusions regarding coverage, as those situations where  $\hat{\rho}_B$  is often +1 or -1 are the same situations where the  $t$ -distribution with  $n-1$  degrees of freedom is a poor approximation to the true sampling distribution. The true degrees of freedom to use here are complex and account for the within-study variances [12]; however this is rarely done in meta-analysis and is beyond the scope of this paper.

#### *BRMA versus URMA for estimating the pooled values*

For complete data, in most scenarios the BRMA was marginally superior to URMA as the pooled estimates had slightly smaller standard errors and mean-square errors, especially given large correlations (Table 4 and Appendix 1 – see Additional file 1). However, the URMA sometimes performed equally well, and occasionally even better in those scenarios where  $\hat{\rho}_B$  was often +1 or -1 as discussed above. We also compared the subset of BRMA results where  $\hat{\rho}_B$  did not equal +1 or -1 with the corresponding URMA results, and again found that BRMA was generally slightly superior to URMA. This finding agrees with previous algebraic results [10], that given complete data there is generally a very small benefit of BRMA over URMA for estimating  $\beta_1$  and  $\beta_2$  themselves. Our focus here is on the individual pooled estimates, but we note that there are also broader reasons why a BRMA may be preferred over URMA for complete data. These are summarised in the Discussion to ensure a more complete picture for practitioners considering BRMA.

For the missing data simulations, the pooled estimate for endpoint  $j = 2$  was of particular interest because of the missing data for this endpoint. Encouragingly, the mean-square error and mean standard error of  $\hat{\beta}_2$  were much smaller in the BRMA than the URMA, although the coverage was comparable (Table 4 and Appendix 1 – see Addi-

tional file 1). For example, in the  $n = 10$  simulations of scenario (xi) the mean standard error was 0.225 in the BRMA compared to 0.262 in the URMA, and the MSE was 0.0708 in the BRMA compared to 0.0921 in the URMA. The reduction in standard error and MSE was larger when both the within- and between-study correlations were high. Even when  $\rho_{Wi}$  was zero there was still a reasonable benefit if  $\rho_B$  was high; for example, in the  $n = 10$  simulations of scenario (viii) the mean standard error of  $\hat{\beta}_2$  was 0.174 in the BRMA compared to 0.209 in the URMA. This finding agrees with algebraic work regarding the benefits of BRMA for when there are data missing at random [10]. Practitioners should again consider this benefit alongside the other broader reasons for using BRMA rather than URMA (see Discussion).

#### **Extended simulations of the normal BRMA model**

In our above simulations of the normal BRMA model we used non-negative within- and between-study correlations; however, in reality negative correlations may arise as in the telomerase and CD4 examples. Also, our simulations took the within-study correlations to be the same in each study, while in reality their value may vary. Further simulations were thus performed to assess negative correlation and discrepant within-study correlations. These gave findings consistent with those identified previously (Appendix 2 – see Additional file 2); the BRMA was still beneficial over URMA for estimating the pooled endpoints, and where the between-study correlation estimate was often +1 or -1 there was again an upward bias in the BRMA between-study variance estimates.

For simplicity, in all our simulations  $s_{i1}^2$  and  $s_{i2}^2$  were generated independently but in reality they are likely to be correlated due to the sample size being similar for both endpoints. We also generated the  $s_{ij}^2$  s independent to the  $Y_{ij}$  s, yet in many situations, such as the synthesis of log-odds ratios, the size of  $s_{ij}^2$  may be related to the size of  $Y_{ij}$ . To address this, we also performed further simulations where we firstly generated individual binary data for diagnostic studies, using simulation code kindly provided by Chu and Cole [12]. From this realistic raw data we then calculated the  $Y_{ij}$  s and their  $s_{ij}^2$  s, before then fitting the normal BRMA model as before. The results again show that the between-study correlation estimate is often +1 or -1 and the BRMA is still preferable to URMA, with improved mean-square error, coverage and, especially, bias of estimates (Table 5). However, the results also

revealed some severe limitations of a normal model for meta-analysis of binary data, as now discussed.

**A generalised model for BRMA of proportions**

So far in this paper we have modelled the summary statistics across studies, i.e. the  $Y_{ij}$ s, and assumed they are normally distributed. Indeed, in our main simulations we generated the  $Y_{ij}$ s directly from the normal BRMA model of equation (1); thus, our conclusions are only valid for  $Y_{ij}$ s that can truly be assumed normally distributed. This normality assumption is common in the meta-analysis field, and will often be suitable (see Discussion). However, in our two motivating examples it is more plausible for the CD4 data than the telomerase data as the latter involves a meta-analysis of proportions, for which the normality assumption is not appropriate when some studies have a small number of patients or the proportions are close to 0 or 1. For this reason recent articles [12] suggest that, rather than modelling the logit-proportions using the normal distribution, one should directly model the binary data using a binomial distribution. This approach also avoids the use of ad hoc continuity corrections in those studies which have zero cells. In terms of diagnostic studies, this generalised model for BRMA of sensitivity and specificity can be written as follows [12]:

no. testing positive,  $\sim$  Binomial (total no. true positives, sensitivity<sub>i</sub>)  $\text{logit}(\text{sensitivity}_i) = \beta_1 + u_1$

no. testing negative,  $\sim$  Binomial (total no. true negatives, specificity<sub>i</sub>)  $\text{logit}(\text{specificity}_i) = \beta_2 + u_2$

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \sim N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{\Omega} \right], \quad \mathbf{\Omega} = \begin{pmatrix} \tau_1^2 & \tau_1 \tau_2 \rho_B \\ \tau_1 \tau_2 \rho_B & \tau_2^2 \end{pmatrix} \quad (3)$$

Equation (3) can be fitted using maximum likelihood estimation in SAS NLMIXED. Chu and Cole [12] show that where the true sensitivity and specificity are large, this generalised BRMA model produces close to unbiased pooled and between-study correlation estimates, whereas the general normal BRMA model produces somewhat biased estimates. This can also be seen in our simulations results for meta-analysis of proportions in Table 5. The mean-square error, coverage, and, most noticeably, bias of estimates are far superior in the generalised BRMA than the normal BRMA. Furthermore, the generalised BRMA is also marginally superior in terms of bias to two separate generalised URMA (i.e. equation (3) where  $\rho_B = 0$ ), emphasising that the BRMA is also beneficial over URMA in the generalised model framework. Note though that, although it is the best method, the generalised BRMA model is itself not without bias (Table 5); to rectify this, extension to REML or other estimation techniques is potentially important.

To conclude our research we applied the generalised BRMA model to the telomerase data. Unfortunately the model would not converge appropriately; different starting values all produced a between-study correlation estimate of -1 but gave markedly different parameter estimates and caused spurious standard errors. For example, for one set of starting values the model gave the standard error of  $\hat{\beta}_1$  as 30.5, whereas for another set the standard error was close to zero. Indeed SAS provides the following warning: 'the final Hessian matrix is not positive definite, and therefore the estimated covariance matrix is not full rank and may be unreliable'. The problem here is again due to the between-study correlation of estimate -1 in  $\mathbf{\Omega}$ , as this causes the determinant of  $\hat{\mathbf{\Omega}}$  to be zero. This has greater implications in equation (3) than for the normal BRMA model of equation (1). The maximum likelihood estimator for equation (1) involves the determinant of  $\delta_i + \hat{\mathbf{\Omega}}$  in each study, which will not be zero unless the within-study correlations are also +1 or -1. However, in equation (3) the maximum likelihood estimator involves the determinant of  $\hat{\mathbf{\Omega}}$  itself, which causes problems akin to dividing by zero, which is why spurious estimates and standard errors are produced. It is clear that there is simply little information to estimate the between-study correlation for the telomerase dataset, due to the small number of studies. This issue is also evident in our  $n = 10$  simulations of the generalised BRMA model (Table 5), where 397 of the 1000 simulations did not converge appropriately. In such situations where estimating  $\rho_B$  is difficult, application of two generalised URMA may be the most appropriate option available (Table 3), although specifically for diagnostic studies other methods may also be valuable [21].

**Discussion**

Multivariate meta-analysis models are increasingly used to synthesise multiple, correlated endpoints of interest, especially in studies of diagnosis [5,21] and surrogate outcomes [7,8]. The Campbell Collaboration suggests that meta-analysts 'should not ignore the dependence among study outcomes'; however, they also note that 'the consequences of accounting for (modelling) dependence or ignoring it are not well understood' [22]. To therefore aid practitioners considering the approach, in this paper we have examined two models for BRMA and compared them to separate univariate syntheses, the traditional approach. We now discuss the main conclusions from our work and suggest future research priorities.

**Table 5: simulation results for meta-analysis of proportions**

Meta-analysis model	<i>n</i>	No. of the 1000 simulations that converged	Bias of mean $\hat{\beta}_1$	Mean s.e. of $\hat{\beta}_1$	MSE of $\hat{\beta}_1$	Coverage of the 95% CIs for $\hat{\beta}_1$	Bias of mean $\hat{\beta}_2$	Mean s.e. of $\hat{\beta}_2$	MSE of $\hat{\beta}_2$	Coverage of the 95% CIs for $\hat{\beta}_2$	Bias of mean $\hat{\tau}_1^2$ (no. of $\hat{\tau}_1^2 = 0$ )	Bias of mean $\hat{\tau}_2^2$ (no. of $\hat{\tau}_2^2 = 0$ )	Bias of mean $\hat{\rho}_B$	% of $\hat{\rho}_B = -1$	% of $\hat{\rho}_B = 1$
Normal URMA	10	995	-0.152	0.311	0.120	93.1%	-0.174	0.310	0.119	93.2%	-0.313 (17)	-0.309 (5)	-	-	-
Normal BRMA	10	995	-0.124	0.318	0.118	93.8%	-0.144	0.317	0.115	94.5%	-0.245 (7)	-0.245 (2)	0.067	53.7%	0%
Generalised URMA	10	1000	-0.013	0.330	0.124	93.8%	-0.037	0.328	0.116	94.0%	-0.186 (20)	-0.189 (21)	-	-	-
Generalised URMA	10	603*	-0.012	0.339	0.119	94.9%	-0.030	0.342	0.118	95.8%	-0.136 (0)	-0.114 (0)	-	-	-
Generalised BRMA	10	603	-0.009	0.339	0.119	95.4%	-0.029	0.341	0.116	96.2%	-0.134 (0)	-0.113 (0)	-0.084	0%	0%
Normal URMA	50	1000	-0.175	0.141	0.049	77.2%	-0.182	0.141	0.050	75.5%	-0.337 (0)	-0.335 (0)	-	-	-
Normal BRMA	50	1000	-0.151	0.142	0.042	81.4%	-0.157	0.143	0.043	81.3%	-0.285 (0)	-0.282 (0)	0.087	17.0%	0%
Generalised URMA	50	1000	-0.018	0.157	0.024	95.1%	-0.026	0.157	0.023	96.1%	-0.091 (0)	-0.084 (0)	-	-	-
Generalised URMA	50	973*	-0.019	0.157	0.024	95.1%	-0.022	0.157	0.023	96.1%	-0.087 (0)	-0.080 (0)	-	-	-
Generalised BRMA	50	973	-0.016	0.157	0.024	96.0%	-0.020	0.158	0.023	96.2%	-0.078 (0)	-0.071 (0)	0.019	0%	0%

MSE = mean-square-error, *n* = number of studies in each meta-analysis, CIs = confidence intervals, s.e. = standard error;

Restricted maximum likelihood estimation was used for the normal models, whereas maximum likelihood estimation was used for the generalised models.

\* These were a subset of the 1000 URMA simulations that converged and were the same ones that converged in the equivalent BRMA; this helps fairly compare the URMA and BRMA results.

### The general normal model for BRMA

A normal meta-analysis model is appropriate when the  $Y_{ij}$ s can be assumed normally distributed; this assumption is commonly used, for example where the  $Y_{ij}$ s are log-odds ratios [15], log-hazard ratios [10], mean differences [1] and log-event rates [11].

#### Between-study covariance parameters

It is clear from our work that maximum likelihood estimation of a normal random-effects meta-analysis model will often truncate the between-study covariance matrix,  $\Omega$ , on the boundary of its parameter space. For URMA this is observed by a between-study variance estimate of zero, whilst in BRMA it is more likely observed by a between-study correlation of +1 or -1. Practitioners are likely to be familiar with the concept of zero variance, but are perhaps less likely to appreciate why a correlation is estimated at unity. However, both arise for the same reason, namely that  $\Omega$  must be a non-negative definite matrix such that  $\tau_j^2$  is not  $< 0$  and  $\rho_B$  is not  $> +1$  or  $< -1$ . Our simulations show that, especially when the number of studies is small and/or the within-study variance is large relative to the between-study variance, such truncation is often necessary to ensure the sensible restrictions are met. In the normal BRMA, we have also shown that a consequence of  $\hat{\rho}_B$  being truncated at +1 or -1 is an upward bias in between-study variance estimates, which are inflated upwards to compensate for the restriction on  $\hat{\rho}_B$ . Practitioners should not, though, be overly concerned by this. We have shown it does not cause any systematic bias in the pooled estimates from BRMA, and it leads to conservative standard errors and mean-square errors.

#### The benefits over URMA for the pooled estimates

Our simulation results highlight that a normal model for BRMA is preferable to two separate URMA for estimating the pooled endpoints, and our results are consistent with previous findings that show how the inclusion of correlation allows the 'borrowing of strength' across endpoints [1,10,19]. We thus recommend practitioners use a BRMA rather than two separate URMA where possible. In particular, when some data are missing at random the BRMA is likely to produce a much smaller standard error and mean-square error of pooled estimates than URMA, even for moderate correlations. Riley et al. [10] give an applied example that shows this. For complete data, practitioners should not expect to see much gain in statistical efficiency over URMA; the mean-square error and standard error of pooled estimates are generally only marginally smaller in

BRMA than URMA, and on the occasion of  $\hat{\rho}_B = +1$  or  $-1$  they may even be slightly worse in BRMA (due to the inflated between-study variances, as in the telomerase and CD4 examples). However, there are broader reasons why BRMA may still be preferable in this situation (see below).

### The generalised model for BRMA

In equation (3) we extended our work to a generalised BRMA model for meta-analysis of two proportions. For synthesis of two proportions, like sensitivity and specificity, this approach is preferable to the general normal BRMA model (Table 5) because the normality assumption breaks down when the proportions are close to 0 or 1 and when there are small patient numbers [12]. It also avoids the use of ad-hoc continuity corrections when there are zero cells in some studies. Practitioners synthesising diagnostic studies are thus encouraged to use the generalised BRMA model, rather than the normal model or indeed two separate generalised URMA. However, they should also be aware that a between-study correlation estimate of +1 or -1 in the generalised BRMA model is likely to be associated with non-convergence and unstable pooled estimates, as discussed for the telomerase data. In such situations there may be little information to estimate the correlation, and so practitioners may wish to consider other methods for synthesising diagnostic studies, such as the hierarchical summary receiver operating characteristic (HSROC) method [21]; if this is also not possible then the best option may be a generalised URMA for sensitivity and specificity separately (Hamza et al., *personal communication*).

### The broader benefits of BRMA

Our simulations focused mainly on the benefits of BRMA over URMA for estimating the pooled endpoints. However, there are also broader reasons why a BRMA may be preferable to URMA, for either complete or missing data. For example, BRMA allows one to describe the bivariate relationship between endpoints [4,5], model, test or make predictions from their association [7], and estimate some function of the two pooled endpoints, like  $\beta_1 - \beta_2$  [10],  $\beta_1 + \beta_2$  [5], or  $\beta_1/\beta_2$  [6]. For instance, for the telomerase data, a BRMA enables a single framework to estimate the pooled sensitivity, pooled specificity, and the pooled diagnostic odds ratio ( $\exp(\beta_1 + \beta_2)$ ) [21]. Furthermore, Reitsma et al. [5] show that a BRMA of diagnostic studies enables the correlation between pooled endpoints to be estimated, which allows one to measure the shape of their bivariate relation and construct confidence ellipses. It also allows calculation of the conditional variance in one parameter given a fixed value of the other parameter, and allows drawing of the summary ROC curve. In terms of the CD4 data, the estimated correlation between pooled endpoints from a BRMA enables one to predict the time of

onset of AIDS or death from a future patient's CD4 level. Thus the BRMA can help establish whether CD4 should be used as a surrogate of disease-free survival [7,8], and further research of multivariate meta-analysis in this context is recommended. A BRMA may also be extended to a bivariate meta-regression by including additional study-level covariates that explain the between-study heterogeneity. For example, one may wish to include a covariate for study quality in meta-analysis of diagnostic studies [23]. Berkey et al [1] show that a bivariate meta-regression is more efficient than separate univariate meta-regressions for assessing such study-level covariates, again due to the inclusion of correlation.

#### Further research suggestions

For the general normal BRMA model, the role of estimation techniques other than REML would be interesting to consider, especially as other potentially better options have just been proposed [24]. For the generalised BRMA model, SAS NLMIXED currently only allows maximum likelihood estimation and so extension to REML is required, especially as the maximum likelihood estimates are not without bias (Table 5). Further research of multivariate meta-analysis within a Bayesian framework is also potentially important, as it would enable the incorporation of prior knowledge about the parameters, which may be valuable when the number of studies is small [15,25]. It would also allow the uncertainty of the  $s_{i1}^2$  s,  $s_{i2}^2$  s and  $\rho_{wi}$  s to be taken into account, as in practice they will only be estimates themselves as mentioned by Daniels and Hughes [7]. Further assessment of the role of the within-study correlations is also required, in particular what should we do when they are non-zero but unavailable? For the meta-analysis of surrogate endpoints it has been suggested that the within-study correlations are likely to be small (between 0 and 0.2) and can plausibly be considered constant across studies [7], or even zero [26]. However, this is not necessarily true in other fields; for example, in a multivariate meta-analysis of longitudinal data the within-study correlations varied between 0.48 and 0.97 (Jones et al., *personal communication*).

The use of individual patient data (IPD) in multivariate meta-analysis should also be considered, especially as IPD is the gold-standard for meta-analysis [27] and it would allow any unavailable within-study correlations to be calculated directly [7]. In practice though, IPD may only be available for a proportion of studies, and so methods for multivariate meta-analysis are required that combine IPD and aggregate data [27,28]. There has also be little consideration of how to assess dissemination bias using a multi-

variate meta-analysis framework [29,30], and this warrants attention as meta-analysis datasets are often fraught with such issues as publication bias [31] and within-study selective reporting [32,33]. In such scenarios some of the missing endpoints may not be missing at random, and so sensitivity analyses to assess how the meta-analysis results change under a variety of missing data assumptions would be potentially valuable [34].

#### Conclusion

In this paper we have used analytic reasoning, two applied examples and a realistic simulation study to highlight the benefits of a normal model for BRMA over two separate URMA, and explain why the between-study correlation is often estimated as +1 or -1. For meta-analysis of proportions, we also extended our work to a generalised model for BRMA, to ensure the binary data is modelled correctly. Our work adds to a growing body of literature indicating the rationale and benefits of a multivariate approach to meta-analysis, and we encourage meta-analysts to consider the approach in practice.

#### Abbreviations

BRMA – Bivariate random-effects meta-analysis

IPD – Individual patient data

URMA – Univariate random-effects meta-analysis

REML – Restricted maximum likelihood

#### Competing interests

The author(s) declare that they have no competing interests.

#### Authors' contributions

RDR, KRA, AJS and PCL developed the study. RDR performed analysis of the telomerase and CD4 data. RDR, PL and JRT performed the normal model simulations; RDR performed the generalized model simulations. All authors examined the simulation results and discussed their interpretations. RDR wrote the paper with contributions from all authors.

#### Additional material

##### Additional file 1

Appendix 1. Simulation results of the normal BRMA and URMA models for scenarios (iii) to (vii), (x) and (xi)

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2288-7-3-S1.doc>]

### Additional file 2

Appendix 2. Simulation results for the normal BRMA and URMA models for some scenarios involving negative correlation

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2288-7-3-S2.doc>]

### Acknowledgements

Richard Riley is funded by the Department of Health National Coordinating Centre for Research Capacity Development as a Post-doctoral Research Scientist in Evidence Synthesis. We would like to thank the three reviewers of our manuscript, whose comments have substantially improved the content of this paper. We also greatly thank Haitao Chu, who kindly provided SAS simulation code that enabled us to assess the generalised bivariate model.

### References

- 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)**:2537-2550.
- Hasselblad V: **Meta-analysis of multitreatment studies.** *Med Decis Making* 1998, **18(1)**:37-43.
- Becker BJ, Tinsley HEA, Brown S: . In *Multivariate Meta-analysis* San Diego , Academic Press; 2000.
- Van Houwelingen HC, Arends LR, Stijnen T: **Advanced methods in meta-analysis: multivariate approach and meta-regression.** *Stat Med* 2002, **21(4)**:589-624.
- Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH: **Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews.** *J Clin Epidemiol* 2005, **58(10)**:982-990.
- Thompson JR, Minelli C, Abrams KR, Tobin MD: **Meta-analysis of genetic studies using Mendelian randomization--a multivariate approach.** *Stat Med* 2005, **24(14)**:2241-2254.
- Daniels MJ, Hughes MD: **Meta-analysis for the evaluation of potential surrogate markers.** *Stat Med* 1997, **16(17)**:1965-1982.
- Gail MH, Pfeiffer R, Van Houwelingen HC, Carroll RJ: **On meta-analytic assessment of surrogate outcomes.** *Biostatistics* 2000, **1(3)**:231-246.
- Kalaian HA, Raudenbush SW: **A Multivariate Mixed Linear Model for Meta-Analysis.** *Psychological Methods* 1996, **1(3)**:227-235.
- 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.** *Statistics in Medicine* 2007, **26**:78-97.
- Arends LR, Vokro Z, Stijnen T: **Combining multiple outcome measures in a meta-analysis: an application.** *Stat Med* 2003, **22(8)**:1335-1353.
- Chu H, Cole SR: **Bivariate meta-analysis for sensitivity and specificity with sparse data: a generalized linear mixed model approach (letter to the Editor).** *Journal of Clinical Epidemiology* 2006, (in press):.
- Glas AS, Roos D, Deutekom M, Zwinderman AH, Bossuyt PM, Kurth KH: **Tumor markers in the diagnosis of primary bladder cancer. A systematic review.** *J Urol* 2003, **169(6)**:1975-1982.
- Hardy RJ, Thompson SG: **A likelihood approach to meta-analysis with random effects.** *Stat Med* 1996, **15(6)**:619-629.
- Nam IS, Mengersen K, Garthwaite P: **Multivariate meta-analysis.** *Stat Med* 2003, **22(14)**:2309-2333.
- Berrington A, Cox DR: **Generalized least squares for the synthesis of correlated information.** *Biostatistics* 2003, **4(3)**:423-431.
- Raudenbush SW, Becker BJ, Kalaian H: **Modeling multivariate effect sizes.** *Psychological Bulletin* 1988, **103(1)**:111-120.
- Gentle JE, Gentle JE: **Cholesky Factorization.** In *Numerical Linear Algebra for Applications in Statistics* Berlin , Springer-Verlag; 1998:93-95.
- Sohn SY: **Multivariate meta-analysis with potentially correlated marketing study results.** *Naval Research Logistics* 2000, **47**:500-510.
- Follmann DA, Proschan MA: **Valid inference in random effects meta-analysis.** *Biometrics* 1999, **55(3)**:732-737.
- Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA: **A unification of models for meta-analysis of diagnostic accuracy studies.** *Biostatistics* 2006.
- Becker BJ, Hedges LV, Pigott TD: **Campbell Collaboration Statistical Analysis Policy Brief.** *A Campbell Collaboration resource document (available at [http://www.campbellcollaboration.org/ECG/policy\\_statsp](http://www.campbellcollaboration.org/ECG/policy_statsp))* 2004.
- Westwood ME, Whiting PF, Kleijnen J: **How does study quality affect the results of a diagnostic meta-analysis?** *BMC Med Res Methodol* 2005, **5(1)**:20.
- Sidik K, Jonkman JN: **A comparison of heterogeneity variance estimators in combining results of studies.** *Stat Med (in press)* .
- Abrams KR, Sutton AJ, Cooper NJ, Sculpher MPS, Ginley L, Robinson M: **Populating economic decision models using meta-analyses of heterogeneously reported studies augmented with expert beliefs.** *Technical Report 05-06, Centre for Biostatistics and Genetic Epidemiology, University of Leicester* 2005.
- Korn EL, Albert PS, McShane LM: **Assessing surrogates as trial endpoints using mixed models.** *Stat Med* 2005, **24(2)**:163-182.
- Riley RD, Look MP, Simmonds MC: **Combining individual patient data and aggregate data in evidence synthesis: a systematic review identified current practice and possible methods.** *Journal of Clinical Epidemiology (in press)* 2007.
- Goldstein H, Yang M, Omar RZ, Turner RM, Thompson SG: **Meta-analysis using multilevel models with an application to the study of class size effects.** *Applied Statistics* 2000, **49**:399-412.
- Riley RD, Sutton AJ, Abrams KR, Lambert PC: **Sensitivity analyses allowed more appropriate and reliable meta-analysis conclusions for multiple outcomes when missing data was present.** *Journal of Clinical Epidemiology* 2004, **57(9)**:911-924.
- Jackson D, Copas J, Sutton AJ: **Modelling reporting bias: the operative mortality rate for ruptured abdominal aortic aneurysm repair .** *Journal of the Royal Statistical Society, Series A* 2005, **168**:737-752.
- Sterne JA, Egger M, Smith GD: **Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis.** *BMJ* 2001, **323(7304)**:101-105.
- Hahn S, Williamson PR, Hutton JL, Garner P, Flynn EV: **Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies.** *Stat Med* 2000, **19(24)**:3325-3336.
- Hutton JL, Williamson PR: **Bias in meta-analysis due to outcome variable selection within studies.** *Appl Stat* 2000, **49**:359-370.
- Copas J, Shi JQ: **Meta-analysis, funnel plots and sensitivity analysis.** *Biostatistics* 2000, **1(3)**:247-262.

### Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2288/7/3/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:

[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

