STATISTICS IN MEDICINE
Statist. Med. 2008; **27**:4381–4396
Published online 9 May 2008 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3311

Meta-analysis of summary survival curve data

Lidia R. Arends[1,*,†], M. G. Myriam Hunink[1,2,3] and Theo Stijnen[1]

[1] *Department of Epidemiology & Biostatistics, Erasmus MC, University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands*
[2] *Department of Radiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands*
[3] *Department of Health Policy & Management, Harvard School of Public Health, Boston, MA, U.S.A.*

SUMMARY

The use of standard univariate fixed- and random-effects models in meta-analysis has become well known in the last 20 years. However, these models are unsuitable for meta-analysis of clinical trials that present multiple survival estimates (usually illustrated by a survival curve) during a follow-up period. Therefore, special methods are needed to combine the survival curve data from different trials in a meta-analysis. For this purpose, only fixed-effects models have been suggested in the literature. In this paper, we propose a multivariate random-effects model for joint analysis of survival proportions reported at multiple time points and in different studies, to be combined in a meta-analysis. The model could be seen as a generalization of the fixed-effects model of Dear (*Biometrics* 1994; **50**:989–1002). We illustrate the method by using a simulated data example as well as using a clinical data example of meta-analysis with aggregated survival curve data. All analyses can be carried out with standard general linear MIXED model software. Copyright © 2008 John Wiley & Sons, Ltd.

KEY WORDS: meta-analysis; time to event; multivariate random effects model

1. INTRODUCTION

Since the introduction of the term ‘meta-analysis’ [1], i.e. the quantitative approach to summarize the outcomes of several studies, it has become an important technique in clinical research. The two main statistical approaches to meta-analysis are the fixed-effects and the random-effects models. Nowadays, an increasing number of meta-analyses are analysed with the univariate random-effects meta-analysis model as proposed by DerSimonian and Laird [2]. The fixed-effects model is still used, despite its many disadvantages, which include ignoring possible between-trial variation, and overestimation of the precision of the estimate [3, 4]. With the increasing use of meta-analysis,

[*] Correspondence to: Lidia R. Arends, Department of Epidemiology & Biostatistics, Erasmus MC, University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands.
[†] E-mail: l.arends@erasmusmc.nl, arends@fsw.eur.nl

Received 16 March 2007
Accepted 24 March 2008

the methodology of meta-analysis as a field is growing and changing. Previously, the main use of meta-analysis was to statistically pool results regarding one specific outcome variable from independent but 'combinable' studies [5]. This is still a common practice in most meta-analyses; however, in recent years a new trend is recognizable. As clinical practice is not decided on the basis of a single outcome variable at a single time point, investigators are interested in combining several outcome variables measured at multiple time points—which can vary between studies—to achieve an overall viewpoint on the (relative) performance of an experimental treatment. Within such a meta-analysis, the relationships between the outcome variables and between the various time points are of special interest.

An important subcategory of multiple time point measurements is survival, which is usually presented as a survival curve (proportion survival *versus* time). Within a given study, the reported survival proportions are correlated over time, meaning that the data have a multivariate nature. The meta-analysis of such data is more complicated than the meta-analysis of simpler, univariate data. This is the central issue of this paper.

Several methods have been proposed for meta-analysis of survival data. The gold standard approach is to obtain individual patient data (IPD) from each study [6, 7]. IPD meta-analysis has big advantages over meta-analysis based on the summary data presented in the articles [8, 9]. IPD meta-analysis of survival data fits well into the random-effects modelling methodology of survival data and is relatively well established [10]. However, often it is not possible or practical to collect IPD, and meta-analysis based on aggregate data extracted from study reports is the only option. In this paper, our focus is on methods for meta-analysis of aggregate survival data.

For a comparison between two treatment groups, the simplest approach would be to summarize the difference between the two treatment arms of each contributing trial by a single number, e.g. the (log) hazard ratio along with its standard error, and use standard methods of meta-analysis to combine them [11, 12]. Whitehead and Whitehead [13] discussed the meta-analysis of survival data, combining efficient score statistics for the hazard ratio of an assumed proportional hazards model. However, they expressed doubt over the chances of finding enough information about the required statistics such as the log-rank test statistics, or the hazard ratio estimates and their standard errors. Parmar *et al.* [14] presented a number of methods to extract estimates of these statistics in a variety of situations. Williamson *et al.* [6] give improved methods to estimate log hazard ratios from published survival curves and life tables. Even these methods do not always allow the calculation of the necessary quantities. Additionally, these approaches cannot process information from single-arm trials [11], making it less useful than a more general approach.

Survival reports are often just a series of survival proportions at discrete time intervals (once or twice a year), but the follow-up time points can be different in different studies. Therefore, the data tend to be very unbalanced and thus difficult to analyse with standard methods. One strategy to circumvent this problem is to reduce the survival curve to one or more fixed time points, e.g. the 3-year survival rate. This allows each time point to be analysed separately, with the standard univariate random- or fixed-effects model [15]. Analysing each time point separately, and thus carrying out multiple meta-analyses on the same data set, is inefficient and could lead to inappropriate conclusions [14]. It can lead to loss of power, because in each analysis only a portion of the data is used. It can also give rise to a multiple testing problem, and it could be difficult to interpret the results. Furthermore, this approach demands that the time points at which survival estimates are available are identical in all studies. As a stopgap, if a given trial does not report survival for the chosen time point, the data can be left missing or an estimate can be calculated using inter- or extrapolation. Alternatively, as in a recent analysis [16], survival data can be assigned to a time category (years 2–3, 4–5, and 6–8) and each time category can be analysed separately. When data are left missing, the comparisons at different time points may be based on completely different trials, negating the benefit of extended follow-up and biasing the results. Moodie *et al.* [17] present a methodology that employs the log(−log) survival function difference as the summary measure for the treatment effect. They show that this measure has a clear interpretation and can be used when survival estimates are reported at different time points across studies. This method is a fixed-effects method and, if survival estimates are reported for more than one time point, it does not use them all simultaneously.

All of the above-mentioned methods deal with one summary measure for the treatment effect. A number of methods that focus on estimating the whole survival curve have been proposed [11, 18–21]. In these meta-analysis methods, data from entire survival curves are combined, instead of artificially compromising the data to just one single survival statistic or to a survival estimate on just one fixed point in time. In an overview article, Earle and Wells [22] assessed five methods to combine published survival curves. The summary survival curve resulting from each method was compared with the curve calculated from the corresponding IPD.

One of the five models is the model of Dear [11]. The model of Dear is an extension of the method of Raudenbush *et al.* [23], who showed how to analyse effect sizes for two or more outcomes jointly in a fixed-effects generalized least-squares (GLS) regression model that allows adjustment for study-level covariates. Dear showed how to estimate correlations among serial survival proportions, allowing survival proportions reported at different times by different trials to be analysed together in a fixed-effects model. Berkey *et al.* [24] demonstrated that fixed-effect regression models for correlated outcomes may seriously underestimate the standard errors of regression coefficients when the regression model does not explain all of the heterogeneity between the trials. Therefore, Berkey *et al.* proposed a random-effects approach for the regression meta-analysis of multiple correlated outcomes. The tutorial of van Houwelingen *et al.* [4] shows how this model can be fitted easily using standard software.

In this paper, we propose a multivariate random-effects model for joint analysis of survival proportions reported at multiple times in different studies. The method makes use of the complete and possibly unbalanced set of reported survival proportions in all studies, and no inter- or extrapolation to common chosen endpoints is needed. The model can be seen as a generalization of the fixed-effects model of Dear [11] or as a combination of the models of Dear [11] and Berkey *et al.* [24] and it is fitted with standard software as described in van Houwelingen *et al.* [4]. Our method can also be considered as a kind of extension of the methods in [17] to random effects and multiple time points. The method is applied to a simulated data example as well as to a clinical data example of meta-analysis with aggregated survival curve data, which are described in Section 2. The clinical data example has also been used by Dear to illustrate his method. In Section 3 the model of Dear is presented, after which we propose our generalization of that model. In Section 4 the results of our model using the simulated data are discussed, and the results using the clinical data are compared with the results of Dear. In Section 5, we discuss the shortcomings and potential pitfalls of our model.

2. DATA SETS

To illustrate our method, we use two data sets. As raw, unbalanced, survival data sets are not freely available, our first data set is simulated. This data set is described in Section 2.1. The second data

DOI: 10.1002/sim

set that was used in the paper of Dear [11] is real clinical data. It is described in Section 2.2. Since Dear used the same clinical data set, we are able to directly compare the results of our method with those of Dear (Section 4).

2.1. Simulated data set

In order to test the model under controlled conditions, we created a meta-analysis data set consisting of 10 trials, in principal each containing two treatment arms: experimental and control (Table I). In order to mimic reality, not all trials have data for all time points. Studies 5 and 7 are created as observational studies, with only one treatment arm.

When confronted with a high degree of structural missing data, many meta-analysts will choose few fixed time points to subject to separate independent analyses. How they then deal with the missing data (ignore it, inter- or extrapolate it from the rest of the data, or categorize it) varies—there is no single method that enjoys prominence. In this paper, we present an approach that will allow meta-analysts to analyse all available survival data simultaneously, irrespective of the multiple time points at which the survival rates are measured.

To simulate the data we assumed Weibull-distributed survival times. The cumulative survival functions of the treatment groups in trial i were $\exp(-\exp(\beta_{0i}+\beta_{1i}Z)t^{\alpha_i})$, with α_i being the shape parameter and $\exp(\beta_{0i}+\beta_{1i}Z)$ the scale parameter, where Z is the treatment indicator. For the three parameters, we choose the following independent normal distributions:

$$\alpha_i \sim \mathrm{N}(1.0, 0.04)$$

$$\beta_{0i} \sim \mathrm{N}(-0.7, 0.04)$$

$$\beta_{1i} \sim \mathrm{N}(-0.5, 0.04)$$

This means that the relation with ln time is linear on the ln-minus-ln survival scale: $\ln(-\ln(S(t))) = \beta_0+\beta_1 Z+\alpha\ln(t)$. We randomly drew 100 survival times per treatment group.

To introduce censoring, we assumed a research project with a total study period of 4 years. The intake period of the project was 2 years, followed by a follow-up period of another 2 years. Thus, the potential follow-up period per patient varied between 2 and 4 years. For each patient we randomly drew a censoring time from a uniform distribution on the interval from 2 to 4 years. When this censoring time was lower than the survival time of that patient, the patient was censored. In our data this resulted in 30 per cent censored patients. To calculate the survival estimates for the different time points together with their standard errors, we used Kaplan–Meier survival analysis.

2.2. Clinical data example: bone marrow transplantation versus chemotherapy

A drawback of simulated data is that the data might not be realistic. To illustrate the proposed method with real-life data, we used the meta-analysis data provided by Begg *et al.* [25], which are based on the paper of Dear [11]. The authors studied the relative efficacy of bone marrow transplantation (BMT) *versus* conventional chemotherapy for young adults with acute non-lymphocytic leukaemia in their first complete remission. Data from 14 studies complied. In six studies, of which four randomized, the two treatment arms were directly compared. Eight observational studies included only patients who had had one of the two treatments (two studies BMT only; six studies chemotherapy only).

Table I. Data from the simulated data set: survival (standard errors in parentheses) of trial i by treatment arm j and year k.

Time	$k=0.5$	$k=1.0$	$k=1.5$	$k=2.0$	$k=2.5$	$k=2.7$	$k=3.0$
Trial i							
1 Control ($j=1$)	0.74 (0.04)	0.58 (0.05)	0.52 (0.05)	0.50 (0.05)	0.47 (0.05)	.	0.43 (0.05)
1 Exp. ($j=2$)	0.80 (0.04)	0.74 (0.04)	0.66 (0.05)	0.62 (0.05)	0.52 (0.05)	.	0.48 (0.05)
2 Control ($j=1$)	.	0.50 (0.05)	.	0.25 (0.04)	.	.	0.16 (0.04)
2 Exp. ($j=2$)	.	0.67 (0.05)	.	0.45 (0.05)	.	.	0.37 (0.05)
3 Control ($j=1$)	0.77 (0.04)	.	0.52 (0.05)	.	0.29 (0.05)	.	.
3 Exp. ($j=2$)	0.80 (0.04)	.	0.53 (0.05)	.	0.40 (0.05)	.	.
4 Control ($j=1$)	.	0.51 (0.05)	.	.	.	.	0.09 (0.03)
4 Exp. ($j=2$)	.	0.68 (0.05)	.	.	.	.	0.31 (0.05)
5 Control ($j=1$)	.	0.63 (0.05)	.	0.45 (0.05)	.	.	.
5 Exp ($j=2$)	.	.	.	.	.	.	.
6 Control ($j=1$)	.	0.57 (0.05)	.	.	.	.	0.23 (0.04)
6 Exp. ($j=2$)	.	0.78 (0.04)	.	0.52 (0.05)	.	.	0.42 (0.05)
7 Control ($j=1$)	.	.	.	.	.	.	
7 Exp. ($j=2$)	0.88 (0.04)	0.58 (0.05)	0.43 (0.05)	0.28 (0.05)	0.19 (0.04)	.	
8 Control ($j=1$)	.	0.69 (0.05)	.	.	.	.	0.23 (0.04)
8 Exp. ($j=2$)	.	0.78 (0.04)	.	.	.	.	0.47 (0.05)
9 Control ($j=1$)	.	0.68 (0.05)	.	.	.	.	.
9 Exp. ($j=2$)	.	.	.	.	.	.	0.42 (0.05)
10 Control ($j=1$)	.	.	.	.	.	0.19 (0.04)	.
10 Exp. ($j=2$)	.	.	.	.	.	0.47 (0.05)	.

The data consist of the Kaplan–Meier probabilities of disease-free survival at a maximum of five 1-year intervals after the start of treatment together with their standard errors (Table II). IPD are not available, and all studies give aggregated estimates for at least 3 years after the start of treatment.

The raw survival curves corresponding to the data in Table II are illustrated in Figure 1.

Table II. Data from Begg *et al.* [25] and Dear [11]: per cent disease-free survival (standard errors in parentheses) of trial i by treatment arm j and year k (1–5).

	Bone marrow transplantation ($j=1$)					Chemotherapy ($j=2$)				
Trial (i)	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
1	49 (12)	46 (12)	42 (12)	40 (12)	40 (12)	54 (8)	25 (8)	23 (7)	23 (7)	23 (7)
2	55 (10)	50 (10)	36 (9)			40 (8)	23 (7)	23 (7)	23 (7)	
3	54 (10)	47 (13)	40 (13)	40 (13)		54 (9)	42 (8)	28 (8)	28 (8)	
4	70 (23)	70 (23)	70 (23)	70 (23)		48 (17)	48 (17)	17 (13)		
5	54 (4)	46 (5)	42 (6)			40 (5)	21 (4)	16 (4)	16 (4)	
6	54 (2)	43 (3)	40 (3)	39 (3)		50 (4)	32 (4)	24 (4)	18 (4)	
7	59 (8)	49 (9)	47 (9)	47 (9)	47 (9)					
8	61 (8)	53 (8)	53 (8)	53 (8)	53 (8)					
9						60 (9)	48 (9)	32 (9)	32 (9)	
10						44 (5)	26 (4)	17 (5)	16 (4)	32 (9)
11						50 (3)	33 (3)	26 (3)	22 (3)	
12						62 (3)	38 (3)	29 (3)	24 (3)	19 (3)
13						50 (10)	24 (8)	16 (7)	12 (6)	22 (3)
14						76 (7)	53 (8)	53 (8)	50 (8)	50 (8)

3. METHODS

The studies to be combined in the meta-analysis are indexed by i; in our clinical data example $i=1,\dots,14$. In each study one or more treatments are considered. For each of the studies survival estimates are available for treatment j, where in our clinical data example $j=1$ for BMT and $j=2$ for chemotherapy. Some of the studies have survival estimates for both $j=1$ and 2, and some of the studies—the observational ones—have survival estimates only for either $j=1$ or $j=2$. For each treatment arm at times t_{ijk}, where the index k counts the time points (at most 5 in the clinical data example), a survival estimate is available together with its standard error. The true survival probability of the jth treatment in the ith trial on time point t_{ijk} is denoted by s_{ijk}. The estimate of s_{ijk} is denoted by $\hat{s}_{ijk}$. The corresponding standard error is denoted by se_{ijk}. The estimated correlations between the survival estimates would preferably also be available, but unfortunately this will seldom be the case. In general, estimates of correlations between multiple outcome measures often come from some external source [26]. However, in the special case of estimated survival probabilities, the correlations can be estimated from the data, as shown by Dear [11].

As shown in the simulated data example, the pattern of the time points might be different across the trials and within the trials across the treatment groups, dependent on the choice of the authors of the time points for which to provide survival estimates and standard errors in their publications. Hence, although in our clinical data example the time points are fixed to years after the start of treatment, this is not always the case.

3.1. The method of Dear

As a stepping stone to our proposed method, we briefly explain and discuss the GLS method proposed by Dear [11]. In this approach, a generalized linear regression model is used to relate

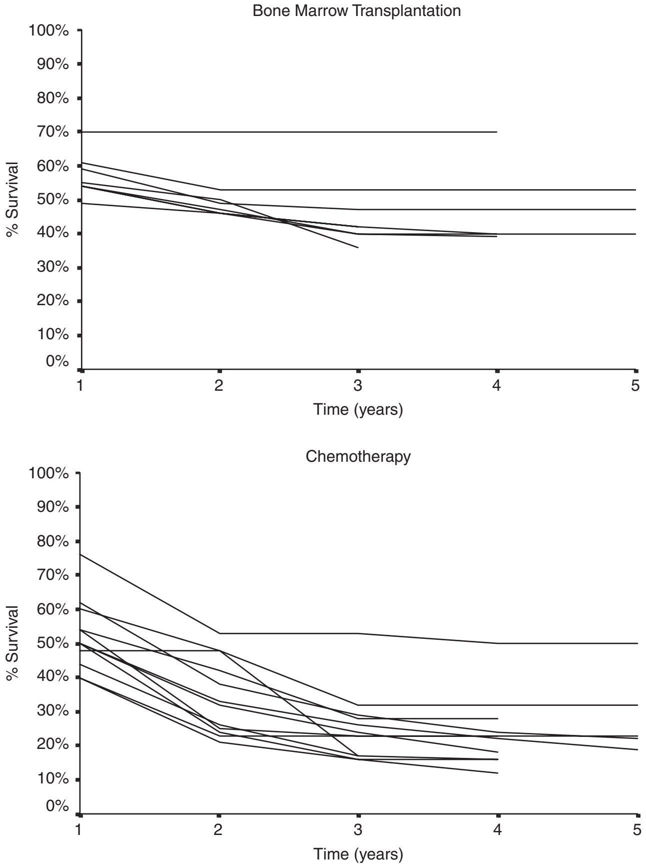

Figure 1. Survival curves based on data from Begg *et al.* [25].

the estimated survival proportions $\hat{s}_{ijk}$ to a design matrix X with both between- and within-study covariates such as time, study, and treatment characteristics, including interaction terms. Dear treated all covariates as categorical and represented them by dummies in the model, but that is not necessary. The model is

$$\hat{s}_i = X_i \beta + \varepsilon_i \tag{1}$$

where $\hat{s}_i$ is a column vector of $\hat{s}_{ijk}$'s and ε_i is a column vector of residuals with

$$\varepsilon_i \sim \mathrm{N}(0, V)$$

The ε_i's are assumed to be independent between studies and treatment arms. Since the errors ε_{ijk} of serial observations in the same treatment arm within the same study are expected to be related, all the off-diagonal elements of matrix V will not be zero. V is block diagonal with blocks corresponding to the treatment groups within studies. The main diagonal is set equal to the reported squared standard errors se^2_{ijk} of the survival proportions (Table II). To estimate the covariances within a treatment group, Dear [11] made use of the fact that the correlations of proportions between time points t_{ijk} and $t_{ijk'}$ are given by

$$\mathrm{corr}(s_{ijk}, s_{ijk'}) = \sqrt{\frac{s_{ijk}(1-s_{ijk'})}{(1-s_{ijk})s_{ijk'}}}$$

This formula is derived by exploiting the fact that the estimated cumulative hazards over different intervals are independent. Based on these correlations, the covariances in matrix V are obtained by multiplying the correlation with the corresponding standard errors given in Table II:

$$\mathrm{cov}(\hat{s}_{ijk}, \hat{s}_{ijk'}) = \mathrm{se}_{ijk} \cdot \sqrt{\frac{s_{ijk}(1-s_{ijk'})}{(1-s_{ijk})s_{ijk'}}} \cdot \mathrm{se}_{ijk'} \tag{2}$$

The se_{ijk}'s are considered fixed and known, while the s_{ijk}'s are to be estimated. The β parameters are estimated in an iterative manner. Iterations start with substituting the observed survival proportions in (2). Then, given this covariance matrix V, the β's are estimated with GLS: $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}s$. With these β's new estimates of the s_{ijk}'s are computed by (1) and substituted into (2). This results in a new matrix V, which is subsequently used to calculate new GLS β estimates. This is repeated until convergence, providing fully efficient maximum likelihood estimators for the β parameters.

Compared with the other methods published in the paper of Earle and Wells [22], the model of Dear [11] is generally applicable. It allows joint analysis of survival proportions at multiple time points, instead of analysing the survival probabilities separately for each time point. In addition, it is possible to combine studies with different numbers of treatment arms, as in our example; however, this assumes that the same treatment-specific profile of survival over time is applied in all studies. When this assumption is doubtful, then study characteristics, such as patient population parameters, should be sought to account for the discrepancies. These characteristics can be represented in the model through study-level covariates [11]. Finally, Dear's model makes it possible to fit and compare different regression models, making it very flexible and informative.

The most important shortcoming in Dear's model is that it is a fixed-effects model. To allow for between-study heterogeneity, Dear introduced a dummy variable for each study, indicating the study's common survival level. Because of this, there are very many parameters in the model relative to the number of data points.

The way in which this method is usually applied in practice also causes disadvantages. Dear presented his model as needing a fixed pattern of time points for which survival estimates are available, e.g. annually. Collections of published articles rarely provide this kind of balanced data; therefore, the meta-analyst needs to intra- and extrapolate estimates for the same set of time points across the studies. Even in the rare case where entire survival curves are published, allowing calculation of survival estimates at whatever time points are desired, the standard error at that time point will not be available. Another weakness in Dear's model is the use of dummies for all covariates including time and study. This can result in an enormous number of parameters to be estimated, especially when one also has to include interaction terms. The final weakness in Dear's model is that he fitted a linear model on the survival probabilities, while the probabilities are restricted to the [0,1] interval. Therefore, a linear model could return nonsense values. These weaknesses may be partially circumvent by modelling time as a continuous covariate, allowing

DOI: 10.1002/sim

different time patterns between studies and reducing the number of parameters. Furthermore, the method could be applied to transformed survival probabilities, e.g. with the logit or ln(− ln) transformation.

3.2. Multivariate random-effects model

In this paper we propose a multivariate, random, mixed-effects model that relates the ln-minus-ln transformed survival estimates $\ln(-\ln(\hat{s}_{ijk}))$ to both fixed and random covariates, such as time or ln(time), treatment group, etc. Any other transformation of the survival probabilities, which maps the interval $[0,1]$ into $(-\infty,\infty)$, for instance, the logit transformation, could also serve, but we prefer the ln(− ln) for reasons to be discussed in Section 5. We assume the following model:

$$\ln(-\ln(\hat{s}_i)) = X_i\beta + Z_i b_i + \varepsilon_i \tag{3}$$

with

$$b_i \sim \mathrm{N}(0, D)$$

$$\varepsilon_i \sim \mathrm{N}(0, V_i)$$

and

$$V_i = \begin{pmatrix} V_{i1} & 0 \\ 0 & V_{i2} \end{pmatrix} \quad \text{with } V_{ij} = \frac{\mathrm{se}_{ijk}}{\hat{s}_{ijk}\ln(\hat{s}_{ijk})}\sqrt{\frac{s_{ijk}(1-s_{ijk'})}{(1-s_{ijk})s_{ijk'}}}\frac{\mathrm{se}_{ijk'}}{\hat{s}_{ijk'}\ln(\hat{s}_{ijk'})} \tag{4}$$

In equation (3), $\hat{s}_i$, X_i, and ε_i have the same meaning as in Dear's model. β is the parameter vector containing the fixed effects (time, treatment, etc.). Compared with Dear's model, the model is extended with the random part $Z_i b_i$. The vectors of random coefficients b_i are assumed to be independently and normally distributed with expectation zero and between-studies covariance matrix D, independent from the ε_i's. Z_i is the design matrix for the random effects, typically containing intercept, time, and possibly treatment effect. The residual components have expectation zero and covariance matrix V_i, which is in fact the within-trial covariance matrix. Since the residual components across time are correlated within a trial arm (or survival curve) but independent between treatment arms, the covariance matrix V_i is a block diagonal matrix existing of blocks corresponding to the treatment arms. In our clinical example, we derive two matrices V_{i1} (for the BMT survival curves) and V_{i2} (for the chemotherapy survival curves). The within-study covariance matrix (4) is completely analogous to (2). Note that the standard error of the ln(− ln) transformed observed survival probability is

$$\mathrm{se}(\ln(-\ln(\hat{s}_{ijk}))) = \frac{\mathrm{se}_{ijk}}{\hat{s}_{ijk}\ln(\hat{s}_{ijk})}$$

This standard error is assumed to be known. The correlation between two transformed survival estimates is equal to

$$\mathrm{corr}(\ln(-\ln\hat{s}_{ijk}), \ln(-\ln\hat{s}_{ijk'})) = \sqrt{\frac{s_{ijk}(1-s_{ijk'})}{(1-s_{ijk})s_{ijk'}}}$$

Analogous to Dear's approach, this correlation is estimated from the data.

DOI: 10.1002/sim

The parameters in the model—the β's and the between-studies covariance matrix D—are estimated in an iterative manner similar to Dear's approach. In the first step, an initial estimate of V_i is obtained by substituting the observed survival probabilities into (4). With this covariance matrix, the mixed model is fitted, and the new survival estimates are used to calculate the new correlations in (4), and so on. This can be done easily in a general linear mixed model program, provided the residual variances can be fixed at arbitrary values per individual survival estimate [4]. Because of the iterative manner of model fitting, it is also convenient if the fitted survival estimates can be saved in order to automatically update the correlations between them and thereby the covariance matrix V_i. These features are available in the procedure MIXED of SAS and the function *lme* of S-Plus, but they may also be available in other statistical software. The SAS Proc Mixed syntax is available from the first author.

By applying a transformation to the observed survival probabilities, our model guarantees that fitted survival probabilities are between 0 and 1. However, as with Dear's method, the fitted survival curves are not necessarily non-increasing. We believe that in practice non-monotonically non-increasing fitted curves will be very rare and therefore will not be a serious problem. In addition, if curves are extrapolated from the smallest t_{ijk} to $t=0$, the fitted survival at $t=0$ might be less than 1.

As an illustration we apply this method to our two data examples and compare the results of the clinical data example with the results of Dear [11] using the same data set.

4. RESULTS

4.1. Results from the simulated data set

Using Proc MIXED in SAS, we fitted the following model on the data of Table I:

$$\ln(-\ln(\hat{s}_{ijk})) = \beta_0 + \beta_1 \text{treat}_{ijk} + \beta_2 \ln(\text{year}_{ijk}) + b_0 + b_1 \text{treat}_{ijk} + b_2 \ln(\text{year}_{ijk}) + \varepsilon_{ijk} \tag{5}$$

Here treat is a dummy variable for the treatment, $0=$control, and $1=$experimental. We allow random effects for the intercept, slope of ln(time), and for treatment effect. We assume a zero mean multivariate normal distribution for the random effects (b_1, b_2, b_3) with a covariance matrix, which is left completely free and has to be estimated. Choosing the $\ln(-\ln)$ transformation for the survival proportions and ln(year) instead of year itself as covariate corresponds to a Weibull distribution assumption. The advantage is that the treatment effect β_1 is expressed as a hazard ratio. Furthermore, survival curves start at level 1 at $t=0$. Of course, in practice the Weibull assumption might not be true, and other covariate specifications could be tried. In addition, another transformation of the survival probabilities than the $\ln(-\ln)$, such as the logit or probit, might be entertained. The parameters are estimated by repeated calls of Proc MIXED, each time updating the correlations. The results are given in Table III.

The overall mean estimated survival curves of both treatment groups together with their confidence intervals are shown in Figure 2.

All between-study variances differed significantly from zero when tested with the likelihood ratio test. As a model check, it was investigated whether adding terms as $\ln(\text{year})^2$ or interaction between ln(year) and treatment improved the model, but no extension was statistically significant.

DOI: 10.1002/sim

Table III. Results of fitting model (5) on the data of Table I.

Regression coefficients	Estimate	Standard error
Intercept	−0.5608	0.0871
ln(year)	0.9384	0.0689
Treat	−0.4911	0.0785
Covariance parameters		
Variance intercept	0.0505	
Covariance intercept∗ln(year)	0.0226	
Variance ln(year)	0.0301	
Covariance intercept∗treat	−0.0071	
Covariance ln(year)∗treat	−0.0283	
Variance treat	0.0171	

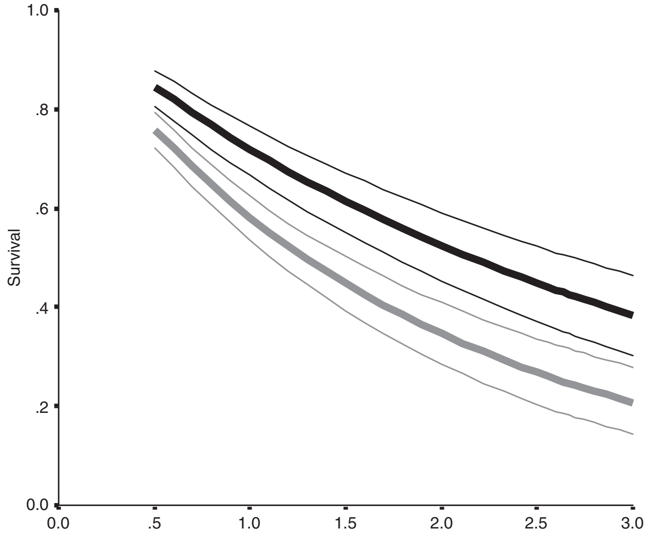

Figure 2. Overall mean survival curves (plus confidence bands) of the treated group (black lines) and the control group (grey lines) estimated from the data in Table I.

As a by-product of the analysis, empirical Bayes estimates are provided for the study-specific survival curves and are illustrated in Figure 3.

4.2. *Results from the clinical data example*

In this section, we present the results of the GLS model of Dear [11] as well as our multivariate random-effects model and indicate the differences and the similarities.

Dear fitted a linear model to relate the estimated survival proportions to several dummy variables, indicating the treatment, year of follow-up, and the study. The final model in the publication of Dear includes dummy variables for ‘study’, ‘treatment’, ‘follow-up year’, and for the interaction

DOI: 10.1002/sim

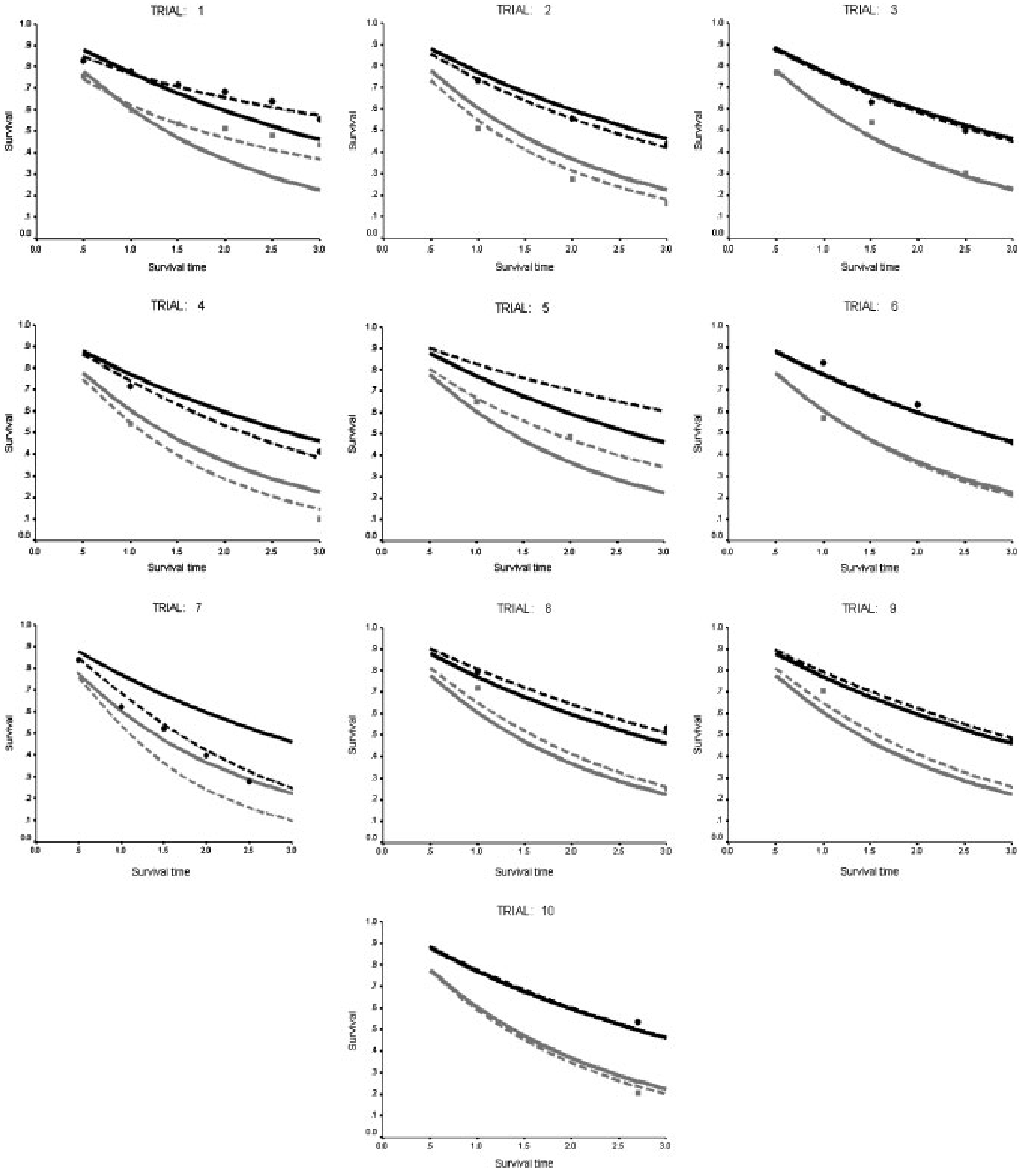

Figure 3. The observed survival estimates per treatment group and per survival time are given by the dots. The estimated mean survival curves (fixed effects) per treatment are plotted (solid lines) next to the estimated study-specific (empirical Bayes) survival curve lines (dotted lines). The black lines represent the treated group and the grey lines represent the control group.

of the dummies 'treatment by follow-up year', resulting in the following model:

$$\hat{s}_{ijk} = \beta_0 + \sum_{i=1}^{14} \beta_i \,\text{study}_i + \sum_{j=1}^{2} \beta_j \,\text{treatment}_j + \sum_{k=1}^{5} \beta_k \,\text{year}_k + \sum_{j=1,k=1}^{j=2,k=5} \beta_{jk} \,(\text{treatment}*\text{year})_{jk} + \varepsilon_{ijk} \tag{6}$$

with constraints $\sum \beta_i = 0, \sum \beta_j = 0, \sum \beta_k = 0$ to ensure estimability. The covariance matrix of $\hat{s}_{ijk}$ is estimated using GLS in an iterative manner, as explained in Section 3.

To compare this model with our multivariate random-effects model, we started to include the same variables as Dear did, but in a continuous manner. Therefore, estimated ln-minus-ln survival was the dependent variable, and treatment (0 = BMT, 1 = chemotherapy), ln(year), and the interaction between ln(year) and treatment were included as covariates, along with a random intercept and regression coefficients for ln(year) and treatment. Adding $\ln(\text{year})^2$ and the interaction between $\ln(\text{year})^2$ and treatment significantly improved the model. The random terms for ln(year) and treatment turned out to be non-significant and were dropped from the model. Thus, we ended up with the following model:

$$\ln(-\ln(\hat{s}_{ijk})) = \beta_0 + \beta_1 \text{treat} + \beta_2 \ln(\text{year}) + \beta_3 \text{treat}*\ln(\text{year}) + \beta_4 \ln(\text{year})^2 + \beta_5 \text{treat}*\ln(\text{year})^2 + b_{0i} + \varepsilon_{ijk} \tag{7}$$

The results are given in Table IV.

In Figure 4 the mean survival estimates per treatment are shown for models (6) and (7), together with their confidence intervals.

Note that the model of Dear included 23 β parameters in the model (to estimate 85 survival probabilities) *versus* only 7 (including only the variance of the intercept) in our model. Although at each time point there is extended overlap of the confidence intervals of both models, this overlap is slightly smaller at 2 and 3 years. At these time points, the model of Dear drops more than the multivariate model. This might be caused by the fact that the model of Dear works with dummies and is thereby slightly more flexible to a non-linear or non-quadratic decrease in the survival rate.

The empirical Bayes study-specific survival curves are depicted in Figure 5. The shrinkage phenomenon is nicely illustrated in this figure. For example, trial 4 has a much higher observed

Table IV. Parameter estimates of model (7).

Regression coefficients	Estimate	Standard error
Intercept	−0.61	0.08
Treatment	0.12	0.08
ln(year)	0.41	0.07
$\ln(\text{year})^2$	−0.10	0.04
Treatment∗ln(year)	0.54	0.10
Treatment∗$\ln(\text{year})^2$	−0.14	0.05
Covariance parameters		
Variance intercept	0.0338	
Covariance intercept∗ln(year)	−0.0001	
Variance ln(year)	0.0028	

DOI: 10.1002/sim

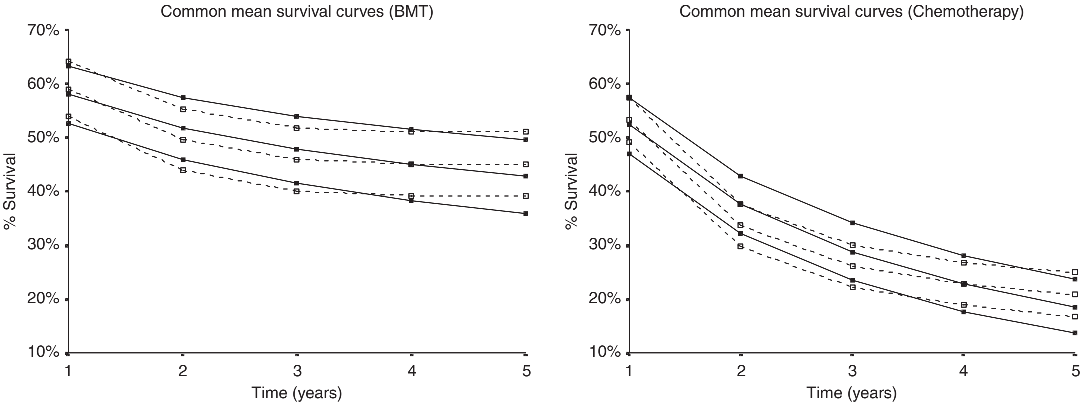

Figure 4. Mean survival estimates per treatment, with 95 per cent confidence limits. The dotted lines represent Dear's model and the solid lines represent the multivariate random-effects model.

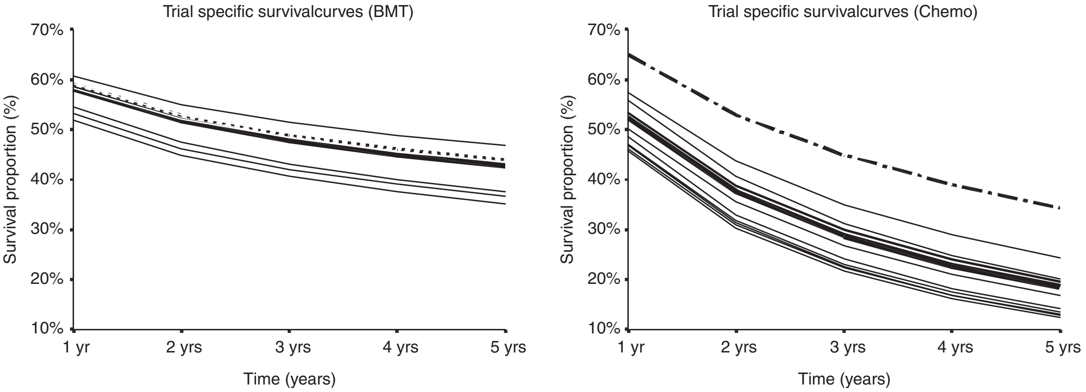

Figure 5. Trial-specific survival curves of both treatment arms. The dashed line in the BMT plot represents the survival curve of trial 4. The dashed line in the chemotherapy plot represents trial 14.

survival curve for the BMT treatment compared with the other trials (see Figure 1 and Table II). However, it is a very small trial with large standard errors (Table II). Therefore, the empirical Bayes estimate of the survival curve for this trial has strongly shrunk towards the average survival curve. On the other hand, trial 14 has an extremely high survival curve for chemotherapy compared with the other trials. Since trial 14 is a large trial with relatively small standard errors, the empirical Bayes estimate of this survival curve has slightly shrunk towards the common mean, but it remains on a higher level than the other survival curves.

5. DISCUSSION

In 1994, Dear proposed a general linear model with the survival estimate as the dependent variable, and follow-up time, treatment, and study as categorical covariates. The parameters are estimated in an iterative manner by weighted GLS. The main drawback of the model is that it is a fixed-effects model. In this paper we generalized the GLS method of Dear [11] to a multivariate random-effects framework. The model can be fitted in standard programs such as SAS Proc MIXED. The method can also be considered as a generalization of the DerSimonian–Laird random-effects model for univariate outcomes [2]. For a fixed time t, our model reduces to the DerSimonian–Laird model. The modelling approach is very flexible in that the data set does not need to be balanced. Different studies may provide survival estimates at different time points. This enables the meta-analyst to analyse all available data as provided in the publications, without need to inter- or extrapolate to fixed time points. There is also a lot of freedom in the modelling process. The ln-minus-ln transformation may be replaced by a different transformation function, e.g. the shape of the survival curves could be modelled using regression splines or fractional polynomials, but the standard programs can still be used. We prefer the ln-minus-ln transformation in combination with ln(time) as covariate, since this allows one to interpret the covariate effects as hazard ratios as in a Cox regression model. Moreover, as pointed out by Moodie *et al.* [17], when the hazard ratio is not constant, the difference in $\ln(-\ln(\text{survival}))$ between two groups can still be interpreted as the natural logarithm of a weighted average of the hazard ratio over time.

As always in the meta-analysis of summary data, it is difficult to judge whether the model fits well. In principle a residual analysis, as for a general linear mixed model could be done, but this is hampered by the small number of (correlated) residuals. One of the assumptions in our approach is that the random effects have a normal distribution. This implies that for a fixed point in time the usual DerSimonian–Laird model holds. It is well known that this model is quite robust against the normal distribution assumption [4], and it is plausible that this property is carried over to our model. Other assumptions, such as the proportional hazards assumption in our model (6), can be investigated by adding other functions of time as covariates and see whether the fit of the model improves. Misspecification of the covariance structure can be investigated by, for instance, adding a third random effect. However, in our experience, models with three or more random effects are very difficult to fit and suffer from non-convergence. Similarly, misspecification of the survival transformation can be investigated by trying other transformations, such as the logit transformation, and see whether the fit improves.

Similar to Dear's method, a disadvantage of our approach is that the estimated curves are not forced to be survival curves. They are not necessarily non-increasing, and when extrapolated to 0, the curve does not necessarily start at 1. We do not believe that this is a serious disadvantage in practice, but there is certainly a need for models in which the curves are forced to be real survival curves.

We fitted our models by iterated linear mixed model fits, updating the estimated correlations between survival estimates of the same curve. An open question is whether the empirical Bayes (study-specific) survival estimates or the estimates based only on the fixed part should be used. Additionally, it might be possible to update the standard errors of the transformed survival estimates as well. In our examples, these alternatives gave similar results. Simulation studies could determine the best method.

ACKNOWLEDGEMENTS

We are very grateful to two reviewers for their detailed comments and valuable suggestions, which have led to considerable improvements of this paper.

DOI: 10.1002/sim

REFERENCES

1. Glass GV. Primary, secondary and meta-analysis of research. *Educational Researcher* 1976; **5**:3–8.
2. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**(3):177–188.
3. Hardy RJ, Thompson SG. A likelihood approach to meta-analysis with random effects. *Statistics in Medicine* 1996; **15**:619–629.
4. van Houwelingen HC, Arends L, Stijnen T. Tutorial in biostatistics: advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine* 2002; **21**:589–624.
5. Egger M, Ebrahim S, Smith GD. Editorial: where now for meta-analysis? *International Journal of Epidemiology* 2002; **31**:1–5.
6. Williamson P, Smith C, Hutton J, Marson A. Aggregate data meta-analysis with time-to-event outcomes. *Statistics in Medicine* 2002; **21**:3337–3351.
7. Sargent D. A general framework for random effects survival analysis in the Cox proportional hazards setting. *Biometrics* 1998; **54**(4):1486–1497.
8. Hutton J, Williamson P. Bias in meta-analysis due to outcome variable selection within studies. *Applied Statistics* 2000; **49**:359–370.
9. Williamson P, Marson A, Tudur C, Hutton J, Chadwick D. Individual patient data meta-analysis of randomized anti-epileptic drug monotherapy trials. *Journal of Evaluation in Clinical Practice* 2000; **6**(2):205–214.
10. Rondeau V, Michiels S, Liquet B, Pignon J. Investigating trial and treatment heterogeneity in an individual patient data meta-analysis of survival data by means of the penalized maximum likelihood approach. *Statistics in Medicine* 2008; **27**(11):1894–1910.
11. Dear KBG. Iterative generalized least squares for meta-analysis of survival data at multiple times. *Biometrics* 1994; **50**:989–1002.
12. Sutton A, Abrams K, Jones D, Sheldon T, Song F. *Methods for Meta-analysis in Medical Research*. Wiley: Chichester, 2000.
13. Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomized clinical trials. *Statistics in Medicine* 1991; **10**:1665–1677.
14. Parmar MKB, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Statistics in Medicine* 1998; **17**:2815–2834.
15. Vokó Z. Etiology and prevention of stroke. *The Rotterdam Study* (*Doctoral*). Erasmus University Rotterdam, Rotterdam, The Netherlands, 2000.
16. Hoffmann ST, TenBrook JA, Wolf MP, Pauker SG, Salem DN, Wong JB. A meta-analysis of randomized controlled trials comparing coronary artery bypass graft with percutaneous transluminal coronary angioplasty: one and eight-year outcomes. *Journal of the American College of Cardiology* 2003; **41**:1293–1304.
17. Moodie P, Nelson N, Koch G. A non-parametric procedure for evaluating treatment effect in the meta-analysis of survival data. *Statistics in Medicine* 2004; **23**:1075–1093.
18. Hunink MGM, Wong JB. Meta-analysis of failure-time data with adjustment for covariates. *Medical Decision Making* 1994; **14**:59–70.
19. Shore T, Nelson N, Weinerman B. A meta-analysis of stages I and II Hodgkin's disease. *Cancer* 1990; **65**: 1155–1160.
20. Voest EE, Houwelingen JCV, Neijt JP. A meta-analysis of prognostic factors in advanced ovarian cancer with median survival and overall survival (measured with the log(relative risk)) as main objective. *European Journal of Cancer Clinical Oncology* 1989; **25**:711–720.
21. Reimold SC, Chalmers T, Berlin JA, Antman EM. Assessment of the efficacy and safety of antiarrhythmic therapy for chronic arterial fibrillation: observations on the role of trial design and implications of drug-related mortality. *American Heart Journal* 1992; **124**:924–932.
22. Earle CC, Wells GA. An assessment of methods to combine published survival curves. *Medical Decision Making* 2000; **20**:104–111.
23. Raudenbush SW, Becker BJ, Kalaian H. Modeling multivariate effect sizes. *Psychological Bulletin* 1988; **103**(1):111–120.
24. 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**:2537–2550.
25. Begg C, Pilote L, McGlave P. Bone marrow transplantation versus chemotherapy in acute non-lymphocytic leukemia: a meta-analytic review. *European Journal of Cancer and Clinical Oncology* 1989; **25**:1519–1523.
26. Berkey CS, Anderson JJ, Hoaglin DC. Multiple-outcome meta-analysis of clinical trials. *Statistics in Medicine* 1996; **15**:537–557.

DOI: 10.1002/sim