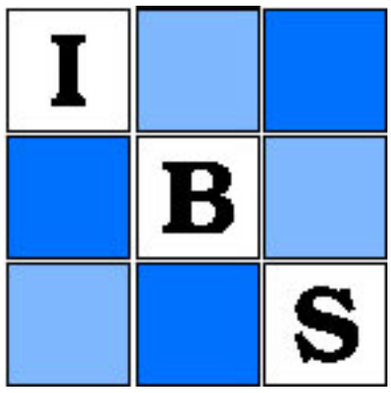

BIOMETRICS 50, 989–1002
December 1994

Iterative Generalized Least Squares for Meta-Analysis of Survival Data at Multiple Times

Keith B. G. Dear*

Technology Assessment Group,
Harvard School of Public Health,
Boston, Massachusetts 02115, U.S.A.

SUMMARY

A method is presented for joint analysis of survival proportions reported at multiple times in published studies to be combined in a meta-analysis. Generalized least squares is used to fit linear models including between-trial and within-trial covariates, using current fitted values iteratively to derive correlations between times within studies. Multi-arm studies and nonrandomized historical controls can be included with no special handling. The method is illustrated with data from two previously published meta-analyses. In one, an early treatment difference is detected that was not apparent in the original analysis.

1. Introduction

Meta-analysts often face the problem that published reports of clinical trials typically provide multiple outcome measures on which a comparison of the treatments might be based. The usual approach to the analysis of such data considers each outcome measure separately, including in each of these restricted analyses only the papers that report that measure. For example, Begg, Pilote, and McGlave (1989) presented separate analyses for each year, from a meta-analysis of six comparative clinical trials and eight single-arm case-series (historical controls), each reporting 3 to 5 years of follow-up. An approach capable of using all the data simultaneously is desirable, if similar effects are found at the different times, to improve power and precision and to provide unified conclusions; but a more important advantage of combined analysis is the ability to fit multiple regression models to the data, using comparisons between the models to test hypotheses about the effects of the treatments on the various outcomes. Such an analysis is necessary, as we illustrate below, to study the behaviour of treatment differences over time, when the same outcome measure is reported repeatedly.

A development in this direction, suggested by Raudenbush, Becker, and Kalaian (1988), requires knowledge of the correlations between the various outcomes. They illustrate their method using data on verbal and math SAT scores, for which they use an exogenous estimate of .6 for the correlation. Here we extend their method to survival analysis, allowing survival reported at multiple times during a trial to be analysed together.

The situation here is similar to the repeated-measures experiments discussed by Verbyla and Cullis (1990). They used parametric models for the covariances, whereas in the present situation approximate theoretical values are provided by the binomial nature of the underlying survival data in the absence of censoring. Verbyla and Cullis used residual maximum likelihood to estimate the dispersion from the data, whereas here the precision of the individual data values is assumed known from the trial reports in the meta-analysis.

1.1 *Choice of Model*

Perhaps the easiest approach to a meta-analysis of survival data would be to summarise each contributing trial by a single number, with its standard error, and use standard methods of meta-analysis to combine them. If a proportional hazards survival model appeared suitable within each

* *Current address:* Department of Statistics, The University of Newcastle; Callaghan, NSW 2308, Australia.

Key words: Clinical trials; Historical controls.

trial, the hazard ratio could be used. Whitehead and Whitehead (1991) discuss the meta-analysis of survival data, combining efficient score statistics for the hazard ratio of an assumed proportional hazards model. Their approach requires statistics such as the log-rank test statistics, or the hazard ratio estimates and their standard errors, and they express doubt over the chances of finding sufficient such details, of adequate quality, in the publications. A second problem is that this approach does not use information from single-arm trials, which cannot provide a summary measure comparing the treatments. In the data presented by Begg et al. (1989), which is Example 1 here, this would mean discarding 40% of the available data (8 of 20 treatment arms were in noncomparative trials); also there is evidence of nonconstant hazard ratios in these data (Table 2). We therefore prefer to model the original survival data.

In modeling proportions, one usually uses a logistic model, partly to avoid nonsensical fitted values but more importantly in the hope that covariate effects may be additive on that scale, so eliminating the need for interaction terms in the model. Our data are not in the form of counts, so we invoke asymptotic normality rather than using binomial error models. Correlated survival data could nevertheless be analysed using a logistic model through the use of generalized estimating equations (Liang and Zeger, 1986). It would not be necessary to use moment estimators of the covariance, since the variances are assumed known from the reports contributing to the meta-analysis, and approximate correlation estimates are available as functions of the fitted values.

In this paper we instead present the use of an ordinary linear model (i.e., with identity link function), estimating parameters by a modified least squares algorithm and assuming multivariate normality for inferences. This ensures easily interpreted parameters, and avoids a strongly parametric model, such as linear dependence on time, which might be the logistic model of choice. It also serves to combine the approaches of Begg and Pilote (1991) and Raudenbush et al. (1988), who both used linear models.

2. Iterative Generalized Least Squares

The general linear model can be written $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{X}$ is a matrix of design and covariate values and $\boldsymbol{\varepsilon}$ is a vector of random errors with expectation zero. If the errors are independent with equal variance, i.e., $\text{var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}_n$, then ordinary least squares is appropriate for estimating the parameters $\boldsymbol{\beta}$. If the errors are normally distributed and correlated, with $\text{var}(\boldsymbol{\varepsilon}) = \mathbf{V}$ for some positive-definite matrix $\mathbf{V}$ other than the identity matrix, then generalized least squares (GLS) provides the fully efficient maximum likelihood estimators.

In a meta-analysis of trials reporting survival proportions, the vector of data $\mathbf{Y}$ will consist of the reported proportions. For example, if a trial provides data y_{ij} for two years ($j = 1, 2$) on treatment $i = 1$, but for three years ($j = 1, 2, 3$) for treatment $i = 2$, then the vector $\mathbf{Y}$ would include five elements for that trial, $y_{11}, y_{12}, y_{21}, y_{22}, y_{23}$. The ordering of the data values is unimportant, except that the rows of $\mathbf{X}$ and $\mathbf{V}$ must correspond. Missing values, such as y_{13} in this example, require no special handling, and noncomparative single-arm studies and case series, serving as historical controls, can be included directly. Begg and Pilote (1991) discuss the importance and validity of including historical controls in any comprehensive assessment of an experimental therapy.

In GLS, the estimate $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ of ordinary least squares is replaced with

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}, \tag{1}$$

with variance $\text{var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$. The generalized residual sum of squares is given by $\text{RSS} = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$, and is distributed proportional to a χ^2_{n-p} distribution, where p is the rank of $\mathbf{X}$. The covariance matrix $\mathbf{V}$ may be assumed known, or known apart from a scale factor, which can be estimated by the residual mean square $\text{MS}_e = \text{RSS}/(n - p)$. In the present application, the variance of the data is assumed known, so MS_e has an expected value of unity, and the goodness of fit of the model can be tested by comparing RSS with the χ^2_{n-p} distribution.

2.1 *Inferences and Residual Analysis*

As in ordinary least squares, composite linear hypotheses can be tested under the assumption of normality by calculating an associated sum of squares and comparing it with either a χ^2 distribution or with an F distribution, depending on whether the variance of the data is known or is estimated. If the hypothesis is represented by $\mathbf{M}\boldsymbol{\beta} = \mathbf{0}$, where $\mathbf{M}$ is a matrix of rank ν, then the sum of squares, on ν degrees of freedom (d.f.), is

$$\hat{\boldsymbol{\beta}}'\mathbf{M}'[\mathbf{M}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{M}']^{-1}\mathbf{M}\hat{\boldsymbol{\beta}}.$$

Here, where the variance of the data is assumed known from the original reports contributing to the meta-analysis, tests will use the appropriate χ^2 distribution, effectively assuming that the expected

generalized error mean square is exactly 1. The sum of squares can most easily be found as the difference in residual sum of squares between two models with and without the terms to be tested. The assumption of normality is approximate, based on each data value having been derived from the survival or not of numerous independent patients in a clinical trial. This approximation will be good in moderate-sized trials, except where the survival proportions are close to 0 or 1. The normality can be examined through a probability plot of the residuals, after transformation to identity covariance. Since the covariance structure of the data is mainly not empirical (the variances are assumed known and the correlations are functions of the fitted means), this should work well. This technique can also identify outliers, as occurred in Example 1 below.

2.2 *Computation*

Fitting a GLS model requires that the correlation matrix of the response variables be known. The method proposed here uses survival probabilities estimated from the model to calculate the correlations. The iterative scheme involves using those correlations to estimate new survival probabilities, and so on until the change is smaller than a specified tolerance. From the correlation matrix $\mathbf{C}^{(m)}$ at iteration m, we obtain the covariance matrix $\mathbf{V}^{(m)}$ from the elementwise product

$$\mathbf{V}^{(m)} = (\mathbf{s}\mathbf{s}') \cdot \mathbf{C}^{(m)},$$

where $\mathbf{s} = \{s_u\}$ is the column vector of reported standard errors of the data elements y_u, and the index u ranges over all the data elements from all the contributing studies. This scaling reproduces the original variances s_u^2 of the data values and assigns covariances to agree with the estimated correlations. Parameter estimates $\hat{\boldsymbol{\beta}}^{(m)}$ are next calculated using (1), and these provide estimated survival probabilities (fitted values) $\hat{\mathbf{Y}}^{(m)} = \mathbf{X}\hat{\boldsymbol{\beta}}^{(m)}$. Finally, an updated correlation matrix, for use in fitting the linear model in the next iteration, is calculated as a function of these probabilities:

$$\mathbf{C}^{(m+1)} = \mathbf{C}(\hat{\mathbf{Y}}^{(m)}).$$

At each iteration the correlations change, and so the covariances used in the model change. The variances of the individual data values remain as reported.

The Appendix derives the correlation function $\mathbf{C}(\mathbf{Y})$, using the natural assumption of conditionally binomial counts of events in consecutive time intervals, and also assuming that there is no censoring in the data. This second assumption will rarely be true, but is convenient to permit the derivation, and also because the rate of censoring is often not reported. The Appendix reports a small simulation study, showing that the correlations will be reduced only slightly by censoring in the data. For example, in one setup where about 6% of observed events were censorings rather than failures, the correlation between two survival estimates was reduced by about 1%, from .614 to .608. The reduction appears to be linear in the rate of censoring. In runs of 10^5 simulated clinical trials, with various degrees of censoring, results identical within the sampling error of the simulation were found whether the trials were of 20 or of 100 patients each. Errors of this magnitude introduced into the covariance calculation will be less than rounding and other errors in the reported variances of the individual estimates, and can reasonably be ignored.

It is advantageous to estimate the correlations in this iterative way for two reasons. First, more precise estimates can be obtained using all the available data. Second, correlation estimates obtained directly from the survival probabilities reported in the individual studies will be exactly 1 whenever a study reports equal survival proportions at two consecutive times. Though not unusual, this feature is troublesome because it imposes additional unwanted structure on the model, effectively requiring that the fitted survival probability remain constant over such intervals. Our present approach avoids the 1's by estimating the correlations from the modeled probabilities, which are less likely to be constant over any interval.

The pseudo-code shown in Figure 1 describes the iterative process. The data are in the vector **y**, standard errors are in **s**, the full-rank design matrix is **X**, and the vector named "group" identifies the treatment arms; data in different groups are uncorrelated. The prime character denotes a transpose, and the "#" operator performs element-wise, row-wise, or column-wise multiplication, as determined by the dimensions of the operands. The "solve" function solves sets of linear equations; if **V** is of full rank, then solve(**V**, **y**) returns $\mathbf{V}^{-1}\mathbf{y}$. The iterations start with an identity correlation matrix **C** and parameters beta ($p \times 1$) set to 0.

Computer code to fit the model, written in the SAS/IML matrix language (SAS Institute, 1988), is available from the author.

```
n = nrow(X)
p = ncol(X)
C = I(n)
beta = repeat(0, p, 1)

do
  V = s#C#s'
  old = beta
  beta = solve(X'*solve(V, X), X'*solve(V, y))
  fv = X*beta
  do i = 1 to n - 1
  do j = i + 1 to n
    if group[i] = group[j] then do
      rho = sqrt(fv[j]#(1 - fv[i])/fv[i]/(1 - fv[j]))
      C[i, j] = rho
      C[j, i] = rho
    end
  end
  end
until (max(abs(beta - old)) < tolerance)
```

Figure 1. Pseudo-code for iterative generalized least squares.

3. Example 1: Chemotherapy vs Bone-Marrow Transplantation in Acute Myelogenous Leukemia

Begg et al. (1989) present a meta-analysis of 14 studies on patients with acute myelogenous leukemia. Six of the studies compared two treatments, allogeneic bone-marrow transplantation (BMT) and chemotherapy. Two included only BMT, and the remaining six included only chemotherapy. The available data (Table 1) were the empirical probabilities of disease-free survival at five

Table 1
Data from Begg et al. (1989): *Percent disease-free survival* (*standard error in parentheses*) *by year* (1 *to* 5)

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

Each line shows results from one clinical trial on patients with acute myelogenous leukemia. The first six trials compared bone-marrow transplantation (BMT) with chemotherapy; the other eight trials tested only one of these alternative therapies.

1-year intervals after the start of treatment. All reports gave estimates for at least the first 3 years. Begg et al. summarised the results separately for only the first 4 years, because only two of the seven studies that reported survival at 5 years recorded any change. Their methods are reported in Begg and Pilote (1991). Here we jointly analyse the results from all 5 years.

3.1 *Results of Example 1*

Analysis of residuals (see later) suggested the presence of an outlier: the 70% survival at year 4 under BMT in trial 4. The results reported below were found omitting this data value. There was no year 5 data from this trial, and no year 4 result from the chemotherapy arm.

The minimal model of interest allows different survival rates by treatment, year, and study, according to the linear model

$$E(Y_{ijk}) = \mu + \tau_i + \gamma_j + \psi_k, \tag{2}$$

where μ is the overall mean, τ_i is an effect for treatment i ($i = 1, 2$), γ_j is an effect for year j ($j = 1, \ldots, 5$), and ψ_k is an effect for study k ($k = 1, \ldots, 14$). The usual constraints, $\sum \tau_i = 0$, $\sum \gamma_j = 0$, and $\sum \psi_k = 0$, can be applied to ensure estimability. The covariance matrix of $\mathbf{Y}$ will be estimated by iterative generalized least squares using the methods of Section 2, and for purposes of significance testing $\mathbf{Y}$ will be assumed to follow a multivariate normal distribution. Fitting models separately to each year, as Begg et al. did, is equivalent to including interaction terms $(\tau\gamma)_{ij}$ between treatments and years, and $(\gamma\psi)_{jk}$ between years and studies, and allowing a different error variance in each year. Applying GLS to all the data together permits testing the hypotheses that these interactions are absent, i.e., that the treatment difference and study effects are the same in all years. If so, that is of interest in itself. Moreover, the treatment effect can then be estimated with greater precision by combining information between years.

Fitting the minimal model (2) gives a residual mean square (MS_e) of 1.88 on 65 d.f. If the covariance matrix of the data is assumed known, this estimated scale factor reflects lack of fit of the model. For a well-fitting model it should be close to 1. The χ^2_{65} statistic of 122.2 is very significantly greater than expected, and we can accordingly reject the minimal model. Fitting a model that allows different treatment effects $(\tau\gamma)_{ij}$ in each year reduces the variance scale-factor to 1.07 on 61 d.f., for $\chi^2_{61} = 65.3$ and $P = .33$. This is a satisfactory model. The significant change in residual sum of squares, 57 on 4 d.f., further confirms that the treatment effect is not constant across years. Allowing different study effects in different years then increases the MS_e to 1.37 on 17 d.f. It is therefore consistent with the data to assume constant additive year effects in all studies, an assumption that is necessary to derive any information from the historical control studies. In fitting the larger model, which includes the $(\gamma\psi)_{jk}$ year-by-study parameters, we are effectively discarding the noncomparative data by allowing a model parameter for each data item from those studies.

The remaining possible model of interest, which allows year-by-study effects but not treatment-by-year effects, yields a MS_e of 2.10 on 21 d.f., which is a poor fit ($P = .0023$). The chosen model is therefore

$$E(Y_{ijk}) = \mu + \tau_i + \gamma_j + \psi_k + (\tau\gamma)_{ij}, \tag{3}$$

where, in addition to the previous constraints, we have that $\sum_i (\tau\gamma)_{ij} = 0\ \forall j$, and $\sum_j (\tau\gamma)_{ij} = 0\ \forall i$. For computation, this is more conveniently reformulated as

$$E(Y_{ijk}) = \psi^*_k + (\tau\gamma)^*_{ij}, \tag{4}$$

where to ensure full rank, set

$$(\tau\gamma)^*_{2,5} \equiv -\sum_{i,j \neq 2,5} (\tau\gamma)^*_{ij}.$$

The parameter estimates found for the data of Begg et al. were, for BMT, $(\widehat{\tau\gamma})^*_{1,j}$ ($j = 1, \ldots, 5$) = 18.9, 9.5, 5.8, 5.0, 5.0, and for chemotherapy, $(\widehat{\tau\gamma})^*_{2,j}$ ($j = 1, \ldots, 5$) = 13.1, −6.5, −14.1, −17.3, −19.3. The units are percent survival. Standard errors ranged from 1.6 to 1.9. Table 2 shows

Table 2
Percent disease-free survival (standard error in parentheses)

Year	BMT	Chemotherapy	Hazard ratio	Survival difference: GLS model	Survival difference: Begg et al.
1	59.0 (2.6)	53.2 (2.1)	1.1	5.7 (3.1)	2 (3)
2	49.6 (2.9)	33.6 (2.0)	2.3	16.0 (3.3)	13 (3)
3	45.9 (3.0)	26.1 (2.0)	3.0	19.8 (3.4)	16 (3)
4	45.1 (3.0)	22.8 (2.0)	7.3	22.3 (3.4)	21 (3)
5	45.1 (3.0)	20.8 (2.1)	∞	24.3 (3.4)	(not analysed)

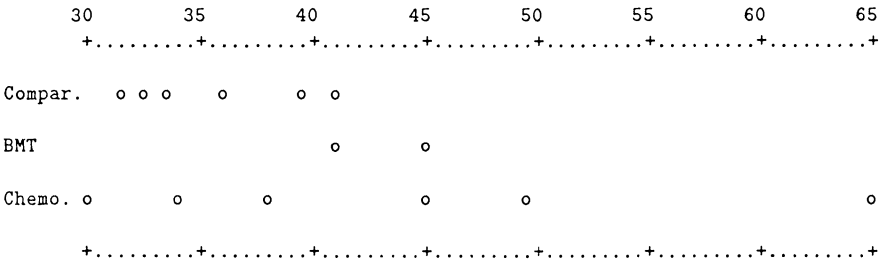

Figure 2. Percent disease-free survival: Trial effects. Values are parameter estimates ψ_k^* from model (4) and represent hypothetical disease-free survival, averaging the effects of BMT and chemotherapy over the first 5 years of therapy. Three types of trial are identified: "Compar." (trials comparing BMT and chemotherapy), "BMT" (trials of BMT only) and "Chemo." (trials of chemotherapy only).

predicted treatment-specific survival profiles from the GLS model, found by adding these zero-sum parameter estimates for the treatment-by-year interaction to the average of the estimated mean-parameters $\hat{\psi}_k^*$ for the 14 studies. The $\hat{\psi}_k^*$ are displayed in Figure 2. Table 2 also shows the estimated hazard ratio between the treatments, a ratio greater than 1 indicating greater hazard under chemotherapy; for example,

$$\left(\frac{53.2 - 33.6}{53.2}\right) \Big/ \left(\frac{59.0 - 49.6}{59.0}\right) \approx 2.3.$$

It appears that the hazard ratio is initially about 1 but increases steeply over the years, and certainly a proportional hazards model is not appropriate here.

3.1.1 *Residual analysis.* Figure 3 shows normal probability plots for the BMT data. The upper plot shows an outlier, with a standardised normal score of 7.1. Although the transformed residuals do not necessarily correspond directly to individual items of data, inspecting the transformation matrix $\mathbf{V}^{-1/2}$ may identify a single data value contributing most of a particular residual. This was the case here; the offending value was the survival rate of 70% to year 4, under BMT in trial 4. Removing this data point and refitting the model gives the lower plot. The plot perhaps still gives some cause for concern over the assumption of normality: the residuals have kurtosis 3.5, like the t-distribution on 16 degrees of freedom. Whether this is sufficient to invalidate the inferences is a matter of judgment.

3.2 *Discussion of Example* 1

Begg et al. (1989) modeled the study effects as random, summarising them in a variance component rather than estimating individual study parameters. However, their estimate of this between-study variance was 0, at least in year 2, the only year reported in Begg and Pilote (1991). The model then effectively becomes an ordinary regression model, with no allowance for study differences. This model can be tested in the GLS setting by calculating the sum of squares for a set of constraints specifying equal trial effects. Doing so gives $\chi^2_{13} = 37.1$, for $P < .001$. Thus it appears that the studies indeed differed in overall disease-free survival, and it is puzzling that the random-effects model found no variation. However, the present model assumes the same study effects in all years, and it may be, despite the absence of a large study-by-year interaction, that the study differences were less apparent in year 2.

Begg et al. discussed in their Section 4 the possibility that the uncontrolled studies were subject to a systematic bias. They used fixed-effect models with parameters expressing this bias separately for the studies of BMT only and the studies of chemotherapy only, and showed that both of these parameters were small. The extra parameters represented differences in the "baseline" survival rates, which are accounted for in the present model by the study parameters which provide for a different baseline survival rate for each separate trial. Similar precautions can nevertheless be taken here, by fitting parameters expressing different year effects for each of the three types of study (controlled, BMT only, and chemotherapy only). Adding such parameters to the model gives an increased RMS of 1.166, and a χ^2_8 statistic of 1.8 for $P = .987$, testing the significance of the reduction in residual sum of squares. It is therefore reasonable to fit models to these data which assume the absence of such bias. Despite the lack of significance, the estimated year effects are

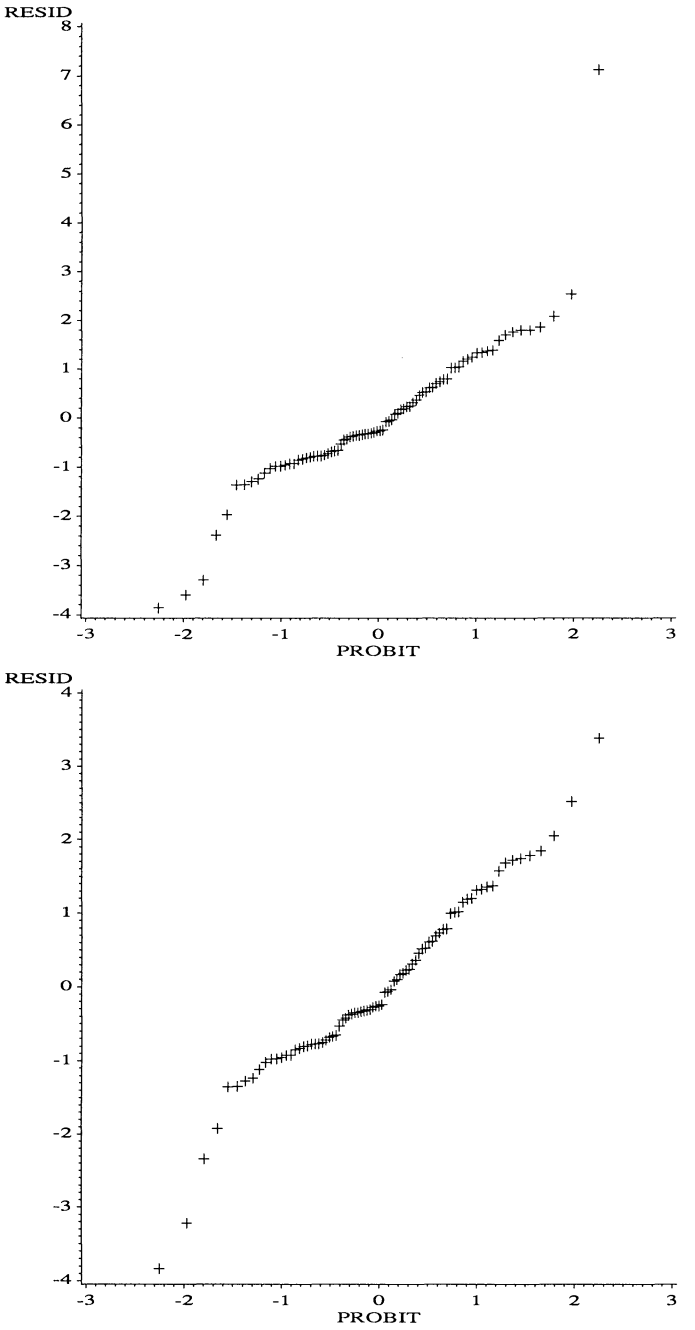

Figure 3. *Upper:* Normal probability plot for BMT data, with outlier. *Lower:* Normal probability plot for refitted model, with outlier excluded.

interesting, in that they appear to suggest that a comparison using only uncontrolled trials would tend to overestimate the benefit of chemotherapy in the first 2 years and underestimate it in years 4 and 5. The five estimated year-effects, summing to 0 for each trial type, were, for the two BMT trials: −3, −2, 1, 2, and 2%, and for the six chemotherapy trials: 3, 2, 0, −1, and −4%. Thus for example it appears that, compared with the survival curve from the controlled trials, survival at year 1 was reduced by 3% relative to the other years in BMT trials, and increased by 3% in chemotherapy trials.

In summarising the data, Begg et al. (1989) write, in part, " ... the studies of allogeneic BMT consistently show 1-year disease-free survival in the region of 55% dropping to a plateau at 3 years of around 40%. By contrast the chemotherapy controls, while having similar one year disease-free survival around 50%, experience substantially poorer long-term disease-free survival in the region

of 20–25%" (p. 1520). The summary for BMT from the GLS analysis is somewhat more favorable, suggesting a 1-year disease-free survival rate of 59%, and an eventual fall only to about 45%. The summary for chemotherapy is essentially as reported by Begg et al.

The model also provides estimates of survival advantage in each year (Table 2) and standard errors of those estimates. The advantage appears to increase with time, at least for four years. The correlations between these estimates are large and positive, ranging from .78 to .98. The differences reported by Begg et al., estimated by the methods of Begg and Pilote (1991), are also shown.

The analysis in Begg et al. is very instructive, and suggests essentially the same broad result as found here, a definite and increasing advantage from BMT. However, our detailed conclusions are somewhat different. Begg et al. (1989) report " ... that the short-term effects of BMT and chemotherapy are essentially equivalent, but that the probability of long-term disease-free survival is substantially higher for allogeneic BMT, at about 40% vs. 20%" (p. 1522). In contrast, the present analysis suggests the possibility of some advantage from BMT even in the first year ($P = .066$, two-sided, comparing 59.0 with 53.3 in Table 2).

This analysis compares two treatments, using data from six studies that tested both treatments and eight more studies that each tested only one of the two. The contribution of the eight studies to the precision of the comparison can be assessed by rerunning the analysis without them. The comparison was made using only four years' data, because the six comparative studies reported no change at all between year 4 and year 5. This analysis of the six studies yielded generally similar estimates of the treatment difference in the four years, but with standard errors larger by 18% on average. That is, including the eight further studies reduced the variances by a factor of about .72. The determinant of the covariance matrix of these four estimates was reduced still more steeply, by a factor of .060 or $(.49)^4$. It therefore appears that substantial benefit, in terms of precision of estimation of treatment difference, has come from including eight studies that did not compare the two treatments!

The benefit gained relies on analysing the various years' data together, and on assuming that the same treatment-specific profile of survival over time applied in all studies: i.e., that there is no time by study interaction. The usefulness of performing each noncomparative study depends in principle on just this assumption, without which the results of the study cannot be applied to medical practice. When the assumption is doubtful, either from external considerations or as a result of a formal test, then study characteristics, such as patient population parameters, should be sought to account for the discrepancies. These characteristics can then be investigated within the present setting by adding study-level covariates to the model.

One concern with this analysis must be that the noncomparative studies might be of lower quality than the controlled trials, in representativeness of the patients and in control of non-experimental factors affecting survival. This might bias their results. However, differences of this sort between the trials, if constant over time, would not affect their contribution to the present model, since such trials contribute information only about differences in survival over time. A constant additive improvement in survival in a trial would be absorbed by the model parameter representing the overall level of that trial, and would have no effect on the estimated treatment difference. It is implicitly assumed, in the additive linear model fitted here, that such trial differences are additive on the untransformed scale of survival probability, as indeed they appear to be in the comparative studies. Other models, such as logistic models, could embody other assumptions.

Figure 2 shows the estimated trial effects, averaged over the five years and the two treatments. The three rows show, respectively, the six comparative trials, the two BMT-only trials, and the six chemotherapy-only trials. The standard errors of these estimates ranged from 2% to 11%. One trial in particular, the last in Table 1, showed surprisingly high average disease-free survival of 65% (standard error 7%). This would not directly influence the results from the GLS model, in which each trial was allowed a parameter to describe its overall level of disease-free survival. However, the random-effects model of Begg et al., particularly when the random component is estimated as 0, would be strongly influenced by this trial, in the direction of improving estimated disease-free survival under chemotherapy.

The random-effects model of Begg and Pilote (1991) permits the analysis to incorporate noncomparative, single-treatment studies, and to derive some benefit from those data despite the lack of any within-trial information on the treatment difference. The present analysis also accomplishes this, despite allowing a fixed-effect parameter for each trial. However, the attractiveness of a random-effects model remains, even in the setting of a combined analysis, as a way of extracting information from the overall level of single-treatment trials, as well as from their year-to-year variation in outcome. As Begg and Pilote point out, the use of such information needs careful justification in the context of each analysis.

4. Example 2: Chemotherapy in Malignant Glioma

Fine et al. (1993) report a meta-analysis of 17 randomized controlled trials that tested the addition of chemotherapy to radiotherapy in postoperative malignant glioma in adults (Table 3). They

Table 3

Data from Fine et al.: Survival percentages (first row) and standard errors (second row)

	Radiotherapy + Chemotherapy				Radiotherapy only			
Months	6	12	18	24	6	12	18	24
Trial								
1	84.2	57.9	21.1	21.0	90.9	54.5	36.4	13.6
	8.4	11.3	9.4	9.4	6.1	10.6	10.3	7.3
2	64.5	51.6	44.4	44.4	62.9	33.3	22.2	16.7
	8.6	9.0	9.1	9.1	8.2	8.1	7.6	7.5
3	61.1	29.2	14.1	4.7	58.8	22.1	4.7	.0
	5.8	5.4	4.2	2.6	6.0	5.1	2.7	2.6
4	85.9	64.4	23.2	7.7	60.0	25.0	18.8	12.5
	7.5	10.9	11.1	7.3	11.0	9.7	9.1	7.9
5	88.3	60.4	37.7	21.3	84.4	40.0	20.0	16.7
	4.0	6.1	6.1	5.1	6.4	8.8	7.2	6.8
6	70.8	43.5	25.5	16.6	69.2	34.6	15.1	12.0
	3.4	3.7	3.2	2.8	4.8	4.9	3.7	3.4
7	92.3	50.0	19.2	11.5	60.0	36.0	20.0	18.0
	5.2	9.8	7.7	11.5	6.9	6.8	5.7	5.4
8	83.6	60.7	30.5	18.6	80.0	54.5	35.3	27.1
	4.7	6.3	6.3	5.7	5.4	6.7	6.6	6.2
9	83.3	63.9	36.1	27.8	68.0	48.0	16.0	16.0
	6.2	8.0	8.0	7.5	9.3	10.0	7.3	7.3
10	96.3	41.9	17.3	12.5	99.0	39.8	12.2	.0
	2.8	7.4	5.6	5.0	4.0	8.3	5.3	5.0
11	68.5	42.9	27.4	20.6	66.8	36.3	20.0	17.0
	2.7	2.9	2.6	2.4	3.0	3.1	2.6	2.6
12	72.2	44.1	25.2	18.8	68.6	32.8	15.1	6.0
	2.3	2.5	2.3	2.0	3.9	4.0	3.1	2.1
13	94.4	57.4	35.0	33.7	92.3	53.9	26.9	21.5
	3.1	6.7	6.5	9.3	5.2	9.8	8.7	8.5
14	93.3	40.0	20.0	27.6	66.7	20.0	6.7	6.7
	3.7	7.3	6.0	8.3	12.2	10.3	6.4	6.4
15	100.0	92.3	84.6	67.7	100.0	76.5	70.6	64.2
	10.0	7.4	10.5	13.6	10.0	10.3	11.1	11.8
16	80.8	44.8	24.4	20.4	79.0	52.6	19.7	6.6
	7.7	9.9	7.3	7.1	9.4	12.4	10.1	6.4
17		56.6		30.7		53.9		39.8
		5.6		5.8		5.4		5.7

estimate the survival advantage separately at 6, 12, 18, and 24 months after surgery. The between-trial variance, estimated by the method of DerSimonian and Laird (1986), was 0 or close to 0 in each of the four analyses, so their conclusions are based on weighted averages of the reported survival probabilities at each time. For most of the contributing studies, the survival estimates were measured from published Kaplan–Meier survival curves, and the variances were derived using Greenwood's formula (see Miller, 1981).

4.1 *Results of Example* 2

Three factors might be considered in models: trials (17 levels), treatments (2 levels), and times (4 levels). Table 4 shows the residual mean square (MS_e) from fitting various models to the data. All the models include the three main-effect terms, and all but the first include one or more interaction terms. There were 132 units of data, four fewer than the full design would indicate, because trial 17 reported at only two of the four time points. The largest model fitted, represented by the last line of Table 4, is

$$E(Y_{ijk}) = \mu + \tau_i + \gamma_j + \psi_k + (\gamma\psi)_{jk} + (\tau\psi)_{ik} + (\tau\gamma)_{ij}, \tag{5}$$

Table 4
GLS *model results: Glioma trials*

Interaction terms included			MS_e	d.f.	Chi-squared *P*-value
(Main effects only)			2.425	111	<.001
		Trt.*Time	2.431	108	<.001
	Trial*Trt.		2.609	95	<.001
	Trial*Trt.	Trt.*Time	2.604	92	<.001
Trial*Time			1.336	65	.037
Trial*Time		Trt.*Time	1.213	62	.121
Trial*Time	Trial*Trt.		1.422	49	.028
Trial*Time	Trial*Trt.	Trt.*Time	1.272	46	.102

The chi-squared test compares the MS_e with its expected value of 1.

where μ is the overall mean, τ_i is an effect for treatment i ($i = 1, 2$), γ_j is an effect for time j ($j = 1, \ldots, 4$), and ψ_k is an effect for trial k ($k = 1, \ldots, 17$). The remaining terms represent two-factor interactions. Because the variance of the data is known, the three-factor interaction is implicitly tested by the goodness of fit of model (5), which is acceptable.

The goodness of fit of a model is measured by comparing its residual sum of squares with the appropriate χ^2 distribution. The evident lack of fit of the first four models in Table 4 indicates the presence of an interaction between trials and times [the $(\gamma\psi)_{jk}$ parameters], suggesting that the differences in survival rates between the trials were not constant over the two years of the trial. This is to be expected, because these trials included different mixtures of patients with two histologies of malignant glioma, one with far worse prognosis than the other, and the trials would therefore be expected to diverge over time.

The significance of the interaction terms is tested in the context of the largest model, to avoid bias in the covariance structure. Comparing the last two models in Table 4, the treatment-by-time interaction on 3 d.f. generates a sum of squares of 11.2 for $P = .011$. This interaction is therefore significant, and we conclude that the difference between the two treatments changes over time. However, the trial-by-treatment interaction, with 16 d.f., generates a sum of squares of only 16.7, for $P = .40$, so there is no evidence of any heterogeneity in treatment effects between the trials. To obtain this sum of squares, refer to Table 4 and subtract the residual sum of squares of the maximal model, 1.272×46, from that of the model without this interaction, 1.213×62.

Table 5
Fitted treatment means: Percent survival (*standard error in parentheses*)

	GLS			Fine et al.		
Months	XRT	Chemo	Difference	XRT	Chemo	Difference
6	77.8 (1.7)	82.0 (1.4)	4.2 (1.9)	78.1 (1.6)	82.1 (1.3)	4.0 (1.8)
12	42.3 (1.9)	52.4 (1.7)	10.2 (2.1)	42.8 (1.7)	52.2 (1.6)	9.4 (2.0)
18	21.6 (1.6)	30.6 (1.6)	9.0 (1.8)	22.1 (1.5)	30.7 (1.5)	8.6 (1.7)
24	16.6 (1.5)	23.5 (1.5)	6.9 (1.6)	16.8 (1.3)	23.5 (1.4)	6.7 (1.5)

Correlations between the four successive GLS estimates of the treatment difference were .44, .59, and .76.

The resulting model of choice is

$$E(Y_{ijk}) = \mu + \tau_i + \gamma_j + \psi_k + (\gamma\psi)_{jk} + (\tau\gamma)_{ij}$$

in conventional terms, or in equivalent full-rank form for computation:

$$E(Y_{ijk}) = \begin{cases} (\gamma\psi)_{jk} + (\tau\gamma)_{ij} & i = 1, \\ (\gamma\psi)_{jk} & i = 2. \end{cases}$$

We are then effectively fitting a separate model, with a single distinct treatment difference, at each time. However, the correlation structure of the data causes the estimates to differ from the equivalent weighted linear model fitted independently at each time. The results from the present model ("GLS") and the separate weighted linear models of Fine et al. are shown in Table 5.

4.2 *Discussion of Example* 2

In the analysis of these data, the chosen model fits the same parameters as are fitted in the separate models of Fine et al. Although the correlation structure assumed in the iterative GLS model perturbs the estimates and standard errors somewhat, the results and conclusions are essentially the same. The advantage of the GLS approach here lies in permitting confirmation that such a full model is really necessary.

5. Conclusion

In the past, medical meta-analysis has mainly been used in statistically rather simple situations, such as the combination of 2×2 tables of discrete outcomes from randomized clinical trials, e.g., Yusuf et al. (1985). With the growing popularity of meta-analysis in recent years (Sacks et al., 1987; Chalmers, Enkin, and Keirse, 1989; Wachter and Straf, 1990; Bailar and Mosteller, 1991), the approach is increasingly being applied to collections of smaller trials, with disparate outcome measures and often with different original objectives. Rather than simply averaging the trial results in one way or another, it then becomes necessary to model the data using trial-level covariates to account for these differences. A recent development incorporated historical controls into such analyses: Begg and Pilote (1991) used a random-effects model to allow for the differences between the contributing studies, which can be expected when different kinds of trial are combined.

The spread of meta-analysis into more complex medical situations also raises the issue of multiple outcome measures, a topic that has recently begun to receive attention (Rosenthal and Rubin, 1986; Raudenbush et al., 1988). The present application of the GLS method provides a way to handle one kind of multiple outcome, namely repeated measures, when each trial contributing to a meta-analysis provides several survival proportions for each treatment, perhaps at yearly intervals as subjects in the trial are followed. Such values clearly cannot be treated as independent data; but once correlations between them are available, they can be modeled using generalized least squares. We have shown how to derive these correlations and incorporate them in linear models.

It is of interest to consider what the contribution of historical controls can be, when the overall level of their data is not permitted to contribute directly to the estimation of the treatment difference. When a single outcome is considered, they can then make no contribution, which is one reason why Begg and Pilote (1991) used instead a random-effects model. In the present method, where a series of outcomes are analysed together, the noncomparative studies contribute information on the shape of the survival curve, i.e., the changes in survival between successive years. This information affects the estimated treatment difference, and because the correlations between data points from the same trial are estimated using the fitted survival probabilities, it will also affect the covariances used in fitting the generalized least squares model.

Consider the limiting situation where there are very many noncomparative trials on each of two treatments, sufficient to fully determine the drops in survival in each year. Because each of these trials will "own" a trial-effect parameter in the model, they provide no information about the treatment effect. This must come entirely from (let us assume) a small number of comparative studies. These comparative studies will then be used to estimate only a single parameter (other than their own trial-effect parameters), which is the treatment effect. None of the information from the comparative studies need be wasted estimating the shape of the survival curves, and the precision of treatment effect estimation will thereby be increased. This increase in precision is the contribution of the historical controls.

The influence of the historical controls may be viewed as a contribution, if one is prepared to accept the model, or as contamination, when one considers the method's robustness to failed assumptions. The precise details of the assumptions made will depend on the model ultimately chosen to represent each data set, but in the first of the two analyses presented above, the model embodied the assumption that, for each treatment, the shape of the survival curve was the same for all studies, whether controlled or uncontrolled. Only the overall level of survival was supposed to vary between studies. An alternative possibility, which was tested above, is that one shape existed in controlled studies, and a different shape in uncontrolled studies of the same treatment. The historical controls can then tell us nothing, even indirectly, about the treatment difference, and the fitted values for the controlled studies will be the same as if only these were analysed. If the assumption is made but is false, then the nature of the induced bias will be that the treatment difference will be overestimated in some years and underestimated in others; the overall magnitude of the estimated treatment effect will not be greatly changed. This high level of robustness is due to the conservative model, which makes no use of the absolute level of survival data from the uncontrolled studies.

A current focus of research in meta-analysis is over the appropriate uses of fixed-effects and

random-effects models [see, for example, Pocock and Hughes (1990, §8); Carlin (1992)]. Both the model of Begg and Pilote (1991), and the GLS model introduced here, treat as fixed the effect of primary interest, namely the treatment difference. The alternative approach, of assuming a random component in the treatment effects, is typified by the method of DerSimonian and Laird (1986). But as Pocock and Hughes suggest, when heterogeneity in treatment difference is evident, simply adding a random component to accommodate it is as unsatisfactory as constraining the fitted differences to be equal.

Begg and Pilote incorporate a random term for the baseline effects of the studies. The present approach instead favors exploratory analysis to identify possible explanations for the observed baseline heterogeneity, and provides a modeling framework to facilitate such exploration. This is made possible by the inclusion of all the available data, from all times reported, in a single analysis, making proper allowance for the covariance of those data.

Acknowledgements

The author gratefully acknowledges valuable contributions from Frederick Mosteller, David C. Hoaglin, Graham A. Colditz, and Philip Hougaard. The work was supported by AHCPR Grant HS-05936-03.

Résumé

On propose une méthode pour l'analyse jointe de taux publiés de survie correspondant à des temps multiples afin de réaliser une méta-analyse. Des méthodes de moindres carrés généralisés sont utilisées pour ajuster des modèles linéaires qui incluent à la fois des covariables inter- et intra-études. La méthode est itérative et elle utilise les valeurs ajustées pour estimer les corrélations entre les temps de survie au sein de chaque étude. Ces méthodes se prêtent aussi aux études à plusieurs bras de traitement ou aux études non randomisées avec témoins historiques. La méthode est appliquée à des données de deux méta-analyses publiées. Dans un cas, la méthode proposée a détecté une différence précoce entre traitements qui n'a pas été mise en évidence dans l'analyse publiée.

References

Bailar, J. C. and Mosteller, F. (1991). *Medical Uses of Statistics*, 2nd edition. Waltham, Massachusetts: NEJM Books.

Begg, C. B. and Pilote, L. (1991). A model for incorporating historical controls into a meta-analysis. *Biometrics* **47,** 899–907.

Begg, C. B., Pilote, L., and McGlave, P. B. (1989). Bone marrow transplantation versus chemotherapy in acute non-lymphocytic leukemia: A meta-analytic review. *European Journal of Cancer and Clinical Oncology* **25,** 1519–1523.

Carlin, J. B. (1992). Meta-analysis for 2 × 2 tables: A Bayesian approach. *Statistics in Medicine* **11,** 141–158.

Chalmers, I., Enkin, M., and Keirse, M. J. N. C. (1989). *Effective Care in Pregnancy and Childbirth*. Oxford: Oxford University Press.

DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* **7,** 177–188.

Fine, H. A., Dear, K. B. G., Loeffler, J. S., and Canellos, G. P. (1993). Meta-analysis of adjuvant chemotherapy for malignant gliomas in adults. *Cancer* **71,** 2585–2597.

Halperin, M., Lan, K. K. G., Ware, J. H., Johnson, N. J., and DeMets, D. L. (1982). An aid to data monitoring in long-term clinical trials. *Controlled Clinical Trials* **3,** 311–323.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73,** 13–22.

Miller, R. G., Jr. (1981). *Survival Analysis*. New York: Wiley.

Pocock, S. J. and Hughes, M. D. (1990). Estimation issues in clinical trials and overviews. *Statistics in Medicine* **9,** 657–672.

Raudenbush, S. W., Becker, B. J., and Kalaian, H. (1988). Modeling multivariate effect sizes. *Psychological Bulletin* **103,** 111–120.

Rosenthal, R. and Rubin, D. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin* **99,** 400–406.

SAS Institute Inc. (1988). *SAS/IML User's Guide, Release 6.03 Edition*. Cary, North Carolina: SAS Institute Inc.

Sacks, H. R., Berrier, J., Reitman, D., Ancona-Berk, V. A., and Chalmers, T. C. (1987). Meta-analysis of randomized clinical trials. *New England Journal of Medicine* **316,** 450–455.

Verbyla, A. P. and Cullis, B. R. (1990). Modelling in repeated measures experiments. *Applied Statistics* **39,** 341–356.

Wachter, K. W. and Straf, M. L. (eds) (1990). *The Future of Meta-Analysis*. New York: Russell Sage.

Whitehead, A. and Whitehead, J. (1991). A general parametric approach to the meta-analysis of randomized clinical trials. *Statistics in Medicine* **10,** 1665–1677.

Yusuf, S., Peto, R., Lewis, J., Collins, R., and Sleight, P. (1985). Beta blockade during and after myocardial infarction: An overview of the randomized trials. *Progress in Cardiovascular Diseases* **27,** 335–371.

Received August 1991; *revised November* 1992 *and March* 1993; *accepted March* 1993.

APPENDIX

Correlations Between Serial Survivorships

The data analysed in the GLS method proposed here consist of the proportions S_j of patients who have not so far experienced some observable event, such as death, at each of a series of times, T_j. We require the correlations between the S_j. Halperin et al. (1982) have also considered the correlation of serial survival estimates, but at calendar-time intervals in the data monitoring of an ongoing clinical trial, rather than the patient-time intervals of survival analysis considered here. For simplicity we here assume no censoring, so that patients leave the group only through occurrence of the event of interest.

The true correlations will be functions of the unknown survival probabilities, which might be estimated directly as the surviving fractions of patients. However, we will not use these probabilities in estimating the correlation. Instead, our best estimate of the survival probability, in a given group of patients, is based on a model fitted jointly to data from these patients and others in the trial. The estimates for each group will then differ from the life-table estimates, and will be more precise if the model is correct.

The correlation between S_j and S_k, where $T_k > T_j$, is the same as the correlation between Y_{0j} and Y_{0k}, where, independently for each patient,

$$Y_{jk} = \begin{cases} 1 & \text{if the patient dies in the interval } T_j \text{ to } T_k, \\ 0 & \text{otherwise,} \end{cases}$$

and the trial started at patient time T_0. (This is because S_j is a linear function of the sum over patients of the independent Y_{0j}.)

Let

$$q_j = E(S_j) = \Pr\{Y_{0j} = 0\}$$

so that

$$\Pr\{Y_{jk} = 1\} = q_j - q_k.$$

Then

$$\begin{aligned} \operatorname{cov}(Y_{0j}, Y_{0k}) &= \operatorname{cov}(Y_{0j}, Y_{0j} + Y_{jk}) \\ &= \operatorname{var}(Y_{0j}) + \operatorname{cov}(Y_{0j}, Y_{jk}) \\ &= q_j(1 - q_j) - (1 - q_j)(q_j - q_k) \\ &= (1 - q_j)q_k. \end{aligned}$$

Thus

$$\operatorname{corr}(S_j, S_k) = \operatorname{corr}(Y_{0j}, Y_{0k}) = \frac{(1 - q_j)q_k}{[q_j(1 - q_j)q_k(1 - q_k)]^{1/2}} = \left[\frac{(1 - q_j)q_k}{q_j(1 - q_k)}\right]^{1/2},$$

which we estimate by replacing the q_j by the fitted values $\hat{S}_j$.

An alternative derivation uses the conditional independence of integrated hazards to write

$$\operatorname{cov}(\hat{\Lambda}_j, \hat{\Lambda}_k) = \operatorname{var}(\hat{\Lambda}_j) \approx V_j/S_j^2,$$

where $\hat{\Lambda}_j$ is the observed hazard to time T_j, and V_j is the variance of the observed survival proportion S_j at T_j. Then

$$\text{corr}(S_j, S_k) \approx \frac{S_j S_k \text{cov}(\hat{\Lambda}_j, \hat{\Lambda}_k)}{(V_j V_k)^{1/2}} = \sqrt{\frac{V_j}{V_k}} \left(\frac{S_k}{S_j}\right).$$

This form does take account of the degree of censoring, through inflation of the reported V_j, but unfortunately it is unusable in practice because the V_j are often known with very low precision, typically being reported as one significant figure of the standard deviation $\sqrt{V_j}$. Estimating the V_j by $S_j(1 - S_j)/n$ returns the previous expression.

A Simulation Study

The formula above was derived under the assumption of no censoring in the data, so that the numbers of failures in successive intervals are binomial random variates. This is unrealistic for most clinical contexts, and the presence of censoring does somewhat reduce the correlations from the value predicted by the simplistic theory. This section reports on a small simulation study conducted to measure this error.

Simulated lifetimes T were exponential with mean 2, and survival estimates were made at times $t = 1$ and $t = 2$. The expected proportions surviving are $e^{-.5} \approx .61$ and $e^{-1} \approx .37$, and the correlation predicted between estimates of these values, in the absence of censoring, is .61444.

Simulated censoring times C were uniform(0, k), with k chosen to provide a desired probability p that an individual will be seen to be censored before $t = 2$; i.e., $C < 2$ and $C < T$. This is satisfied by $k = 2(1 - e^{-1})/p$ provided that $p < .63$ so that $k > 2$. Of events observed before time $t = 2$, a proportion $q = p/[1 - 1/e + p/(e - 1)]$ are expected to be censorings. The resulting correlations r were estimated by simulating 100,000 trials, each of $N = 20$ then each of $N = 100$ patients, with $p = 0(.01).1$.

Using Fisher's approximation that $z = \frac{1}{2}\log[(1 + r)/(1 - r)]$ has variance $1/(n - 3)$, we find standard errors for these estimates of about .0020. Fitting a single straight line through the origin to the change from .61444 gave a slope of $-.1642 \pm .0055$ ($R^2 = .977$), with no observable pattern in the residuals. The fitted correlations are shown in Table 6. The residual standard deviation from this model was .0015 on 21 d.f., compatible with the expected precision of the data.

Table 6
Simulation results

$100p$	0	1	2	3	4	5	6	7	8	9	10
$100q$	0	1.6	3.1	4.6	6.1	7.6	9.0	10.4	11.8	13.1	14.5
r ($N = 20$)	.6120	.6138	.6133	.6119	.6093	.6072	.6052	.6024	.6002	.6000	.5972
r ($N = 100$)	.6138	.6110	.6111	.6112	.6066	.6092	.6041	.6009	.5994	.6007	.5982
Fitted	.6144	.6128	.6112	.6095	.6079	.6062	.6045	.6030	.6013	.5997	.5980

Although only one pattern of survival and censoring processes has been investigated, for only one pair of survival probabilities, these results are sufficient to suggest that the effect of moderate amounts of censoring on the correlation will not be large, and so to justify the use of the simple expressions derived above for the correlation. Assuming that the linearity observed in these data continues, a 10% drop in the correlation would require that nearly 40% of the patients accrued be censored before the second timepoint, an unusually high rate of loss.