STATISTICS IN MEDICINE
Statist. Med. **17**, 2815–2834 (1998)

EXTRACTING SUMMARY STATISTICS TO PERFORM META-ANALYSES OF THE PUBLISHED LITERATURE FOR SURVIVAL ENDPOINTS

MAHESH K. B. PARMAR[1]*, VALTER TORRI[2] AND LESLEY STEWART[1]

[1] *MRC Cancer Trials Office, 5 Shaftesbury Road, Cambridge, U.K.*
[2] *Istituto Mario Negri, Milan, Italy*

SUMMARY

Meta-analyses aim to provide a full and comprehensive summary of related studies which have addressed a similar question. When the studies involve time to event (survival-type) data the most appropriate statistics to use are the log hazard ratio and its variance. However, these are not always explicitly presented for each study. In this paper a number of methods of extracting estimates of these statistics in a variety of situations are presented. Use of these methods should improve the efficiency and reliability of meta-analyses of the published literature with survival-type endpoints. © 1998 John Wiley & Sons, Ltd.

1. INTRODUCTION

Meta-analyses (MAs) are becoming increasingly accepted as the best means of summarizing the results of a number of randomized controlled trials (RCTs) addressing similar questions. Well conducted systematic MAs aim to analyse data from all patients entered into all relevant randomized trials. Those which centrally collect, check and re-analyse raw individual patient data perhaps make the greatest efforts to achieve this aim.[1] Advantages of this approach stem from the increased accuracy and updating of data and the flexibility and extent of analyses that can be done. Less rigorous meta-analyses based only on data extracted from published reports can give different results.[2] However, collecting individual patient data requires considerable time, resources and perseverance, and thus to date this approach has only been adopted in a relatively small number of meta-analyses. Some meta-analyses use summary data provided directly by trialists. This avoids some potential difficulties as unpublished material may be sought (both trials not published and patients excluded from the published analysis) and follow-up may be updated.[3] However, most meta-analyses rely on extracting data from the published literature. Such projects can be done relatively quickly and cheaply and may be used as precursors to designing new trials[4] or indeed for performing meta-analyses collecting individual patient data.[5] Unfortunately, in general the quality of such projects has been poor,[6] and it should be possible to improve further the approach to, and analysis of, meta-analyses based on the published literature.

* Correspondence to: Mahesh K. B. Parmar, MRC Cancer Trials Office, 5 Shaftesbury Road, Cambridge, U.K.

CCC 0277–6715/98/242815–20$17.50

The Cochrane Collaboration has done much to promote good methodology for systematic reviews[7] and the subsequent quantitative synthesis of the results into a meta-analysis, in particular with regard to identifying trials and making the methods used explicit, and promoting the use of an overall systematic and scientific approach. These general methods should be used in all meta-analyses.

The statistical methods adopted for time to event outcomes, such as length of survival, in many meta-analyses of published data do not use all the available information to the full. For example, researchers often adopt the method of comparing the numbers of events in each group or the odds of survival at one or two fixed points in time.[8,9] This is inefficient and could lead to inappropriate conclusions.

This paper presents some methods for performing more reliable analyses of the published literature for survival type data. The methods focus on approaches to extracting the data from publications and are illustrated throughout with real examples. Difficulties and problems with the methods are discussed and some solutions and guidance for performing meta-analyses of the published literature are presented.

2. SUMMARY STATISTICS

When reporting a randomized controlled trial with survival-type data it is recommended that the appropriate summary statistics are the log hazard ratio and its variance.[10] The log hazard ratio has been specifically designed for comparing two survival curves, because it is the only summary statistic which allows for both censoring and time to an event. Note that if time to an event and censoring are not included in the calculation of the log hazard ratio then it just becomes the log relative risk.

Although it is often stated that the use of the log hazard ratio implicitly assumes proportional hazards for the two groups being compared, this is not strictly true. The hazard ratio is a summary of the difference between two Kaplan–Meier curves and represents the overall reduction in the risk of death on treatment compared to control over the period of follow-up of patients. It is most easily interpretable when the hazards are proportional, but it is still a valid and useful statistic when they are not. To allow appropriate interpretation of this statistic, however, it is important to have a graphical summary in the form of survival curves and perhaps also in the form of hazard function plots. An approach to forming such a graphical summary is presented in Section 7.

From these summary statistics, an estimate of the treatment effect in the form of the hazard ratio can be calculated, together with a p-value and confidence intervals. These summary statistics can also be used to combine results from each trial to perform a meta-analysis. Throughout this paper it is assumed that the aim is to obtain numerical estimates of these summary statistics from each trial and then to perform a stratified analysis to combine the results. A number of methods have been proposed for combining the summary statistics from each trial into an overall summary statistic.[11] This is not the focus of this paper, and, therefore, here we shall concentrate on the simplest.

For simplicity it is assumed that each trial within an MA is comparing a research arm, r, with a control arm, c. Suppose that there are K trials, and for each trial, $i = 1, \ldots, K$, an estimate of the log hazard ratio, $\ln(\mathrm{HR}_i)$ and its variance $\mathrm{var}(\ln(\mathrm{HR}_i))$, are available, then an estimate of the overall log hazard ratio across trials, $\ln(\mathrm{HR})$, and its variance, $\mathrm{var}[\ln(\mathrm{HR})]$, are given by the

expressions:

$$\ln(\mathrm{HR}) = \frac{\sum_{i=1}^{K} \dfrac{\ln(\mathrm{HR}_i)}{\mathrm{var}[\ln(\mathrm{HR}_i)]}}{\sum_{i=1}^{K} \dfrac{1}{\mathrm{var}[\ln(\mathrm{HR}_i)]}} \tag{1}$$

and

$$\mathrm{var}[\ln(\mathrm{HR})] = \left[\sum_{i=1}^{K} \frac{1}{\mathrm{var}[\ln(\mathrm{HR}_i)]} \right]^{-1}. \tag{2}$$

The overall log hazard ratio is therefore just a weighted average of the individual log hazard ratios, with the weights inversely proportional to the variance of the log hazard ratio of each trial.

However, many trials do not directly report these statistics[1,2] and thus a number of methods of estimating them are presented.

3. ESTIMATING THE LOG HAZARD RATIO AND ITS VARIANCE DIRECTLY

For each individual trial, two direct methods for estimating the log hazard ratio, $\ln(\mathrm{HR}_i)$, and its variance are:

$$\ln(\mathrm{HR}_i) = \ln\left(\frac{O_{\mathrm{r}i}/E_{\mathrm{r}i}}{O_{\mathrm{c}i}/E_{\mathrm{c}i}}\right) \tag{3}$$

and

$$\ln(\mathrm{HR}_i) = \left(\frac{O_{\mathrm{r}i} - E_{\mathrm{r}i}}{V_{\mathrm{r}i}}\right) \tag{4}$$

where, for trial i:

$O_{\mathrm{r}i}$ = observed number of events in the research group;
$O_{\mathrm{c}i}$ = observed number of events in the control group;
$E_{\mathrm{r}i}$ = logrank expected number of events in the treated group;
$E_{\mathrm{c}i}$ = logrank expected number of events in the control group; and
$1/V_{\mathrm{r}i}$ = Mantel–Haenszel variance of the log hazard ratio.

Of the two estimates, the expression in equation (3) is preferred because that is the formal definition of the log hazard ratio. Nevertheless, these two estimates will only differ markedly when the total number of events in a trial is small, perhaps less than 25.[13]

Estimates of the variance of the two log hazard ratio estimates given in (3) and (4) are respectively given by:

$$[(1/E_{\mathrm{r}i}) + (1/E_{\mathrm{c}i})] \tag{5}$$

and

$$1/V_{ri}. \tag{6}$$

Again these will only differ markedly when the total number of events in a study is small.

3.1. Example 1

The Veterans Administration and Surgical Group (VASOG) randomized 865 patients with lung cancer to receive either surgery alone or surgery plus chemotherapy.[14] On the primary endpoint of length of survival the authors give the observed and logrank expected number of deaths on surgery alone (control) as 191 and 204·6 and on surgery plus chemotherapy (research) as 212 and 198·4, respectively. (Note, there is an obvious typographical error in this paper which gives the expected number of deaths on control as 304·6). Using the expression in equation (3) the estimated log hazard ratio can be calculated as $\ln[(212/198{\cdot}4)/(191/204{\cdot}6)] = 0{\cdot}135$ and an estimated variance of this estimate is given by $(1/198{\cdot}4) + (1/204{\cdot}6) = 0{\cdot}00993$.

4. ESTIMATING THE LOG HAZARD RATIO OR ITS VARIANCE INDIRECTLY

4.1. Indirect variance estimation

Sometimes an estimate of the log hazard ratio is given but its variance is not given. In this case if a $(1 - \alpha_i) \times 100$ per cent confidence interval is given for the log hazard ratio, then this can be used to estimate the variance of the log hazard ratio. In particular an estimate of the variance of the estimated log hazard ratio is

$$\mathrm{var}(\ln(\mathrm{HR}_i)) = \left[\frac{\mathrm{UPPCI}_i - \mathrm{LOWCI}_i}{2\Phi^{-1}(1 - \alpha_i/2)}\right]^2 \tag{7}$$

where UPPCI_i and LOWCI_i are the value for the upper and lower ends of the confidence interval for $\ln(\mathrm{HR}_i)$. In practice it is more likely that the confidence interval for the hazard ratio is given. In this case UPPCI_i and LOWCI_i are just the logarithm of the upper and lower limits of the hazard ratio, respectively. Clearly, whichever form they are given in, confidence intervals need to be given to sufficient level of accuracy to allow reliable calculations to be made, and we believe that this would usually require an accuracy of at least two significant figures. We note that in practice the 95 per cent intervals are usually given, and thus the denominator inside the square brackets in expression (7) will usually take the value of $2 \times 1{\cdot}96$.

4.2. Example 2

The British Medical Research Council reported the results of a three-arm randomized trial in patients with superficial bladder cancer.[15] The aim of this trial was to compare resection alone against resection plus one instillation of Mitomycin-C (MM-C) against resection plus 5 instillations of MM-C. The three groups were compared pairwise on the endpoint of recurrence-free interval. The results of the comparison of the resection plus one instillation of MM-C versus resection alone are presented in Table I.

Using expression (7) the variance of the estimated log hazard ratio is

$$\mathrm{var}(\ln(\mathrm{HR}_i)) = \left[\frac{\ln 0{\cdot}91 - \ln 0{\cdot}48}{2 \times 1{\cdot}96}\right]^2$$

$$= 0{\cdot}0266.$$

To assess the accuracy of this estimate an estimate of the logrank χ^2 statistic is given by

$$\left[\frac{\ln(\mathrm{HR}_i)}{\sqrt{\{\mathrm{var}(\ln(\mathrm{HR}_i))\}}}\right]^2 = \left[\frac{\ln 0{\cdot}66}{\sqrt{0{\cdot}0266}}\right]^2 = 6{\cdot}48.$$

This gives an associated p-value of 0·011, very close to the actually reported p-value of 0·010.

4.3. Indirect log hazard ratio and variance estimation

In our experience the statistic most frequently quoted with survival type endpoints is the p-value for the logrank or Mantel–Haenszel test. An equivalent statistic is the estimated chi-squared statistic. We assume that there is approximate equivalence between the logrank and Mantel–Haenszel χ^2 statistics, and in fact, in our experience, many researchers do not distinguish between the two. A further important statistic which is often reported is the total number of events observed in the two groups.

For the Mantel–Haenszel version of the logrank test a reasonable estimate of $V_{\mathrm{r}i}$ (as long as the treatment effect is not too large and the randomization ratio in the two groups is 1 : 1) is given by $O_i/4$, where O_i is the total number of deaths in the two groups.[16]

Thus

$$V_{\mathrm{r}i} \approx O_i/4 \tag{8}$$

and

$$\frac{O_{\mathrm{r}i} - E_{\mathrm{r}i}}{\sqrt{V_{\mathrm{r}i}}} = \sqrt{\chi^2_{\mathrm{M-H}}} = \Phi^{-1}(1 - p_i/2) \tag{9}$$

where, $\chi^2_{\mathrm{M-H}}$ is the Mantel–Haenszel version of the logrank statistic, p_i is the reported (two-sided) p-value associated with it, and Φ is the cumulative distribution function of the Normal distribution.

Combining the equations in (8) and (9) gives

$$(O_{\mathrm{r}i} - E_{\mathrm{r}i}) = 1/2 \times \sqrt{O_i} \times \Phi^{-1}(1 - p_i/2). \tag{10}$$

If the randomization ratio in the two groups is not 1 : 1, a general extension of (8) is given by[17]

$$O_i R_{\mathrm{r}i} R_{\mathrm{c}i}/(R_{\mathrm{r}i} + R_{\mathrm{c}i})^2. \tag{11}$$

Another reasonable estimate of $V_{\mathrm{r}i}$ is given by[17]

$$O_{\mathrm{r}i} O_{\mathrm{c}i}/O_i. \tag{12}$$

In equations (11) and (12), $R_{\mathrm{r}i}$ and $R_{\mathrm{c}i}$ are the numbers randomized to the research and control groups, respectively. The best choice of (8)/(11) and (12) for an estimate of $V_{\mathrm{r}i}$ is based on a combination of the magnitude of the log hazard ratio and the number of events observed. Given that these will not usually be known, there are no obvious reasons to prefer one estimate over another. In this situation it is probably prudent to calculate both these estimates and perhaps

Table I. Results of the comparison of resection plus one instillation of MMC versus resection alone in a trial for patients with superficial bladder cancer (see text for details)

Treatment group	Number of patients	Hazard ratio (HR)	95% CI for HR	p-value
Resection + MMC	149	0·66	0·48–0·91	0.010
Resection alone	157			

compare them. In our own experience, however, the estimate given in (8)/(11) has performed very well. Expressions (11) and (12) can be used similarly to before with expression (9) to produce expressions

$$O_{ri} - E_{ri} = \frac{\sqrt{(O_i R_{ri} R_{ci})}}{(R_{ri} + R_{ci})} \times \Phi^{-1}\left(1 - \frac{p_i}{2}\right) \tag{13}$$

and

$$O_{ri} - E_{ri} = \sqrt{\frac{O_{ri} O_{ci}}{O_i}} \times \Phi^{-1}\left(1 - \frac{p_i}{2}\right). \tag{14}$$

Finally, using any of (10), (13) or (14) together with (4) gives an estimate of the log hazard ratio. An estimate of the variance of these estimates is given by the respective estimate of $1/V_{ri}$ from (8), (11) and (12).

4.4. Example 3

The Medical Research Council Working Party on Advanced Cancer of the Cervix have reported the results of a randomized controlled trial comparing radiotherapy (RTX) plus a radiosensitizer (RO-8799) against radiotherapy alone.[18] The results on the endpoint of overall survival are reproduced in Table II. One direct and a number of indirect methods for estimating the log hazard ratio and its variance are clearly available, and these are presented in Table III. It should be noted that to use indirect methods to calculate the log hazard ratio involves calculating its variance first. It can be seen that all of the indirect methods perform well, when compared to the direct method and as a consequence the average of the indirect methods performs well also. The investigators also report the 95 per cent confidence interval for the hazard ratio as 1·01 to 2·48. For comparative purposes the estimate of $\mathrm{var}(\ln(\mathrm{HR}_i))$ using equation (7) is 0·0529, which again compares favourably with the direct method.

The problems with this approach are as follows. It may be that a p-value is not given, or is not given to a sufficient level of accuracy to allow reliable calculations to be made. A p-value (or chi-square) with two significant figures is probably required, and it may not be reported to this level of accuracy. It may not be clear whether the reported p-value is one-or two-sided. However, we suspect that one-sided p-values are rarely used without a clear statement in the methods section. Further, the total number of events are not always given. Finally, the direction of the effect, whether in favour of the research arm or in favour of control, may not be clear, especially where the difference between the treatments is small.

Table II. Results of a randomized trial in advanced cancer of the cervix (see text for details)

Treatment	Number of patients	Deaths		O/E	Log hazard ratio	Logrank	
		Observed (O)	Expected (E)			χ^2	p-value
RTX + RO-8799	91	45	36·20	1·24	0·461	4·05	0·044
RTX alone	92	32	40·80	0·78			

Table III. Estimates of the ln (HR) and associated var(ln(HR)) for a randomized trial in advanced cancer of the cervix

Equation number(s)	(3)	(4)	(10) and (4)	(13) and (4)	(14) and (4)	Average of indirect methods
ln(HR)	0·461	0·458	0·458	0·458	0·465	0·460
Equation number(s)	(5)	(6)	(8) and (6)	(11) and (6)	(12) and (6)	Average of indirect methods
var(ln(HR))	0·0521	0·0521	0·0519	0·0519	0·0535	0·0524

4.5. Use of proportional hazards model estimates

It should be noted that a direct estimate of the log hazard ratio is given by the coefficient for the treatment control comparison in the Cox proportional hazards model, and if this is available it can be used. Occasionally, only the p-value may be available from the proportional hazards model after adjustment for prognostic factors. In this instance this p-value can be used in the expressions above. If the associated standard error for this estimate (which is an estimate of the square root of the variance, $\mathrm{I}/V_{\mathrm{r}}$) is not available, the expression (7) or (8) can be used to obtain an estimate of the variance. There is evidence to suggest, somewhat counterintuitively, that the Cox model does not materially improve the precision of the estimate of the log-hazard ratio compared with the precision given by the Mantel–Haenszel method, and that it largely influences the size of the estimate only.[19]

4.6. Example 4

The Cancer and Leukaemia Group B randomized 180 patients with non-small-cell lung cancer to receive either radiotherapy alone or radiotherapy plus chemotherapy.[20] In the 155 eligible patients there were 68 deaths on radiotherapy alone and 58 deaths on radiotherapy plus chemotherapy. Neither the log hazard ratio nor the hazard ratio are given. However, the (two-sided) p-value for the comparison of the two groups is given from the Cox model as 0·0075.

Inserting these values in equations (10) and (8) gives an estimated $O_r - E_r = 1/2 \times \sqrt{126} \times 2{\cdot}67 = 14{\cdot}99$ and $V_r = 126/4 = 31{\cdot}5$, respectively. These figures give a log hazard ratio of 0·476 with a variance of 0·0317. The authors also give the p-value from the logrank test as 0·0066. Using this value a log hazard ratio of 0·485 is obtained.

5. ESTIMATING THE LOG HAZARD RATIO AND ITS VARIANCE FROM SURVIVAL CURVES

If none of the above approaches can provide estimates of the log hazard ratio and its variance then it may be possible to use the published survival curves to obtain them. A general approach to doing this is as follows. For each trial split the time axis into T non-overlapping intervals. Estimate the log hazard ratio for each interval and then combine them in a stratified way across intervals to obtain an overall log hazard ratio for each trial. The steps in this process and the derivation of the necessary formulae to do this are detailed in (i) to (iv) below:

(i) From the Kaplan–Meier curves for trial i ($i = 1, \ldots, K$) read off survival probabilities for each arm at T prespecified non-overlapping time points ($t = 1, \ldots, T$). Also from reading the manuscript, estimate the minimum and maximum follow-up of patients. This can be done from extracting the information on dates of accrual (if given) and from the date of submission, or perhaps publication of the manuscript. In what follows there is a need to distinguish between the number of patients at risk at the start of a time interval and the number of patients at risk during a time interval. The latter clearly allows for censoring during the time interval. To make this distinction clear we have employed extra notation. In particular for the time interval $(t-1, t)$ we use the notation t_s and t_e to represent the start and end of the interval, retaining the term t to represent the whole interval $(t-1, t)$.

(ii) Calculate the number alive and at risk for each time interval. To do this a model for censoring during each time interval must be assumed. One such model using estimates of the minimum and maximum follow-up has been proposed[2] and a modified form is presented in the Appendix. This method assumes that patients are censored at a constant rate during a time interval and thus contribute a half a person's worth of risk during the time interval. The number of patients at risk of death on the research arm during the time interval $(t-1, t)$ is given by

$$R_{ri}(t) = R_{ri}(t-1) - D_{ri}(t-1) - C_{ri}(t) \tag{15}$$

where $R_{ri}(t-1)$ is the effective number of patients at risk on research arm during time interval $(t-2, t-1)$, $D_{ri}(t-1)$ is the effective number of deaths on research arm during time interval $(t-2, t-1)$ and $C_{ri}(t-1)$ is the effective number of patients censored on research arm during time interval $(t-2, t-1)$, with similar definitions and equations for the control arm. We use the convention that $R_{ri}(0)$ and $R_{ci}(0)$ are the total number of patients included in the analysis for the research and control arms, respectively. Further, that $D_{ri}(0) = D_{ci}(0) = C_{ri}(0) = C_{ci}(0) = 0$.

To calculate the effective number censored during a particular time interval $(t-1, t)$ then we need the effective number of patients alive and at risk at the start of the interval $(t-1, t)$. For the research arm this is given by

$$R_{ri}(t_s) = R_{ri}(t-1) - D_{ri}(t-1)$$

with a similar expression for the control arm. The effective number of patients censored during time interval $(t-1,\ t)$ can then be estimated by using the algorithm set out in the Appendix. Once we have the effective number at risk at the start of the time interval $(t-1,\ t)$ and the effective number censored during the same interval, the effective number at risk during the time interval can be calculated. For the research arm this is given by

$$R_{ri}(t) = R_{ri}(t_s) - C_{ri}(t)$$

with similar obvious expressions for the control arm. We should note that if the numbers at risk are published along with the survival curves then these can be used as estimates of $R_{ri}(t_s)$ and $R_{ci}(t_s)$ at the appropriate times and also may serve as a guide to the accuracy of the methods suggested. The effective number of deaths during time interval $(t-1,\ t)$ on the research arm can then be calculated as follows:

$$D_{ri}(t) = \left[R_{ri}(t) \times \left(\frac{S_{ri}(t_s) - S_{ri}(t_e)}{S_{ri}(t_s)} \right) \right] \tag{16}$$

where $S_{ri}(t_s)$ is the estimate of the survival probability on the research arm at the start of time interval $(t-1,\ t)$ read from the Kaplan–Meier curve and $S_{ri}(t_e)$ is the estimate of the survival probability on the research arm at the end of time interval $(t-1,\ t)$ read from the Kaplan–Meier curve.

(iii) We can now exploit the fact that when time to an event and censoring are not formally included in the calculation, then the hazard ratio can be estimated by the relative risk.[10] Thus an estimate of the log hazard ratio during time interval $(t-1,\ t)$ is given by

$$\ln(\mathrm{HR}_i(t)) = \ln\left(\frac{D_{ri}(t)/R_{ri}(t)}{D_{ci}(t)/R_{ci}(t)} \right). \tag{17}$$

An estimate of the variance of this estimate is given by

$$\mathrm{var}[\ln(\mathrm{HR}_i(t))] = \frac{1}{D_{ri}(t)} - \frac{1}{R_{ri}(t)} + \frac{1}{D_{ci}(t)} - \frac{1}{R_{ci}(t)}. \tag{18}$$

Difficulties with calculating the log hazard ratio and its variance will arise whenever $D_{ri}(t)$ or $D_{ci}(t)$ are equal to zero, that is, there are no deaths in that interval on either the research or control arms. To calculate the log hazard ratio and its variance in such circumstances we suggest that the zero is replaced by a small number of deaths, say 10^{-6}, in that interval. In our experience this approach provides the best estimate of the total number of deaths and overall variance in each arm. We have also found this approach to be preferable to, for example, concatenating time intervals such that there is none with zero deaths in it.

(iv) We note that the hazard ratios $\mathrm{HR}_i(t)$, for $t = 1, \ldots, T$, are independent. Thus an estimate of the overall log hazard ratio for each trial is given by an expression analogous to expression (1). The overall log hazard ratio for each trial can thus be given by a weighted sum of the individual estimates of the log hazard ratio during each time interval $(t-1,\ t)$, where the weights are inversely proportional to the variance of each estimate. In particular

$$\ln(\mathrm{HR}_i) = \frac{\displaystyle\sum_{t=1}^{T} \frac{\ln(\mathrm{HR}_i(t))}{\mathrm{var}[\ln(\mathrm{HR}_i(t))]}}{\displaystyle\sum_{t=1}^{T} \frac{1}{\mathrm{var}[\ln(\mathrm{HR}_i(t))]}}. \tag{19}$$

By analogy with expression (2) an estimated variance of this estimate is given by

$$\mathrm{var}[\ln(\mathrm{HR}_i)] = \left[\sum_{t=1}^{T} \frac{1}{\mathrm{var}[\ln(\mathrm{HR}_i(t))]}\right]^{-1}. \tag{20}$$

An alternative approach to estimating the variance would be to total the number of calculated deaths in each group and use equation (8).

There are a number of problems in estimating the log hazard ratio for a trial from the published survival curves. One important practical point is that the published curves are often too small to read accurately, and enlarging them may not help. The choice of length of time intervals is an important component to this approach. For the best results the time intervals should be chosen such that the event rate within the interval is relatively small. As a rule of thumb, it is suggested that the event rate within each time interval is no more than 20 per cent of those at the beginning of the time interval, although of course this may not always be possible. Of particular interest is the final interval and this should be chosen such that the end of the interval is equal to the estimated maximum follow-up time ($F_{\max}$ in the Appendix). Note that within each trial the time intervals do not need to be of equal length and in fact they will generally not be of equal length because the event rates are likely to change over time. However, there is good reason to choose the time intervals across the trials in the meta-analysis to be the same if overall survival curves for the research and treatment arms are to be constructed. This is discussed further in Section 7.

An assumption in this approach is the one of uniform censoring over time. Further, an assumption is made that information on patients is collected at regular intervals in the trial and that little information on the endpoint is missing because of poor conduct of the trial or because of poor follow-up. Of course this assumption could be relaxed but it is not obvious how this might be done to reflect a realistic situation. A simple solution, for example, would be to add 10 per cent to those censored in any given time interval, which would imply that the term $R_{ri}(t_s)$ would become $1.1 \times R_{ri}(t_s)$ in equation (26).

To illustrate the general approach proposed in this section the calculations and results for a single trial are presented.

5.1. Example 5

Ingle and colleagues[21] randomized 102 women to tamoxifen (TAM) alone or tamoxifen combined with aminoglutethimide and hydrocortisone (TAM + AG); 100 eligible patients were analysed for the survival endpoint and the published Kaplan–Meier curves for the two randomized groups are reproduced in Figure 1:

(i) For this trial $T = 14$ non-overlapping time intervals were chosen and are given in the first column of Table IV. From reading the paper and assessing dates of accrual and publication the minimum and maximum follow-ups were estimated as 12 months and 72 months, respectively. The time intervals were therefore chosen as follows. Three monthly intervals were chosen for the first 30 months, subsequently 6 monthly intervals were chosen from 30 to 48 months. The final interval was chosen to be from 48 to 60 months. There were no events after 60 months, and thus although the maximum follow-up was 72 months there was no need to continue calculations beyond 60 months.

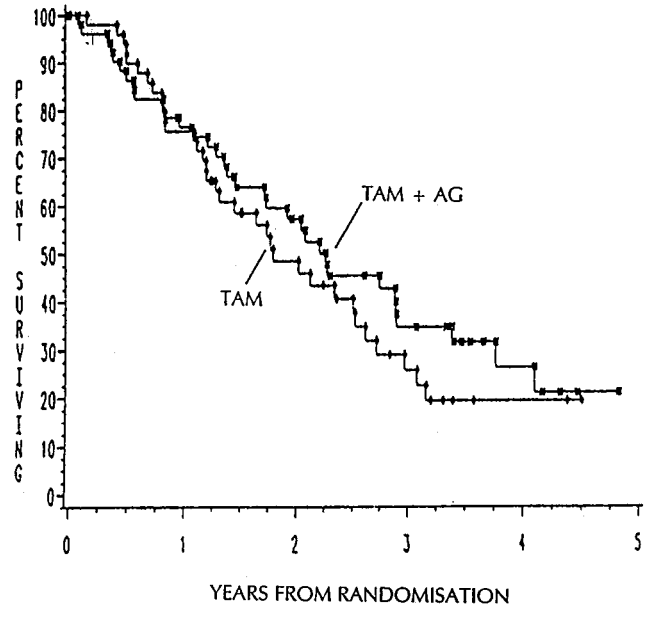

Figure 1. Survival curves for the trial by Ingle *et al.*[18] (reproduced by kind permission of the *Journal of Clinical Oncology*)

(ii) Table IV shows separately for the TAM and TAM + AG arms for each time interval: the survival probability at the start of that interval; the effective number of patients alive at the start of the time interval; the effective number censored in the time interval; the effective number at risk in the time interval, and finally the effective number of deaths in the time interval.

(iii) Table V shows the calculation of the log hazard ratio and its variance for each interval, together with the statistics necessary to calculate the log hazard ratio and its variance for the trial. Note that although no deaths were observed on the TAM arm in the intervals of 42–48 and 48–60 months we have assigned nominal values of 0·000001 deaths in each of these intervals to allow calculation of the log hazard ratio and its variance.

(iv) For the trial then the data from Table V and applying equations (19) and (20) gives a log hazard ratio of $-0{\cdot}244$ (hazard ratio = 0·78) in favour of TAM + AG, with a variance of 0·0550. The authors report a log-hazard ratio of $-0{\cdot}235$ (HR = 0·79). The authors do not give a variance estimate, but an estimated variance for this estimate using the effective total number of deaths from Table V and the indirect methods of Section 4 gives a value of 0·0574.

6. COMPARISON OF SURVIVAL CURVES AND DIRECT/INDIRECT APPROACH

A number of meta-analyses have been performed which included 209 randomized comparisons of the length of survival of treatments for women with advanced breast cancer.[22] These 209 comparisons were studied to assess how frequently the different methods introduced above would

Table IV. Results of the survival curve approach described in point (ii) of Section 5 for the study of Ingle *et al.*[18]

Time interval months (t)	TAM					TAM + AG				
	Survival probability at start of time interval $S_{ci}(t_s)$	Effective number alive at start of time interval $R_{ci}(t_s)$	Effective number censored in time interval $C_{ci}(t)$	Effective number at risk in time interval $R_{ci}(t)$	Effective number of deaths in time interval $D_{ci}(t)$	Survival probability at start of time interval $S_{ri}(t_s)$	Effective number alive at start of time interval $R_{ri}(t_s)$	Effective number censored in time interval $C_{ri}(t)$	Effective number at risk in time interval $R_{ri}(t)$	Effective number of deaths in time interval $D_{ri}(t)$
0–3 (1)	100	49	0	49	0·98	100	51	0	51	2·04
3–6 (2)	98	48·02	0	48·02	3·92	96	48·96	0	48·96	4·08
6–9 (3)	90	44·10	0	44·10	2·94	88	44·88	0	44·88	3·06
9–12 (4)	84	41·16	0	41·16	3·92	82	41·82	0	41·82	2·55
12–15 (5)	76	37·24	0·93	36·31	4·78	77	39·27	0·98	38·29	1·99
15–18 (6)	66	31·53	0·83	30·70	3·26	73	36·30	0·96	35·34	4·36
18–21 (7)	59	27·44	0·76	26·68	2·26	64	30·99	0·86	30·13	1·41
21–24 (8)	54	24·42	0·72	23·70	1·32	61	28·71	0·84	27·87	1·37
24–27 (9)	51	22·38	0·70	21·68	2·98	58	26·50	0·83	25·67	3·10
27–30 (10)	44	18·70	0·62	18·08	0·82	51	22·57	0·75	21·82	1·71
30–36 (11)	42	17·26	1·23	16·03	6·11	47	20·11	1·44	18·67	4·37
36–42 (12)	26	9·92	0·83	9·09	2·10	36	14·30	1·19	13·11	1·09
42–48 (13)	20	7·00	0·70	6·30	0	33	12·02	1·20	10·82	1·97
48–60 (14)	20	6·30	1·57	4·72	0	27	8·85	2·21	6·64	1·23
Total			8·90		35·37			11·27		34·33

Table V. Results of the survival curve approach to obtain estimates of the log hazard ratio and its variance for the study of Ingle *et al.*[18]

Time interval months (t)	Effective number of deaths on TAM in time interval	Effective number at risk on TAM in time interval	Effective number of deaths on TAM + AG in time interval	Effective number at risk on TAM + AG in time interval	Log hazard ratio for time interval	Variance of log hazard ratio for time interval	Log hazard ratio divided by its variance for time interval	Reciprocal of the variance of the log hazard ratio
	$D_{ci}(t)$	$R_{ci}(t)$	$D_{ri}(t)$	$R_{ri}(t)$	$\ln(\mathrm{HR}_i(t))$	$\mathrm{var}[\ln(\mathrm{HR}_i(t))]$	$\frac{\ln(\mathrm{HR}_i(t))}{\mathrm{var}[\ln(\mathrm{HR}_i(t))]}$	$\frac{1}{\mathrm{var}[\ln(\mathrm{HR}_i(t))]}$
0–3	0·98	49	2·04	51	0·69315	1·470588	0·47134	0·68
3–6	3·92	48·02	4·08	48·96	0·02062	0·458951	0·04493	2·178884
6–9	2·94	44·10	3·06	44·88	0·02247	0·621976	0·03613	1·607779
9–12	3·92	41·16	2·55	41·82	−0·445906	0·599051	−0·744354	1·669306
12–15	4·78	36·31	1·99	38·29	−0·929401	0·65806	−1·412334	1·519617
15–18	3·26	30·70	4·36	35·34	0·14999	0·475236	0·31562	2·104216
18–21	2·26	26·68	1·41	30·13	−0·593382	1·081027	−0·548906	0·925046
21–24	1·32	23·70	1·37	27·87	−0·124897	1·409428	−0·088615	0·709508
24–27	2·98	21·68	3·10	25·67	−0·129454	0·57307	−0·225896	1·744989
27–30	0·82	18·08	1·71	21·82	0·54692	1·703168	0·32112	0·587141
30–36	6·11	16·03	4·37	18·67	−0·48762	0·276554	−1·763198	3·615928
36–42	2·10	9·09	1·09	13·11	−1·02196	1·207333	0·846461	0·828272
42–48	0·000001	6·30	1·97	10·82	13·9527	1000000	0·000014	0·000001
48–60	0·000001	4·72	1·23	6·64	13·6812	1000000	0·000014	0·000001
Total	35·37		34·33				−4·440601	18·17069

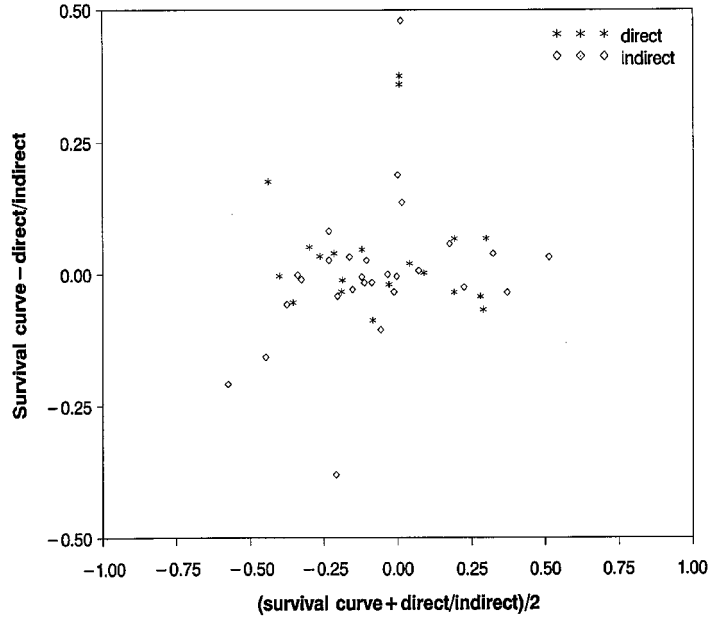

Figure 2. Comparison between the log(HR) estimated from survival curves and the direct/indirect approach (based on 48 trials)

be useful. In 15 comparisons both the log hazard ratio and its variance were given or could be calculated directly. In 50 comparisons, the indirect methods introduced in Section 4 could be used. In 145 comparisons estimation was possible by extracting data from survival curves. To obtain these, estimates of survival were read off at intervals and the methods of Section 5 were applied to these data. Either direct or indirect methods and the survival curve approach could be employed in a total of 48 comparisons.

Of the 209 comparisons it was possible to retrieve some summary data either using the direct/indirect or the survival curve approach in 147 comparisons. Most of the remaining 62 comparisons reported none or little relevant data on overall survival, concentrating mainly on other endpoints such as response of tumour and toxicity.

Figures 2 and 3 show a comparison of the estimates which are obtained using the survival curves and the direct/indirect approach. The survival curve estimate of the log hazard ratio appears to perform reasonably well except in a few cases. Overall there was no evidence of systematic bias in the survival curve estimate with an estimated mean difference between approaches of 0·002 (standard error $= 0{\cdot}029$). For estimating the variance of the log hazard ratio, the survival curve approach again appears to perform reasonably well. However, this approach appears generally to underestimate slightly values given by the direct/indirect approach with a mean difference between approaches of $-0{\cdot}006$ (standard error $= 0{\cdot}002$). This slight underestimate might be expected because the uniform censoring scheme is perhaps a little idealized and a somewhat heavier censoring scheme is likely in practice.

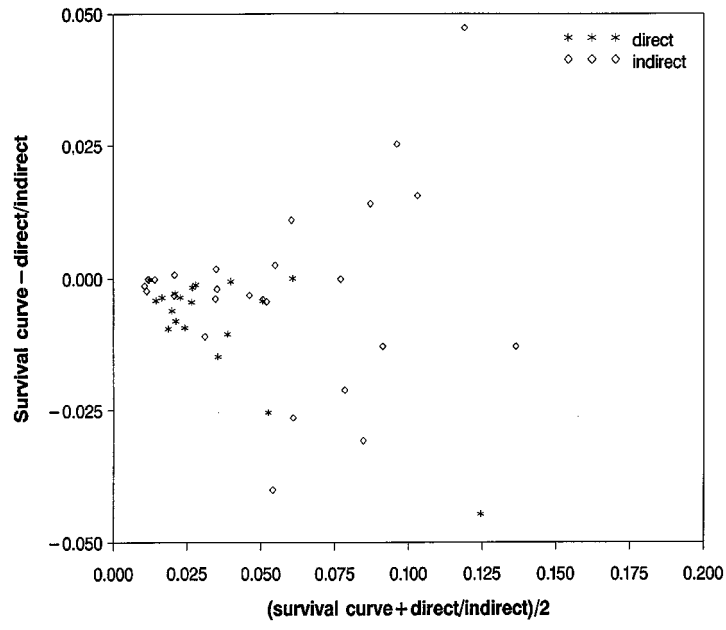

Figure 3. Comparison between the variance of the log(HR) estimated from the survival curves and the direct/indirect approach (based on 48 trials)

7. ESTIMATING THE OVERALL SURVIVAL CURVES ACROSS TRIALS FOR THE RESEARCH AND CONTROL ARMS

Adapting the methods in Section 5, it is possible to produce a graphical estimate of the overall survival curves for the two groups in the following way:

1. Follow steps (i) to (iii) in Section 5.
2. Obtain an estimate of the overall hazard ratio, across all the trials, during each time interval. For the $i = 1, \ldots, K$ trials in the meta-analysis, by analogy with expression (1) for time interval $(t-1, t)$ this can be obtained by

$$\ln(\mathrm{HR}(t)) = \frac{\sum_{i=1}^{K} \dfrac{\ln(\mathrm{HR}_i(t))}{\mathrm{var}[\ln(\mathrm{HR}_i(t))]}}{\sum_{i=1}^{K} \dfrac{1}{\mathrm{var}[\ln(\mathrm{HR}_i(t))]}} \tag{21}$$

and

$$\mathrm{var}[\ln(\mathrm{HR}(t))] = \left[\sum_{i=1}^{K} \frac{1}{\mathrm{var}[\ln(\mathrm{HR}_i(t))]}\right]^{-1}. \tag{22}$$

3. For the end of each time interval calculate an estimate of the proportion alive in the control group across all trials by a weighted average of the control group survivals in each of the

trials. One reasonable weighting scheme is to weight by the number at risk at the end of each interval. Thus, the overall probability of survival on control at time t_e, is given by

$$S_c(t_e) = S_c(t_c - 1) \times \sum_{i=1}^{K} S_{ci}(t_e) \times w_{ci}(t_e) \tag{23}$$

where

$$w_{ci}(t_e) = \frac{R_{ci}(t_e)}{\sum_{i=1}^{K} R_{ci}(t_e)} \tag{24}$$

with the convention that $S_c(0) = S_{ci}(0) = 1$. To calculate the overall probability of survival on treatment we assume that during each time interval the relative risk of death on the research arm to the control arm is constant, that is, the hazards are proportional. Under this assumption, the overall probability of survival on the research arm at time t_e is given by

$$S_r(t_e) = \exp\left[\frac{\ln S_c(t_e)}{\mathrm{HR}(t)}\right] \tag{25}$$

with the convention that $S_r(0) = 1$. We emphasize that this method effectively assumes that the log hazard ratio is constant within each time interval, but can vary across time intervals. The main problems with this approach are obviously similar to the problems for the approach in Section 5. However, here there is the added difficulty that one or more trials may not present a survival curve. Such trials clearly cannot be used to form the overall survival curves. A further discrepancy may arise from the order in which things are done. For example, we could adapt expressions (23) and (24) to first evaluate $S_r(t_e)$ rather than $S_c(t_e)$ and then rearrange expression (25) to calculate $S_c(t_e)$. This approach may give different results to the approach presented above. However, we believe that the approach presented is sufficiently accurate to produce an approximate and useful graphical summary of the two survival curves. Nevertheless, only empirical evidence over a interval of time can reliably substantiate this claim.

8. DISCUSSION

Meta-analyses, and systematic reviews generally, are increasingly being recommended as the best means of assessing the efficacy of medical interventions.[2,3] It should be recognized that such projects are not just mechanical exercises and require considerable care and judgement in their design, conduct and analysis. Guidelines have been proposed to aid in many of these aspects.[7] Some of the most difficult methodological problems arise when meta-analyses of time-to-event data are performed, and it could be argued that this is a strong enough case alone for collecting and analysing individual patient data from all relevant trials.[2] However, this will not always be possible and a meta-analysis based on data extracted from the published literature may in some cases be the only practical alternative. Even if a meta-analysis of individual patient data is practical and planned, it may be useful to perform a meta-analysis of the published literature as a precursor to the larger project.

The main aim of this paper is to present some simple methods which can be used to extract the relevant statistics to perform meta-analyses from published papers. Emphasis has been placed on

providing a summary using the log hazard ratio, and its variance to construct point estimates and confidence intervals. There is clear hierarchy in the approaches presented. If a hazard ratio and its variance are available directly for an individual trial, then these values should be used. If either statistic is not available directly then as many of the indirect methods as possible should be used. We suggest a simple average of the indirect estimates should be used as a reasonable alternative to the absence of direct estimates. If insufficient information for direct or indirect estimation is possible for both or either statistic, then we recommend that the published curves are used to try and construct a log hazard ratio. In this case it is important to adjust the number at risk and the number of events to take account of censoring. To do this a model for censoring has to be assumed. One model which assumes non-informative censoring with constant 'drop out' of patients over time is presented in the Appendix.

An aspect of meta-analysis that we have not discussed in any detail is the appropriate methods for combining data across trials. This has been done elsewhere.[11] However, we have presented a technique to combine published survival curves to obtain overall 'meta-analysis' survival curves for the research and control arms. We have done this because our methods to estimate the hazard ratio for a single trial from survival curves for that trial extend easily to forming such overall survival curves. There are two main issues involved in forming such curves – the methods of combining the different sections of the curve across trials and also the practical relevance of the curves. The methods we have used employ a fixed effect type model for each time interval, but they can easily be adapted to employ a random effects model. Perhaps a more important issue is the practical relevance of the curves. As for a single trial the curves present a graphical summary of the hazard ratio also allowing a visual inspection of the appropriateness or otherwise of the proportional hazards assumption over time. However, as for an individual randomized trial there should be no direct assumption that these curves are necessarily representative of a population of patients.

Other researchers have addressed different components of the problems discussed in this paper. Messori and Rampazzo[24] present an algorithm for calculating an overall '"logrank" odds ratio' for a meta-analysis. However, details of the procedure and means of extracting the relevant data are not given in their paper. Hunink and Wong[25] present a method combining data from a number of studies adjusted for differences in case-mix among studies. Their approach has some items in common with the methods presented in Section 5 of this paper. However, their approach assumes proportional hazards for the survival curves within each trial and exponential distributions within each time interval. Chene and Thompson,[26] employ techniques similar in spirit to those presented in Section 4 of this paper for Normally distributed continuous observational data and binary data. Dear[27] proposes an iterative generalized least squares approach to combine survival data at multiple times, but does not deal with the problem of extracting the data from survival curves or methods to allow for censoring. A'hern *et al.*[28] have suggested some similar techniques to those in Section 5 to extract and estimate summary statistics from survival curves, in particular estimating a log hazard ratio at successive timepoints. However, they provide no methods for calculating the variance of estimates from survival curves, nor do they provide a model for censoring. The Early Breast Cancer Trialists Collaboration[29] have presented a method for odds ratios which is similar in spirit to equations (23) and (25) for calculating overall survival curves for the research and control arms.

Our aim in this paper has been to present a series of simple methods to try and extract the relevant data from publications with the aim of performing a meta-analysis. Clearly, such an approach will never be as complete and may not be as secure as collecting individual patient data

from all relevant trials. However, the use of the methods presented in this paper may improve the standard and reliability of meta-analyses based on the published literature.

APPENDIX

If it is assumed that the rate of censoring in a trial is constant and non-informative, then given the numbers randomized and duration of follow-up, the effective number of patients at risk can be calculated. This in effect works out the number of patients with complete follow-up that is equivalent to the sum of the partial follow-ups during a given interval. Intuitively, if 100 patients are censored at a constant rate over a year, then the average follow-up is six months and the effective number of patients is half the total, that is, 50. When the minimum and maximum follow-up are known and if we assume that after the minimum follow-up patients are censored at a constant rate until the point of maximum follow-up, then the number of patients with complete follow-up can be represented as shown in Figure 4. We show the calculations for the research arm for the time interval $(t-1, t)$ where N is the total number of patients on research arm in the analysis, $F_{\min}$ is the minimum follow-up in the trial, $F_{\max}$ is the maximum follow-up in the trial, t_s is the start of time interval $(t-1, t)$, t_e is the end of time interval $(t-1, t)$ and $R_{ri}(t_s)$ is the number at risk on research arm at beginning of time interval $(t-1, t)$.

If $t_s \geqslant F_{\min}$ and $F_{\min} \leqslant t_e \leqslant F_{\max}$ then the effective number of patients censored during time interval $(t-1, t) = \frac{1}{2}x$ (assuming that patients are censored at a constant rate during the time interval).

Using similar triangles

$$\frac{(F_{\max} - t_s)}{R_{ri}(t_s)} = \frac{(t_e - t_s)}{x}$$

$$\Rightarrow x = R_{ri}(t_s)\frac{(t_e - t_s)}{(F_{\max} - t_s)}.$$

Thus, the number of patients censored during time interval $(t-1, t)$

$$= R_{ri}(t_s)\left\{\frac{1}{2}\frac{(t_e - t_s)}{(F_{\max} - t_s)}\right\}. \tag{26}$$

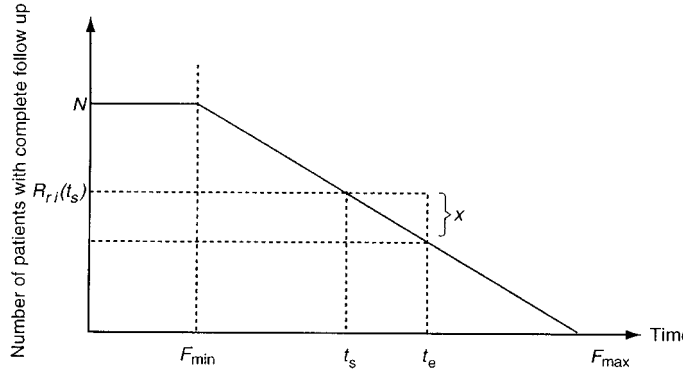

Figure 4. Patients with complete follow-up

If $t_s < F_{min}$ and $t_e < F_{min}$ number censored $= 0$.
If $t_s < F_{min}$ and $F_{min} \leqslant t_e \leqslant F_{max}$ then set $t_s = F_{min}$ in (26).
If $t_s < F_{min}$ and $t_e > F_{max}$, set $t_s = F_{min}$ and $t_e = F_{max}$ in (26).
If $t_s > F_{min}$ and $t_e > F_{max}$, set $t_e = F_{max}$ in (26).

ACKNOWLEDGEMENTS

We thank Angelo Tinazzi for performing the analyses comparing the survival curves and direct/indirect approaches, Doug Altman for comments on an early draft, and Paula Williamson for pointing out that the numbers at risk may be used in the survival curve approach. We also thank both referees for very helpful and constructive comments.

REFERENCES

1. Stewart, L. A. and Clarke, M. J. on behalf of the Cochrane Working Group on Meta-Analysis using Individual Patient Data. 'Practical methodology of meta-analyses (overviews) using updated individual patient data', *Statistics in Medicine*, **14**, 2057–2079 (1995).
2. Stewart, L. A. and Parmar, M. K. B. 'Meta-analysis of literature or of individual patient data: is there a difference?', *Lancet*, **341**, 418–422 (1993).
3. King, J. F., Grant, A., Keirse, M. J. N. C. and Chalmers, I. 'Beta-mimetics in pre-term labour: an overview of the randomised controlled trials', *British Journal of Obstetrics and Gynaecology*, **95**, 211–222 (1988).
4. Gray, R., James, R., Mossman, J. and Stenning, S. on behalf of the UK Coordinating Committee on Cancer Research (UKCCCR) Colorectal Cancer Subcommittee. 'AXIS - a suitable case for treatment', *British Journal of Cancer*, **63**, 841–845 (1991).
5. Tierney, J. F., Mosseri, V., Stewart, L. A., Souhami, R. L. and Parmar, M. K. B. 'Adjuvant chemotherapy for soft tissue sarcoma: review and meta-analysis of published results of randomised clinical trials', *British Journal of Cancer*, **72**, 469–475 (1995).
6. Sacks, H. S., Berrier, J., Reitman, D., Ancona-Berk, V. A. and Chalmers, T. C. 'Meta-analyses of randomised controlled trials', *New England Journal of Medicine*, **316**, 450–455 (1987).
7. Chalmers, I. 'The Cochrane Collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care', *Annals of New York Academy of Science*, **703**, 156–165 (1993).
8. Bhansali, M. S., Vaidya, J. S., Bhatt, R. G., Patil, P. K., Badwe, R. P. and Desai, P. B. 'Chemotherapy for carcinoma of the oesophagus: a comparison of evidence from meta-analysis of randomized trials and historical control studies', *Annals of Oncology*, **7**, 355–359 (1996).
9. Warde, P. and Payne, D. 'Does thoracic irradiation improve survival and local control in limited-stage small cell carcinoma of the lung? A meta-analysis', *Journal of Clinical Oncology*, **10**(6), 890–895 (1992).
10. Parmar, M. K. B. and Machin, D. *Survival Analysis*: *Practical Approach*, Wiley, Chichester, 1995.
11. Whitehead, J. and Whitehead, A. 'A general parametric approach to the meta-analysis of randomised clinical trials', *Statistics in Medicine*, **10**, 1665–1677 (1991).
12. Altman, D. G., De Stavola, B. L., Love, S. B. and Stepniewska, K. A. 'Review of survival analyses published in cancer journals', *British Journal of Cancer*, **72**, 511–518 (1995).
13. Peto, R. and Pike, M. C. 'Conservatism of the approximation $\Sigma(O\text{-}E)^2/E$ in the logrank test for survival data or tumour incidence data', *Biometrics*, **29**, 579–584 (1973).
14. Shields, T. W., Higgins, G. A., Humphrey, E. W., Humphrey, W., Matthews, M. J. and Keehn, R. J. 'Prolonged intermittent adjuvant chemotherapy with CCNU and hydroxyurea after resection of carcinoma of the lung', *Cancer*, **50**, 1713–1721 (1982).
15. Tolley, D. A., Parmar, M. K. B., Grigor, K. M., Lallemand, G. and the Medical Research Council Superficial Bladder Cancer Working Party. 'The effect of intravesical Mitomycin C on recurrence of newly diagnosed superficial bladder cancer: a further report with 7 years follow up', *Journal of Urology*, **155**, 1233–1238 (1996).
16. Tsiatis, A. A. 'The asymptotic joint distribution of the efficient scores test for the proportional hazards model calculated over time', *Biometrika*, **68**, 311–315 (1981).

17. Kalbfleisch, J. D. and Prentice, R. L. *The Statistical Analysis of Failure time Data*, Oxford University Press, Oxford, 1980.
18. Medical Research Council Working Party on Advanced Carcinoma of the Cervix. 'A trial of Ro03-8799 (piminidazole) in carcinoma of the uterine cervix: an interim report from the Medical Research Council Working Party on advanced carcinoma of the cervix', *Radiotherapy and Oncology*, **26**, 93–103 (1993).
19. Anderson, G. L. and Fleming, T. R. 'Model mis-specification in proportional hazards regression', *Biometrika*, **82**, 527–541 (1995).
20. Dillman, R. O., Seagren, S. L., Propert, K. J., Guerra, J., Eaton, W. L., Perry, M. C., Carey, R. W., Frei, E. and Green, M. R. 'A randomised trial of induction chemotherapy plus high-dose radiation versus radiation alone in Stage III non-small cell lung cancer', *New England Journal of Medicine*, **323**, 940–945 (1990).
21. Ingle, J. N., Green, S. J., Ahmann, D. L., Long, H. J., Edmonson, J. H., Rubin, J., Chang, M. N. and Creagan, E. T. 'Randomized trial of tamoxifen alone or combined with aminoglutethimide and hydrocortisone in women with breast cancer', *Journal of Clinical Oncology*, **6**, 958–964 (1986).
22. Cancer Guidance Sub-group of the Clinical Outcomes Group. 'Improving outcomes in breast cancer. The research evidence', Department of Health, London, 1996.
23. Advisory Group on Health Technology Assessment for the Director of Research and Development. 'Assessing the effects of health technologies', Department of Health, London, 1992.
24. Messori, A. and Rampazzo, R. 'Meta-analysis of clinical trials based on censored end-points: simplified theory and implementation of the statistical algorithms on a microcomputer', *Computer Methods and Programs in Biomedicine*, **40**, 261–267 (1993).
25. Hunink, M. G. M. and Wong, J. B. 'Meta-analysis of failure-time data with adjustment for covariates', *Medical Decision Making*, **14**, 59–70 (1994).
26. Chene, G. and Thompson, S. G. 'Methods for summarising the risk associations of quantitative variables in epidemiologic studies in a consistent form', *American Journal of Epidemiology*, **144**, 610–621 (1996).
27. Dear, K. B. G. 'Iterative generalised least squares for meta-analysis of survival data at multiple times', *Biometrics*, **50**, 989–1002 (1994).
28. A'hern, R. P., Smith, I. E. and Ebbs, S. R. 'Chemotherapy and survival in advanced breast cancer: the inclusion of doxorubicin in Cooper type regimens', *British Journal of Cancer*, **67**, 801–805 (1993).
29. Early Breast Cancer Trialists' Collaborative Group. *Treatment of Early Breast Cancer*, *Vol 1*: *Worldwide Evidence 1985–1990*, Oxford University Press, Oxford, 1990.

