Methodology

# Practical methods for incorporating summary time-to-event data into meta-analysis

Jayne F Tierney*[1], Lesley A Stewart[2], Davina Ghersi[3], Sarah Burdett[1] and Matthew R Sydes[4]

Address: [1]Meta-analysis Group, MRC Clinical Trials Unit, London, UK, [2]Centre for Reviews and Dissemination, University of York, UK, [3]School of Public Health, NHMRC Clinical Trials Centre, Sydney, Australia and [4]Cancer Group, MRC Clinical Trials Unit, London, UK

Email: Jayne F Tierney* - jt@ctu.mrc.ac.uk; Lesley A Stewart - ls536@york.ac.uk; Davina Ghersi - ghersid@who.int; Sarah Burdett - sb@ctu.mrc.ac.uk; Matthew R Sydes - ms@ctu.mrc.ac.uk

* Corresponding author

## Abstract

**Background:** In systematic reviews and meta-analyses, time-to-event outcomes are most appropriately analysed using hazard ratios (HRs). In the absence of individual patient data (IPD), methods are available to obtain HRs and/or associated statistics by carefully manipulating published or other summary data. Awareness and adoption of these methods is somewhat limited, perhaps because they are published in the statistical literature using statistical notation.

**Methods:** This paper aims to 'translate' the methods for estimating a HR and associated statistics from published time-to-event-analyses into less statistical and more practical guidance and provide a corresponding, easy-to-use calculations spreadsheet, to facilitate the computational aspects.

**Results:** A wider audience should be able to understand published time-to-event data in individual trial reports and use it more appropriately in meta-analysis. When faced with particular circumstances, readers can refer to the relevant sections of the paper. The spreadsheet can be used to assist them in carrying out the calculations.

**Conclusion:** The methods cannot circumvent the potential biases associated with relying on published data for systematic reviews and meta-analysis. However, this practical guide should improve the quality of the analysis and subsequent interpretation of systematic reviews and meta-analyses that include time-to-event outcomes.

## Background

Time-to-event outcomes take account of whether an event takes place and also the time at which the event occurs, such that both the event and the timing of the event are important. For example, in cancer a cure may not be possible, but it is hoped that a new intervention will increase the duration of survival. Therefore, although the same or similar number of deaths may be observed, it is hoped that a new intervention will decrease the rate at which they take place. Other examples of outcomes where the timing of events may be vital in assessing the value of an intervention include: time free of seizures in epilepsy; time to conception in fertility treatment; time to resolution of symptoms of flu and time to fever in chickenpox.

Odds ratios (ORs) or relative risks (RRs) that measure only the number of events and take no account of when they occur are appropriate for measuring dichotomous outcomes, but less appropriate for analysing time-to-event outcomes. Using such dichotomous measures in a meta-analysis of time-to-event outcomes can pose additional problems. If the total number of events reported for each trial is used to calculate an OR or RR, this can involve combining trials reported at different stages of maturity, with variable follow up, resulting in an estimate that is both unreliable and difficult to interpret. Alternatively, ORs or RRs can be calculated at specific points in time making estimates comparable and easier to interpret, at least at those time-points. However, interpretation is difficult, particularly if individual trials do not contribute data at each time point. Furthermore, bias could arise if the time points are subjectively chosen by the systematic reviewer or selectively reported by the trialist at times of maximal or minimal difference between intervention groups.

Time-to-event outcomes are most appropriately analysed using hazard ratios (HRs), which take into account of the number and timing of events, and the time until last follow-up for each patient who has not experienced an event i.e. has been censored. HRs can be estimated by carefully manipulating published or other summary data [1,2], but currently such methods are under-used in meta-analyses. For example, Issue 3, 2006 of the Cochrane Library contained 43 cancer meta-analyses based on published data that included an analysis of survival and were not conducted by the current authors. Only sixteen of these estimated HRs and the remainder calculated ORs or RRs. This may reflect that the trials included in these meta-analyses did not report the necessary statistical information [3,4] to allow estimation of HRs. However, if there is sufficient data available to estimate an OR or RR, there is usually sufficient data to estimate a HR. Therefore, we suspect that use of the methods is limited because awareness is limited or because the statistical notation used to describe them may be difficult to follow for those with little formal statistical training. Furthermore, it is common for information on the effects of interventions to be presented in a number of different ways and it may not be clear which of the published methods is most appropriate.

Our aim in this paper is to provide step-by-step guidance on how to calculate a HR and the associated statistics for individual trials, according to the information presented in the trial report. To facilitate this we have translated the relevant equations (Appendix 1) from the previously reported statistical methods [1,2] into more descriptive versions, using familiar terms and explaining all arithmetic manipulations as simply as possible. We illustrate their use with data extracted from two cancer trial reports [5,6].

### Basic requirements for a meta-analysis based on hazard ratios

A meta-analysis of HRs, in common with meta-analyses of other effect measures, such as the RR or OR, usually involves a 2-stage process. In the first stage, a HR is estimated for each trial and in the second stage, these HRs are pooled in a meta-analysis. A fixed-effect meta-analysis of HRs, can use the method of Peto[7]:

$$\text{pooled lnHR} = \left[ \frac{\sum logrank\ Observed - Expected\ events(O-E)}{\sum logrank\ Variance(V)} \right]$$

(1)

where $\sum$ is the "sum of" the respective values for each trial and "ln" is the natural logarithm (log). The *logrank Observed minus Expected events (O-E)* and the *logrank Variance (V)* are derived from the number of events and the individual times to event on the research arm of each trial. Alternatively, the inverse variance approach can be used [1]:

$$\text{pooled lnHR} = \left[ \frac{\sum \dfrac{log\ Hazard\ Ratio(\text{lnHR})}{Variance\ of\ the\ \text{lnHR}(V^*)}}{\sum \dfrac{1}{Variance\ of\ the\ \text{lnHR}(V^*)}} \right]$$

(2)

which uses the *Variance of the lnHR (V\*)* and the *log Hazard Ratio (lnHR) for each trial*.

If the HR and V or lnHR and V\* are presented in a trial report, they can be used directly in a fixed effect meta-analysis using (1) or (2) respectively. Similarly, if the coefficient of the treatment effect and the variance from a Cox model are provided, which correspond to the lnHR and V\*, they too be used directly in a fixed effect meta-analysis using (2). These same statistics can be employed if a random effects meta-analysis [8] is required. Where they are not reported however, it is necessary to estimate the *O-E* and *V* or the lnHR and V\* for each trial, in order to combine them in a meta-analysis.

### Generating the O-E, V, HR and lnHR from reported summary statistics

There are many ways to use the summary statistical data presented in trial reports to estimate the *O-E, V, V\*,* HR and lnHR. Some methods use the reported information to directly calculate the HR or lnHR and *V or V\** and are described in Sections 1–2. However, it is more likely that a trial report will only provide sufficient information to estimate some or all of the HR, lnHR, *O-E, V* and *V\** by indirect methods that make certain assumptions, and these indirect methods are described in sections 3–9. For some of these methods, it is necessary to estimate the *V*

and then derive *V\** and others the converse approach. Each is the reciprocal of the other:

$$\textit{Variance of the } \ln HR(V^*) = \frac{1}{\textit{logrank Variance}(V)}$$

(3)

$$\textit{logrank Variance}(V) = \frac{1}{\textit{Variance of the } \ln HR(V^*)}$$

(4)

*V* is used to denote the *logrank Variance* and *V\** to denote the *variance of the lnHR*.

If even these indirect methods cannot be applied, then it may be possible to generate the necessary statistics from published Kaplan-Meier curves (sections 10–11). For any set of trials, it is likely that a number of these methods will be required, and for any one trial, it may be possible to use more than one method.

### Extraction of summary statistics from trial reports

At the outset, it is worthwhile extracting all the necessary descriptive and statistical information for the outcome of interest for each trial [9], using a standard form (e.g. Table 1). The term "research" is used to denote the research intervention and "control" to denote the standard or control arm. Numbers have been rounded to two decimal places for presentation, but not for the underlying calculations. Rounding should in fact be avoided when making these calculations.

### 1. Report presents O & E or hazard rates on research and control arm

If both the observed (*O*) and logrank expected events (*E*) on the research and control arm are presented in a trial report, then the HR can be calculated directly as the ratio of the hazard rates:

$$HR = \left[ \frac{\textit{Observed events research} / \textit{logrank Expected events research}}{\textit{Observed events control} / \textit{logrank Expected events control}} \right]$$

(5)

The associated *V* can also be calculated directly:

$$V = \frac{1}{[(1 / \textit{Expected events research}) + (1 / \textit{Expected events control})]}$$

(6)

These statistics were included in our example report of an ovarian cancer trial [5]:

Observed events research = 34 Expected events research = 28.0

Observed events control = 24 Expected events control = 29.9

Using these data and equations (5) and (6), the HR and *V* can be calculated directly:

$$HR = \frac{34 / 28.0}{24 / 29.9} = 1.51 \quad V = \frac{1}{[(1 / 28.0) + (1 / 29.9)]} = 14.46$$

The *O-E* is the number of observed events minus the logrank expected events on the research arm.

$$O - E = 34 - 28.0 = 6.00$$

If a hazard rate for each of the research and control arms is presented in a trial report they can replace the top and bottom of equation (5). Based on the example above, the hazard rate on the research arm of 1.21 and on control of 0.80 would be used to obtain a HR of 1.51. Such hazard rates cannot be used to calculate directly the associated *V*, which would need to be estimated using an indirect method (see below).

### 2. Report presents O-E on research arm and logrank V

If a trial report presents the *O-E* events on the research arm and *V*, the HR can be calculated directly:

$$HR = \exp \left[ \frac{\textit{Observed} - \textit{Expected events research}(O - E)}{\textit{Variance}(V)} \right]$$

(7)

Note that "exp" represents the exponential or inverse of the natural log. HRs calculated using formula (7) will not differ markedly from the formal definition described previously (5), unless the event rate in a trial is low [1].

For illustration purposes, the data derived from the ovarian cancer trial report [5] are shown:

$$O\text{-}E = 6.00 \quad V = 14.46$$

Using the calculated *O-E* and *V* in equation (7) gives a HR of 1.51:

$$HR = \exp \left[ \frac{6.00}{14.46} \right] = 1.51$$

Note that equation (7) can be re-arranged by simple algebra thus:

$$V = \left[ \frac{O - E}{\ln(HR)} \right]$$

(8)

$$O\text{-}E = \ln(HR) \times V$$

(9)

**Table 1: Suggested data collection form completed with data extracted from the report of the example trial in bladder cancer [6]**

| Trial Reference: *BA06* | (Chemotherapy) | (No chemotherapy) |
|---|---|---|
| Randomisation ratio (e.g. 1:1) | 1 | 1 |
| Patients randomised | 491 | 485 |
| Patients analysed | 491 | 485 |
| Observed events | 229 | 256 |
| Logrank expected events | Not reported | Not reported |
| Hazard ratio, confidence interval (& level e.g. 95%) | 0.85, CI 0.71 to 1.02 (95%) | |
| Logrank variance | Not reported | |
| Logrank observed minus-expected events | Not reported | |
| Hazard ratio and confidence interval (& level e.g. 95%) or standard error or variance from adjusted or unadjusted Cox | Not reported | |
| Test statistic, 2-sided p-value to 2 significant figures (& test used e.g. logrank, Mantel-Haenzsel or Cox) | Not reported, 0.075 (logrank) | |
| Advantage to research or control? | Research | |
| Actuarial or Kaplan Meier curves reported? | Yes, Kaplan Meier | |
| Numbers at risk reported | Yes | |
| Follow-up details | Min = 14 months, Max = 82 months (Estimated from recruitment of 69 months, 11/9 – 7/95 and median follow-up of 48 months) | |

If the HR and *O-E* are reported, you can calculate *V*. Alternatively, if the HR and *V* are reported, you can calculate the *O-E*. Equations (8) and (9) are useful for some of the indirect methods presented later.

Equation (5) is the preferred estimate for the HR, although it will only differ markedly from (7) when the total number of events in a trial is small [1].

### 3. Report presents HR and confidence intervals

Where the HR and its associated confidence interval (CI) are presented in a trial report, *V** (variance of the ln(HR)) and subsequently, if necessary, *V*, can be estimated from the confidence interval (CI) provided the CI is given to two significant figures:

$$V^* = \left[ \frac{\ln(\text{upper CI}) - \ln(\text{lower CI})}{2 \times z \text{ score for upper CI boundary}} \right]^2$$

(10)

The top half of the equation uses the log of the upper and lower CI and the bottom half the z-score for the upper boundary of the confidence interval. In the usual situation of a 95% CI being presented, the corresponding z-score is 1.96. Thus, whenever a trial reports a HR and associated a 95% CI, this version of equation (10) can be used to calculate *V**:

$$V^* = \left[ \frac{\ln(\text{upper 95\%CI}) - \ln(\text{lower 95\%CI})}{2 \times 1.96} \right]^2$$

For a 99% CI, the z-score is 2.58 and for a 90% CI the z-score is 1.64.

To demonstrate this and the rest of the indirect methods we use a report of a trial of chemotherapy versus no chemotherapy for bladder cancer [6]. The data extracted from the trial report data are shown in Table 1.

Inserting the 95% CI (0.71–1.02, Table 1) and the z-score of 1.96 into equation (10):

$$V^* = \left[ \frac{\ln(1.02) - \ln(0.71)}{3.92} \right]^2 = 0.0085$$

and using the estimated *V** (without rounding) in equation (4):

$$V = \frac{1}{0.0082} = 117.07$$

Gives an estimate of the logrank *V* of 117.07. Having both the reported HR of 0.85 and the estimated *V*, the *O-E* equation (9) can be used to obtain an O-E of -19.03

$$O - E = \ln(0.85) \times 117.07 = -19.03$$

Note that if a HR of an event on control versus the research arm is reported rather than vice versa, then a HR of the research arm versus control is obtained by taking the reciprocal of the HR i.e. 1/HR and associated CI.

## 4. Report presents HR and events in each arm (and the randomisation ratio is 1:1)

Where a HR is reported, without the associated CI, but with the numbers of events on each arm, and the randomisation ratio is 1:1, a reasonable approximation of $V$ may be obtained using equation (11):

$$V = \frac{Observed\ events\ research \times Observed\ events\ control}{Total\ events}$$

(11)

Using the relevant data from the bladder cancer trial (Table 1), equation (11) and then equation (9):

$$V = \frac{229 \times 256}{485} = 120.87 \quad O - E = \ln(0.85) \times 120.87 = -19.64$$

Gives an estimate of 120.87 for $V$ and -19.64 for the $O$-$E$.

## 5. Report presents HR and total events (and the randomisation ratio is 1:1)

If only the total number of events is reported along with the HR, the variance can be approximated simply using the total number of events, provided again that the randomisation ratio is 1:1:

$$V = \frac{Total\ observed\ events}{4}$$

(12)

where the total observed events is the sum of the observed events on the research and control arms.

Using the total number of events from the bladder cancer trial report (Table 1) gives an estimate 121.25 for $V$. Using this together with the reported HR and equation (9) gives a figure of -19.70 for the $O$-$E$:

$$V = \frac{485}{4} = 121.25 \quad O - E = \ln(0.85) \times 121.25 = -19.70$$

This particular method of estimating $V$ also provides a simple way of checking (approximately) the plausibility of estimates of $V$ derived using other equations.

## 6. Report presents HR, total events and the numbers randomised on each arm

If the randomisation ratio is not 1:1, methods 4 and 5 are not appropriate and one that accounts for the proportion of patients randomised to each arm is needed. If a report describes an analysis that is not based on all randomised patients; some patients being excluded subsequent to randomisation, then the HR and $V$ should be based on the numbers analysed in the report rather than the numbers randomised, otherwise the precision of the estimate will be exaggerated:

$$V = \frac{Total\ observed\ events \times Analysed\ research \times Analysed\ control}{(Analysed\ research + Analysed\ control)^2}$$

(13)

If more than one analysis is presented, for example, one based on eligible patients and one based on all randomised patients, it is preferable to use the analysis based on all randomised patients.

This method can also be used if the randomisation ratio is 1:1. In the bladder cancer trial report, all randomised patients were included in the analysis and so the number randomised in each arm equals the number analysed (Table 1). Equation (13) can be used to estimate $V$ and equation (9) to estimate the $O$-$E$:

$$V = \frac{485 \times 491 \times 485}{(491 + 485)^2} = 121.25 \quad O - E = \ln(0.85) \times 121.25 = -19.70$$

For a trial that randomised patients according to a 1:1 ratio, but analysed unequal numbers of patients on each arm because, for example, patients were excluded differentially by arm, equation (13) is the preferred indirect method of estimating the variance.

## 7. Report presents p-value and events in each arm (and the randomisation ratio is 1:1)

If only the logrank, Mantel Haenszel or even the Cox regression p-value, and numbers of events on each arm are reported and the randomisation ratio is 1:1, these data can be used to estimate the O-E using:

$$O - E = \sqrt{\frac{Observed\ events\ research \times Observed\ events\ control}{Total\ observed\ events}} \times (z\ \text{score for } p\ value \div 2)$$

(14)

For reliability, it is probably wise to use this method only when the exact p-value is given to at least 2 significant figures [1,2]. As well as the events on each arm and overall, a z-score for the 2-sided p-value divided by 2 is required. If a 1-sided p-value is reported it can be used directly to obtain the z-score. Such a z-score can be derived from either statistical tables or statistical or spreadsheet software (e.g. MS Excel).

A decision to assign a positive or negative value to $O$-$E$ is needed and this depends on whether the direction of the effect is in favour of the research or control arm. This in turn will depend on whether the outcome is positive or negative. For a positive outcome, such as time to pregnancy, more pregnancies and/or a shorter the time to pregnancy on the research arm compared to the control arm, will indicate that the effect is in favour of the research arm. For a negative outcome, such as time to death, fewer deaths and/or a longer time to death on the research com-

pared to the control arm will indicate that the effect is in favour of the research arm. If the results are not statistically significantly in favour of either the research or control arm or if the relative numbers of events on each arm are not provided, it is possible to look for other indicators of the direction of the results, such as the relative numbers of events on each arm, separation of Kaplan-Meier curves or textual descriptions of the results.

The logrank p-value of 0.075 gives a z-score of 1.78 and incorporating this with the number of events on each arm (Table 1) into equation (14):

$$O - E = \sqrt{\frac{229 \times 256}{485}} \times 1.78 = 19.57$$

gives an *O-E* of 19.57. It is clear from the report of the bladder cancer trial that survival favours the research treatment, with fewer deaths and a longer time to death in the research arm. Therefore, the *O-E* will be made negative (-19.57). Then, using equations (11) and (7):

$$V = \frac{229 \times 256}{485} = 120.87 \quad HR = \exp\left[\frac{-19.57}{120.87}\right] = 0.85$$

*V* is estimated as 120.87 and the HR as 0.85.

*8. Report presents p-value and total events (and the randomisation ratio is 1:1)*
A similar equation to (14) can be used if just the p-value and the total number of events are reported, provided the randomisation ratio (or the ratio of patients analysed) is 1:1:

$$O - E = 1/2 \times \sqrt{Total\ observed\ events} \times (z\ score\ for\ p\ value \div 2) \tag{15}$$

Using equation (15):

$$O - E = 1/2 \times \sqrt{485} \times 1.78 = 19.60$$

As before, a sign needs to be applied based on the direction of the results, giving -19.60. Then using (12) and (8):

$$V = \frac{485}{4} = 121.25 \quad HR = \exp\left[\frac{-19.60}{121.25}\right] = 0.85$$

give estimates of 121.25 for *V* and 0.85 for the HR.

*9. Report presents p-value, total events and numbers randomised to each arm*
Where the report presents the p-value, the total events and the numbers randomised on each arm, another equation similar to (14) allows estimation of the *O-E* for trials where the randomisation (or analysis) ratio is not 1:1:

$$O - E = \frac{\sqrt{(Total\ observed\ events \times Analysed\ research \times Analysed\ control)}}{(Analysed\ research + Analysed\ control)} \times (z\ score\ for\ p\ value \div 2) \tag{16}$$

Using (16):

$$O - E = \frac{\sqrt{(485 \times 491 \times 485)}}{(491 + 485)} \times 1.78 = 19.60$$

Applying a negative sign on the basis of the direction of the results (-19.60) and equations (13) and (8):

$$V = \frac{485 \times 491 \times 485}{(491 + 485)^2} = 121.25 \quad HR = \exp\left[\frac{-19.60}{121.25}\right] = 0.85$$

Provides an estimate of 121.25 for the *V* and 0.85 for the HR.

### Generating the O-E, V, HR and lnHR from published Kaplan-Meier curves
Some time-to-event analyses are presented solely in the form of Kaplan-Meier curves [1,10]. It is possible to estimate the HR, lnHR, *O-E* and *V* from a number of time intervals from such curves and pool across these time intervals within a trial to estimate a HR or lnHR that represents the whole curve (section 10–11). Alongside, the reported minimum and maximum follow-up times or the reported numbers at risk can be used, to estimate the amount of censoring in a trial. Otherwise, the estimate of effect would be based on too many patients and so be erroneously precise. If a trial report does not present either the numbers at risk or the actual minimum and maximum follow-up, then it may be possible to estimate the level of follow-up from other information provided (Appendix 2).

### Extraction of curve data from trial reports
A sufficiently large, clear copy of the curve needs to be divided up into a number of time intervals, which give a good representation of event rates over time, whilst limiting the number of events within any time interval. Parmar *et al*. [1], suggest that, as far as possible, the event rate within a time interval should be no more than 20% of those at the start of the time interval. If the curve starts to level off, then few (or no) events are taking place and there is little value in extracting data from this area of a curve. Also, the final interval should not extend beyond the actual or estimated maximum follow-up.

For example, in a trial of metastatic breast cancer, many events (deaths) will occur in the first 3 months, so the curve would need to be split into smaller intervals at the beginning then gradually larger time intervals (e.g. monthly for the first 12 months, 3-monthly to 24 months and then 6-monthly thereafter). However, the curve from

the bladder cancer trial (Figure 1) shows an event (death) rate that is quite high in the earlier parts of the curve, but is subsequently fairly steady. Therefore, the curve was divided into 3-monthly intervals for the first 3 years and 6-monthly intervals thereafter (Figure 1). The percentage survival for each arm at the start of each time interval, for each arm, was then extracted into Table 2.

### 10. Report presents Kaplan-Meier curve and information on follow-up

For each time interval and for each arm a number of iterative calculations are required. It is necessary to estimate the number of patients who were: 1) event-free at the start of the interval, 2) censored during the interval and 3) at risk during the interval. Also, 4) the number of events during each interval needs to be estimated. Together these items are used to: 5) estimate the *O-E*, *V* and HR for each time interval. Finally, 6) the *O-E*, *V* and HR for the whole curve are derived from combining the estimates across time interval.

The numbers of patients at risk at the start of the first time interval is simply the total number analysed on each arm, making step 1 redundant for the first time interval of any curve. Therefore, in the bladder cancer trial, at the start of the 0–3 month time period, there are 491 and 485 patients at risk on the research and control arms, respectively (Table 2).

Based on the median follow-up of 48 months and accrual period of 69 months (Table 1), the minimum follow-up is estimated (Appendix 2) to be 14 months for this trial, and so all patients have complete follow-up and no patients are censored in the 0–3, 3–6, 6–9 and 9–12

month intervals. Therefore, for these time intervals, estimating the number of patients censored (step 2) is not relevant. Beyond 14 months patients are censored and this must be taken into account. Going through the steps 1, 3, 4 and 5 for the prior time intervals, the following were estimated for the 12–15 month time interval:

Event-free at start of prior time interval (12–15 month), research = 382.98

Event-free at start of prior time interval (12–15 month), control = 363.75

Events in prior time interval (12–15 month), research = 24.55

Events in prior time interval (12–15 month), control = 24.25

Censored in prior time interval (12–15 month), research = 0.00

Censored in prior time interval (12–15 month), control = 0.00

Note that these estimated values differ somewhat from the actual reported numbers at risk at 12 months (Table 2), but they can be used to illustrate all the steps of the method, in the presence of censoring, for the 15–18 month interval:

*Step 1. Numbers event-free at the start of the current interval*
This is in fact the number of patients that were event-free at the end of the prior time interval:

**Table 2: Example data extraction form with data extracted from bladder cancer Kaplan-Meier plot in Figure 1.**

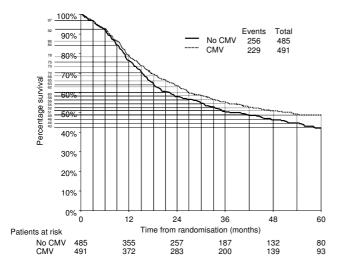| Time at start of interval (months) | % Event-free on research | % Event-free on control | Reported numbers at risk on research | Reported numbers at risk on control |
|---|---|---|---|---|
| 0 | 100 | 100 | 491 | 485 |
| 3 | 97 | 97 | - | - |
| 6 | 92 | 92 | - | - |
| 9 | 86 | 84 | - | - |
| 12 | 78 | 75 | 372 | 355 |
| 15 | 73 | 70 | - | - |
| 18 | 68 | 63 | - | - |
| 21 | 65 | 60 | - | - |
| 24 | 62 | 58 | 283 | 257 |
| 27 | 60 | 56 | - | - |
| 30 | 58 | 54 | - | - |
| 33 | 56 | 52 | - | - |
| 36 | 54 | 51 | 200 | 187 |
| 42 | 52 | 49 | - | - |
| 48 | 51 | 46 | 139 | 132 |
| 54 | 49 | 44 | - | - |
| 60 | 49 | 43 | 93 | 80 |

**Figure 1**
Bladder cancer trial Kaplan-Meier plot (modified with permission [6]), schematically divided into time intervals for data extraction into Table 2.

*Event free at start of current interval = Event free at start of prior interval - Events in prior interval - Censored during prior interval* (17)

Using the data from 12–15 month time interval, the numbers of patients event-free in the current 15–18 month time interval are estimated:

*Event free at start (15–18 month), research* = 382.98 - 24.55 - 0 = 358.43

*Event free at start (15–18 month), control* = 363.75 - 24.25 - 0 = 339.5

*Step 2. Numbers censored during the current interval*
Assuming that censoring is non-informative and that patients are censored at a constant rate within a given time interval, a simple method can be used to estimate numbers censored [1]:

$$At\ risk\ during\ current\ interval \times \left( \frac{1}{2} \times \frac{(End\ of\ time\ interval - Start\ of\ time\ interval)}{(Maximum\ follow\ up - Start\ of\ time\ interval)} \right)$$ (18)

Using the data from step 1, the estimated maximum follow-up of 82 months and equation (18):

$$Censored\ (15-18\ month),\ control = 339.50 \times \left( \frac{1}{2} \times \frac{(18-15)}{(82-15)} \right) = 7.60$$

$$Censored\ (15-18\ month),\ research = 358.43 \times \left( \frac{1}{2} \times \frac{(18-15)}{(82-15)} \right) = 8.02$$

around 8 patients in the research arm and 7 patients in the control arm were estimated to be censored during 15–18 month time interval:

*Step 3. Numbers at risk during the current interval, adjusted for censoring*
The numbers censored can be used to adjust (reduce) the numbers at risk during the time interval:

*At risk during current interval, adjusted for censoring = Event free at start of current interval - Censored during current interval* (19)

Based on the data from step 1 and 2, the numbers at risk during the current 15–18 month time interval are:

*At risk during, adjusted for censoring (15 – 18 month), research* = 358.43 - 8.02 = 350.41

*At risk during, adjusted for censoring (15 – 18 month), control* = 339.50 -7.60 = 331.90

*Step 4. Number of events during the current interval*
The number of events during the interval is then estimated from the reduced numbers at risk:

$$Events\ in\ current\ interval = At\ risk\ during\ current\ interval \times \left( \frac{\%\ Event\ free\ at\ start - \%\ Event\ free\ at\ end}{\%\ Event\ free\ at\ start} \right)$$ (20)

Using the numbers at risk during the interval from step 3 and the data extracted from the curve (Table 2) in equation (20), allows estimation of the number of events in the 15–18 month interval:

$$Events\ during\ (15-18\ month),\ research = 350.41 \times \left( \frac{73-68}{73} \right) = 24.00$$

$$Events\ during\ (15-18\ month),\ control = 331.90 \times \left( \frac{70-63}{70} \right) = 33.19$$

*Step 5. Estimate the HR, V and O-E for the current interval*
As time to event and censoring have already been accounted for, the hazard ratio can be estimated by using the equation for calculating a relative risk:

$$HR = \left( \frac{Events\ research / At\ risk\ research}{Events\ control / At\ risk\ control} \right)$$ (21)

with associated *V*:

$$V = \frac{1}{[(1/Events\ research) - (1/At\ risk\ research) + (1/Events\ control) - (1/Events\ control)]}$$ (22)

Using the data from steps 3 and 5 and equations (21), (22) and (8) above, but without rounding:

$$HR = \left[\frac{24.00/350.41}{33.19/331.90}\right] = 0.68$$

$$V = \frac{1}{[1/24.0 - 1/350.41 + 1/33.19 - 1/331.90]} = 15.17$$

$$O - E = \ln(0.68) \times 15.17 = -5.74$$

Gives estimates of the HR, *V* and *O-E* as 0.68, 15.17 and -5.74, respectively for the 15–18 month time interval. Note that if censoring had not been taken into account, the estimate of the HR for this time interval would still have been 0.68, but the *V* would be slightly greater at 15.52.

These steps are repeated for all time intervals.

### Step 6, combining all time intervals
The final step is to calculate the overall HR for the trial using the formula for calculating a pooled HR shown previously (1). Taking all time intervals and accounting for censoring a pooled HR of 0.88 and *V* of 128.81 (95%CI of 0.74–1.05) is obtained:

$$HR = \exp\left[\frac{\sum O - E}{\sum V}\right]$$

$$= \exp\left[\frac{[(0.00) + (0.00) + (-5.21) + (-3.25) + (-0.51) + (-5.74)...etc.]}{[(7.55) + (12.86) + (18.10) + (22.96) + (13.05) + (15.17)...etc.]}\right]$$

$$= \exp\left[\frac{-16.35}{128.81}\right]$$

$$HR = 0.88$$

In this example, if the censoring model had not been applied the same HR, a smaller, but similar *V* (136.23) and a similar CI (0.74–1.04) would have been estimated. This is probably because it is a large trial with good follow-up, making both estimates fairly precise. In contrast, the ovarian cancer trial [5] accrued far fewer patients and had poorer follow-up. Using the curve method and accounting for censoring, gives a HR estimate of 1.21 (95% CI 0.62–2.36), but discounting censoring, the HR is slightly more extreme (1.26), with overly precise confidence intervals (95% CI 0.69–2.28). I n other situations the differences may be more pronounced.

### 11. Report presents Kaplan-Meier curve and the numbers at risk
The presentation of the numbers at risk at particular time points with a Kaplan-Meier curve, offers a more direct means of assessing the level of censoring [2], which is taken into account when the HR, *V* and *O-E* are estimated. However, this necessarily limits the division of the curve to these time points, which may be relatively few. Further this approach may be problematic when the event rate between time points is large, e.g. greater than 20% [1].

The number of patients event-free at each time point i.e. the numbers of patients event-free at the start and end of the each time interval is known, and so they do not need to be estimated. For each time interval for each arm, assuming that the level of censoring is constant within each interval, it remains to calculate the number of patients who were: 1) at risk during the interval and 2) the number of events during the interval. These can be used to 4) estimate the *O-E*, *V* and HR for the time interval and the data from all the intervals can be combined in 5) to obtain the *O-E*, *V* and HR for the complete curve. Although not required to estimate the HR, the number of patients who were 3) censored during the interval can also be calculated and is useful for comparison with the other curve method.

The bladder cancer trial report gave the numbers at risk annually until 5 years. These data, and the percentage survival (i.e. event-free) for each arm at the start of each time interval, are given in Table 2 and can be used to illustrate the steps of the method for the 0–12 month time period:

### Step 1. Numbers at risk during the current interval
The same data can be used to quantify the numbers of patients at risk during an interval:

$$At\ risk\ during\ current\ interval = \frac{(At\ risk\ at\ start + At\ risk\ at\ end) \times \%\ Event\ free\ at\ start}{(\%\ Event\ free\ at\ start + \%\ Event\ free\ at\ end)}$$

(23)

For the 0–12 month interval:

$$At\ risk\ during\ (0-12\ month),\ research = \frac{(491+372) \times 100}{(100+78)} = 484.83$$

$$At\ risk\ during\ (0-12\ month),\ control = \frac{(485+355) \times 100}{(100+75)} = 480.00$$

### Step 2. Number of events during the current interval
Again, the same published data can be used to estimate the number of events in an interval:

$$Events\ in\ current\ interval = \frac{(At\ risk\ at\ start + At\ risk\ at\ end)(\%\ Event\ free\ at\ start - \%\ Event\ free\ at\ end)}{(\%\ Event\ free\ at\ start + \%\ Event\ free\ at\ end)}$$

(24)

For the 0–12 month interval:

$$Events\ during\ (0-12\ month),\ research = \frac{(491+372)(100-78)}{(100+78)} = 106.67$$

$$Events\ during\ (0-12\ month),\ research = \frac{(485+355)(100-75)}{(100+75)} = 120.00$$

There were approximately 106 events estimated on the research arm and 120 on the control arm.

*Step 3. Numbers censored during the current interval*
The numbers censored are obtained from the reported numbers at risk and the event rate at the start and end of an interval:

$$\textit{Censored during current interval} = \frac{2 \times (\textit{At risk at start} \times \% \textit{ Event free at end} - \textit{At risk at end} \times \% \textit{ Event free at start})}{(\% \textit{ Event free at start} + \% \textit{ Event free at end})}$$

(25)

Using event rates extracted from the curve at 0 and 12 months and the associated numbers at risk:

$$\textit{Censored during } (0-12 \textit{ month}), \textit{ research} = \frac{2 \times (491 \times 78 - 372 \times 100)}{(100 + 78)} = 12.33$$

$$\textit{Censored during } (0-12 \textit{ month}), \textit{ control} = \frac{2 \times (485 \times 75 - 385 \times 100)}{(100 + 75)} = 10.00$$

approximately 12 and 10 patients were estimated to be censored on the research and control arm respectively. Note that in section 10, by estimating the minimum follow-up to be 14 months and using the censoring model, we failed to take accurate account of censoring in the 0–12 month period.

*Step 4a. Estimate the HR and V for the current interval using the number of events and the numbers at risk during the current interval*
The results from steps 1 and 2 can then be used to estimate the HR, *V* and *O-E* for the time interval using equations (21), (22) and (8), as in section 10.

*Step 4b. Estimate the O-E and V and HR for the current interval using the numbers of events and the numbers at risk during the current interval*
An alternative method estimates *E* and then *O-E* within in each interval:

$$\textit{Expected events during, research} = (\textit{Events research} + \textit{Events control}) \times \frac{\textit{At risk during, research}}{\textit{At risk during, research} + \textit{At risk during, control}}$$

(26)

Using the data for the 0–12 month interval gives the *E* as:

$$\textit{Expected events during, research} = (106.67 + 120.00) \times \frac{484.83}{484.83 + 484.00} = 113.90$$

And the *O-E*:

$$O - E = 106.67 - 113.90 = -7.23$$

Either equation (12) or (13), described earlier can be use to estimate *V*. However, equation (13) is preferred if the randomisation ratio is not 1:1, or the numbers at risk during intervals are very different, e.g. because there is a big difference in effect between arms of the trial.

$$V = \frac{226.67 \times 484.83 \times 480}{(484.83 + 480)^2} = 56.67$$

Using equation (6) we can estimate a HR of 0.88 for the interval.

$$\text{HR} = \exp\left[\frac{-7.23}{56.67}\right] = 0.88$$

*Step 6, combining all time intervals*
Taking all time intervals and censoring into account and using equation (1) as in section 10, gives a pooled HR of 0.88 and *V* of 119.80 (95%CI of 0.74–1.05).

**Interpreting the hazard ratio (HR)**
Usually a HR calculated for a trial or a meta-analysis is interpreted as the relative risk of an event on the research arm compared to control. However, it can also be translated into an absolute difference in the proportion of patients who are event-free at a particular time point or for particular groups of patient, assuming proportional hazards:

exp [ln(proportion of patients event-free) × HR] - proportion event-free

Alternatively, it can be translated into an absolute difference in the median time event free, assuming exponential distributions, by first calculating the median time event free on the research arm:

$$\frac{\text{Median time event free on control}}{\text{HR}}$$

and then the difference between medians:

Median time event free on research - Median time event free on research

These measures require an estimate of the proportion of patients that are event-free in the control group or subgroup of interest and an estimate of the median time event-free in the control group, respectively. Such data may be obtained from a Kaplan-Meier curve of a representative trial or individual patient data meta-analysis, or even from epidemiological data. Alternatively, it may be possible to use 'typical' values from other literature.

Using the bladder cancer example, the HR of 0.85 and an estimated 2-year survival of 58% for patients on the control arm, gives an absolute improvement:

exp [ln(0.58) × 0.85] - 0.58 = 0.05

in survival of 5% at 2 years, taking it from 58% to 63%.

The median survival on control was estimated to be 37 months and so the median survival on the research arm is:

$$37.0/0.85 = 43.6$$

43.6 months, giving an absolute improvement in median survival:

$$43.6 - 37.0$$

of 6.6 months with the research treatment.

### Calculations spreadsheet

Some of the methods described are computationally more complex than others and performing all the calculations by hand for each and every trial can be laborious, lead to errors and require extra data checking. We have therefore developed spreadsheet in Microsoft Excel that carries out the calculations for all of the methods described. The user enters all the reported summary statistics and the spreadsheet estimates the HR, 95%CI, lnHR, *V*, and *O-E* by all possible methods. The user can also input data extracted from Kaplan-Meier curves and estimate censoring using the minimum and maximum follow-up or the reported numbers at risk, to obtain similar summary statistics. Graphical representations of the input data are produced for comparison with the published curves, to assist with data extraction or to highlight data entry errors. Results from all methods are provided in a single output screen, which facilitates comparison. The main features of the calculations spreadsheet are illustrated in Figure 2 and the spreadsheet itself is freely available to readers (see Additional file 1).

## Discussion

We have presented methods for calculating a HR and/or associated statistics from published time-to-event-analyses [1,2] into a practical, less statistical guide. A corresponding, easy-to-use calculations spreadsheet, to facilitate the computational aspects, is available from the authors. The resulting summary statistics can then be used in the meta-analysis procedures found in statistical and meta-analysis software.

There is a hierarchy in the methods described [1,2]. The direct methods make no assumptions and are preferable, followed by the various indirect methods based on reported statistics. The curve methods are likely to be the least reliable and it is not yet clear which method of adjusting for censoring is most reliable. If both curve methods are possible, the choice between the two may be a pragmatic one, depending on whether the minimum and maximum follow-up are reported or need to be estimated, and how many time points the number at risk are reported for and the event rate between those time points. The development of a hybrid of the two curve methods might optimise use of available data. Also, it is not clear how different schemes for dividing up the Kaplan Meier

curves may impact on the resulting statistics. In fact, further research is required to assess how well all of the methods perform according to variations in, for example, trial size, levels of follow-up or event rates.

Although the methods provide a means of analysing time-to-event outcomes for individual trials, they cannot circumvent the other well-known problems of relying on only published data for systematic reviews and meta-analyses. For example, it may not be possible to include all relevant trials, either because trials are not published or because the trial report does not include the outcome of interest, situations which could lead to publication bias [11-13] or selective outcome reporting bias [14], respectively. Similarly, these methods cannot correct common problems with the original reported analyses, such as the exclusion of patients [15,16], analyses which are not by intention-to-treat [17] or analyses confined to particular patient subgroups, which may also lead to bias [16]. Furthermore, if the time-to-event outcome of interest is a long-term outcome, such as survival, then any HR estimation for an individual trial or meta-analysis will be limited by the extent of follow-up at the time that trials are reported. Such issues are relevant to all trials, systematic reviews and meta-analyses and so they should always be taken into account in interpreting results of these studies. Their relative impact is likely to vary between outcomes, trials, meta-analyses and healthcare areas and some may be addressed by obtaining further or updated information direct from trial investigators.

While the methods described previously [1,2] and elaborated here are not a substitute for the re-analysis IPD from all randomised patients, they offer the most appropriate way of analysing time-to-event outcomes, when IPD is not available or the approach is infeasible. Thus, whenever possible they should be used in preference to using a pooled OR or RR or a series of ORs or RRs at fixed time points. This should improve the quality of the analysis and subsequent interpretation of systematic reviews and meta-analyses that include time-to-event outcomes.

## Competing interests

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

## Authors' contributions

This manuscript is based on workshops demonstrating these methods to systematic reviewers. JT helped develop the workshops and the spreadsheet to carry out the calculations, and drafted the manuscript. LS had the idea for the workshops, helped develop the initial methods paper and workshops and helped draft this manuscript. DG had the idea for the workshops, helped develop them and commented on the manuscript. SB helped test the spread-
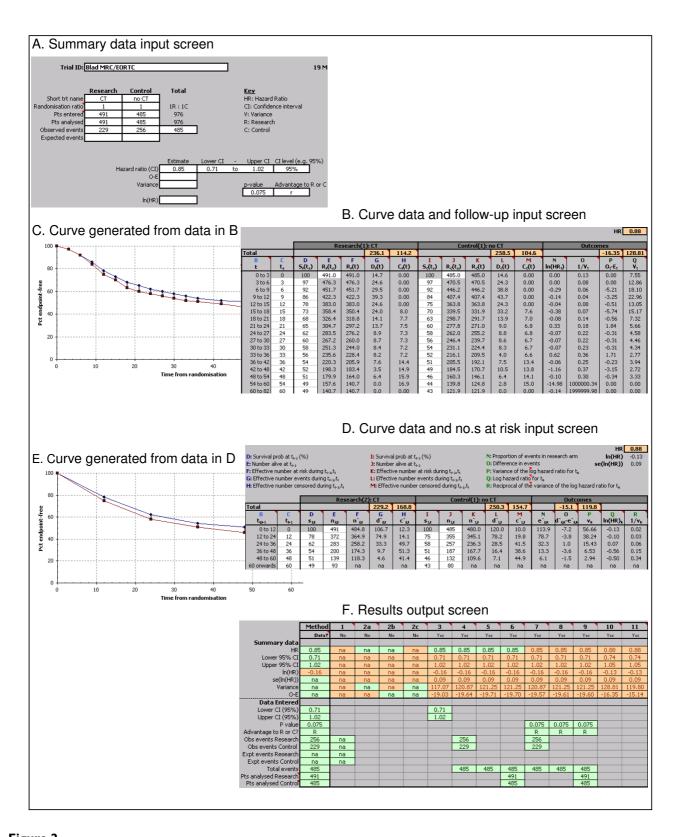
## A. Summary data input screen

Trial ID: Blad MRC/EORTC　　　　　　19 M

|  | Research | Control | Total |
|---|---|---|---|
| Short trt name | CT | no CT |  |
| Randomisation ratio | 1 | 1 | 1R : 1C |
| Pts entered | 491 | 485 | 976 |
| Pts analysed | 491 | 485 | 976 |
| Observed events | 229 | 256 | 485 |
| Expected events |  |  |  |

**Key**
HR: Hazard Ratio
CI: Confidence interval
V: Variance
R: Research
C: Control

|  | Estimate | Lower CI | - | Upper CI | CI level (e.g. 95%) |
|---|---|---|---|---|---|
| Hazard ratio (CI) | 0.85 | 0.71 | to | 1.02 | 95% |
| O-E |  |  |  |  |  |
| Variance |  | p-value | Advantage to R or C |  |  |
|  |  | 0.075 | r |  |  |
| ln(HR) |  |  |  |  |  |

## B. Curve data and follow-up input screen

HR 0.88

| | | | | Research(1): CT | | | | 236.1 | 114.2 | | Control(1): no CT | | | | 258.5 | 104.6 | | Outcomes | | -16.35 | 128.81 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

**Total** values shown: Research 236.1, 114.2; Control 258.5, 104.6; Outcomes -16.35, 128.81

| B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $t$ | $t_k$ | $S_r(t_k)$ | $R_r(t_k)$ | $R_r(t)$ | $D_r(t)$ | $C_r(t)$ | $S_c(t_k)$ | $R_c(t_k)$ | $R_c(t)$ | $D_c(t)$ | $C_c(t)$ | $\ln(HR_k)$ | $1/V_t$ | $O_t$-$E_t$ | $V_t$ |
| 0 to 3 | 0 | 100 | 491.0 | 491.0 | 14.7 | 0.00 | 100 | 485.0 | 485.0 | 14.6 | 0.00 | 0.00 | 0.13 | 0.00 | 7.55 |
| 3 to 6 | 3 | 97 | 476.3 | 476.3 | 24.6 | 0.00 | 97 | 470.5 | 470.5 | 24.3 | 0.00 | 0.00 | 0.08 | 0.00 | 12.86 |
| 6 to 9 | 6 | 92 | 451.7 | 451.7 | 29.5 | 0.00 | 92 | 446.2 | 446.2 | 38.8 | 0.00 | -0.29 | 0.06 | -5.21 | 18.10 |
| 9 to 12 | 9 | 86 | 422.3 | 422.3 | 39.3 | 0.00 | 84 | 407.4 | 407.4 | 43.7 | 0.00 | -0.14 | 0.04 | -3.25 | 22.96 |
| 12 to 15 | 12 | 78 | 383.0 | 383.0 | 24.6 | 0.00 | 75 | 363.8 | 363.8 | 24.3 | 0.00 | -0.04 | 0.08 | -0.51 | 13.05 |
| 15 to 18 | 15 | 73 | 358.4 | 350.4 | 24.0 | 8.0 | 70 | 339.5 | 331.9 | 33.2 | 7.6 | -0.38 | 0.07 | -5.74 | 15.17 |
| 18 to 21 | 18 | 68 | 326.4 | 318.8 | 14.1 | 7.7 | 63 | 298.7 | 291.7 | 13.9 | 7.0 | -0.08 | 0.14 | -0.56 | 7.32 |
| 21 to 24 | 21 | 65 | 304.7 | 297.2 | 13.7 | 7.5 | 60 | 277.8 | 271.0 | 9.0 | 6.8 | 0.33 | 0.18 | 1.84 | 5.66 |
| 24 to 27 | 24 | 62 | 283.5 | 276.2 | 8.9 | 7.3 | 58 | 262.0 | 255.2 | 8.8 | 6.8 | -0.07 | 0.22 | -0.31 | 4.58 |
| 27 to 30 | 27 | 60 | 267.2 | 260.0 | 8.7 | 7.3 | 56 | 246.4 | 239.7 | 8.6 | 6.7 | -0.07 | 0.22 | -0.31 | 4.46 |
| 30 to 33 | 30 | 58 | 251.3 | 244.0 | 8.4 | 7.2 | 54 | 231.1 | 224.4 | 8.3 | 6.7 | -0.07 | 0.23 | -0.31 | 4.34 |
| 33 to 36 | 33 | 56 | 235.6 | 228.4 | 8.2 | 7.2 | 52 | 216.1 | 209.5 | 4.0 | 6.6 | 0.62 | 0.36 | 1.71 | 2.77 |
| 36 to 42 | 36 | 54 | 220.3 | 205.9 | 7.6 | 14.4 | 51 | 205.5 | 192.1 | 7.5 | 13.4 | -0.06 | 0.25 | -0.23 | 3.94 |
| 42 to 48 | 42 | 52 | 198.3 | 183.4 | 3.5 | 14.9 | 49 | 184.5 | 170.7 | 10.5 | 13.8 | -1.16 | 0.37 | -3.15 | 2.72 |
| 48 to 54 | 48 | 51 | 179.9 | 164.0 | 6.4 | 15.9 | 46 | 160.3 | 146.1 | 6.4 | 14.1 | -0.10 | 0.30 | -0.34 | 3.33 |
| 54 to 60 | 54 | 49 | 157.6 | 140.7 | 0.0 | 16.9 | 44 | 139.8 | 124.8 | 2.8 | 15.0 | -14.98 | 1000000.34 | 0.00 | 0.00 |
| 60 to 82 | 60 | 49 | 140.7 | 140.7 | 0.0 | 0.00 | 43 | 121.9 | 121.9 | 0.0 | 0.00 | -0.14 | 1999999.98 | 0.00 | 0.00 |

## C. Curve generated from data in B

(Plot: Pct endpoint-free vs Time from randomisation)

## D. Curve data and no.s at risk input screen

HR 0.88　ln(HR) -0.13　se(ln(HR)) 0.09

D: Survival prob at $t_{l-1}$ (%)
E: Number alive at $t_{l-1}$
F: Effective number at risk during $t_{l-1}, t_l$
G: Effective number events during $t_{l-1}, t_l$
H: Effective number censored during $t_{l-1}, t_l$

I: Survival prob at $t_{l-1}$ (%)
J: Number alive at $t_{l-1}$
K: Effective number at risk during $t_{l-1}, t_l$
L: Effective number events during $t_{l-1}, t_l$
M: Effective number censored during $t_{l-1}, t_l$

N: Proportion of events in research arm
O: Difference in events
P: Variance of the log hazard ratio for $t_k$
Q: Log hazard ratio for $t_k$
R: Reciprocal of the variance of the log hazard ratio for $t_k$

| | | | | Research(2): CT | | | | 229.2 | 168.8 | | Control(1): no CT | | | | 250.3 | 154.7 | | Outcomes | | -15.1 | 119.8 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

| B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $t_{l-1}$ | $t_{l-1}$ | $s_{2,l}$ | $n_{2,l}$ | $n'_{2,l}$ | $d'_{2,l}$ | $c'_{2,l}$ | $s_{1,l}$ | $n_{1,l}$ | $n'_{1,l}$ | $d'_{1,l}$ | $c'_{1,l}$ | $e'_{2,k}$ | $d'_{2,k}$-$e'_{2,k}$ | $v_k$ | $\ln(HR)_k$ | $1/v_k$ |
| 0 to 12 | 0 | 100 | 491 | 484.8 | 106.7 | 12.3 | 100 | 485 | 480.0 | 120.0 | 10.0 | 113.9 | -7.2 | 56.66 | -0.13 | 0.02 |
| 12 to 24 | 12 | 78 | 372 | 364.9 | 74.9 | 14.1 | 75 | 355 | 345.1 | 78.2 | 19.8 | 78.7 | -3.8 | 38.24 | -0.10 | 0.03 |
| 24 to 36 | 24 | 62 | 283 | 258.2 | 33.3 | 49.7 | 58 | 257 | 236.3 | 28.5 | 41.5 | 32.3 | 1.0 | 15.43 | 0.07 | 0.06 |
| 36 to 48 | 36 | 54 | 200 | 174.3 | 9.7 | 51.3 | 51 | 187 | 167.7 | 16.4 | 38.6 | 13.3 | -3.6 | 6.53 | -0.56 | 0.15 |
| 48 to 60 | 48 | 51 | 139 | 118.3 | 4.6 | 41.4 | 46 | 132 | 109.6 | 7.1 | 44.9 | 6.1 | -1.5 | 2.94 | -0.50 | 0.34 |
| 60 onwards | 60 | 49 | 93 | na | na | na | 43 | 80 | na | na | na | na | na | na | na | na |

## E. Curve generated from data in D

(Plot: Pct endpoint-free vs Time from randomisation)

## F. Results output screen

| | Method | 1 | 2a | 2b | 2c | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Summary data** | Data? | No | No | No | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| HR | 0.85 | na | na | na | na | 0.85 | 0.85 | 0.85 | 0.85 | 0.85 | 0.85 | 0.85 | 0.88 | 0.88 |
| Lower 95% CI | 0.71 | na | na | na | na | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 | 0.71 | 0.74 | 0.74 |
| Upper 95% CI | 1.02 | na | na | na | na | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.05 | 1.05 |
| ln(HR) | -0.16 | na | na | na | na | -0.16 | -0.16 | -0.16 | -0.16 | -0.16 | -0.16 | -0.16 | -0.13 | -0.13 |
| se(ln(HR)) | na | na | na | na | na | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 |
| Variance | na | na | na | na | na | 117.07 | 120.87 | 121.25 | 121.25 | 120.87 | 121.25 | 121.25 | 128.81 | 119.80 |
| O-E | na | na | na | na | na | -19.03 | -19.64 | -19.71 | -19.70 | -19.57 | -19.61 | -19.60 | -16.35 | -15.14 |
| **Data Entered** | | | | | | | | | | | | | | |
| Lower CI (95%) | 0.71 | | | | | 0.71 | | | | | | | | |
| Upper CI (95%) | 1.02 | | | | | 1.02 | | | | | | | | |
| P value | 0.075 | | | | | | | | | 0.075 | 0.075 | 0.075 | | |
| Advantage to R or C? | R | | | | | | | | | R | R | R | | |
| Obs events Research | 256 | na | | | | 256 | | | | 256 | | | | |
| Obs events Control | 229 | na | | | | 229 | | | | 229 | | | | |
| Expt events Research | na | na | | | | | | | | | | | | |
| Expt events Control | na | na | | | | | | | | | | | | |
| Total events | 485 | | | | | 485 | 485 | 485 | 485 | 485 | 485 | 485 | | |
| Pts analysed Research | 491 | | | | | | | | | 491 | | | 491 | |
| Pts analysed Control | 485 | | | | | | | | | 485 | | | 485 | |

**Figure 2**
Data input screens (A, B and D), generated curves (C and E) and output screen (F) from the calculations spreadsheet.

sheet and run the workshops and commented on the manuscript. MS developed the spreadsheet and commented on the manuscript. All authors read and approved the final manuscript.

## Appendix 1: Previously published formulae for generating hazard ratios from published time-to-event data [1,2]. The number in brackets link these to their descriptive equivalent in the text

*1. Generating the O-E, V, HR and lnHR from reported summary statistics*

For equations 1–16 and following the notation of Parmar *et al.* [1], for trial *i*:

$O_{ri}$ = observed number of events in the research group

$E_{ri}$ = logrank expected events in the research group

$O_{ci}$ = observed number of events in the control group

$E_{ci}$ = logrank expected events in the control group

$O_r$ - $E_r$ observed minus expected events in the research group

$O_i$ = total observed events ($O_{ri} + O_{ci}$)

$V_{ri}$ = logrank variance

$\ln(HR_i)$ = log HR

var[$\ln(HR_i)$ = variance of the log hazard ratio

UPPCI$_i$ = Value for the upper end of the confidence interval

LOWCI$_i$ = Value for the lower end of the confidence interval

$\Phi^{-1}(1-\alpha_i/2)$ = z score for the upper end of the confidence intreval

$R_{ri}$ = number randomised to the research group

$R_{ci}$ = number randomised to the control group

$p_i$= reported two-sided p-value associated with the logrank or Mantel-Haenszel test (or Cox model)

*Estimating a pooled lnHR from a series of trials*

Estimating a pooled lnHR using the inverse variance method:

$$\ln(HR) = \frac{\sum_{i=1}^{K} \dfrac{\ln(HR_i)}{var[\ln(HR_i)]}}{\sum_{i=1}^{K} \dfrac{1}{var[\ln(HR_i)]}} \qquad (AP2)$$

*Estimating the O-E, V, HR and lnHR from reported summary statistics*

The reciprocal nature of the variance of the lnHR and the logrank variance:

$$var(\ln(HR_i)) = \frac{1}{V_{ri}} \qquad (AP3)$$

$$V_{ri} = \frac{1}{var(\ln(HR_i))} \qquad (AP4)$$

Directly estimating the lnHR and associated variance using the formal definition:

$$\ln(HR_i) = \ln\left[\frac{O_{ri}/E_{ri}}{O_{ci}/E_{ci}}\right] \qquad (AP5)$$

$$V_{ri} = \frac{1}{[(1/E_{ri})+(1/E_{ci})]} \qquad (AP6)$$

Direct estimation of the lnHR using the alternative definition:

$$\ln(HR_i) = \left[\frac{O_{ri}-E_{ri}}{V_{ri}}\right] \qquad (AP7)$$

Indirect estimation of the variance of the lnHR from the confidence interval:

$$var(\ln HR_i) = \left[\frac{UPPCI_i - LOWCI_i}{2\Phi^{-1}(1-\alpha_i/2)}\right]^2 \qquad (AP10)$$

Indirect estimation of the variance of the lnHR from the number of events:

$$V_{ri} = O_{ri}O_{ci}/O_i \qquad (AP11)$$

$$V_{ri} = O_i/4 \qquad (AP12)$$

Indirect estimation of the variance of the lnHR from the number of events and the numbers randomised (analysed) on each arm:

$$V_{ri} = \frac{O_i R_{ri} R_{ci}}{(R_{ri} + R_{ci})^2} \qquad (AP13)$$

Indirect estimation of the observed minus expected events from the observed events and the p-value:

$$(O_{ri} - E_{ri}) = \sqrt{\frac{O_{ri} O_{ci}}{O_i}} \times \Phi^{-1}(1 - \frac{p_i}{2}) \qquad (AP14)$$

$$(O_{ri} - E_{ri}) = 1/2 \times \sqrt{O_i} \times \Phi^{-1}(1 - \frac{p_i}{2}) \qquad (AP15)$$

Indirect estimation of the observed minus expected events from the observed events, the p-value and the numbers randomised (analysed) on each arm:

$$(O_{ri} - E_{ri}) = \frac{\sqrt{(O_i R_{ri} R_{ci})}}{(R_{ri} + R_{ci})} \times \Phi^{-1}(1 - \frac{p_i}{2}) \qquad (AP16)$$

*2. Generating the HR and V from published Kaplan-Meier curves and follow-up*

For equations 17–22, and following the notation of Parmar *et al.* [1], for trial *i* and *T* non-overlapping time points ($t = 1, ...,T$) :

$t$ = whole time interval ($t - 1, t$)

$t_s$ = start of the time interval ($t - 1, t$)

$t_e$ = end of the time interval ($t - 1, t$)

$R_{ri}(t)$ = effective number of patients at risk on the research arm during time interval ($t - 1, t$)

$R_{ri}(t - 1)$ = effective number of patients at risk on the research arm during time interval ($t - 2, t - 1$)

$D_{ri}(t)$ = effective number of events on the research arm during time interval ($t - 1, t$)

$D_{ci}(t)$ = effective number of events on the control arm during time interval ($t - 1, t$)

$D_{ri}(t - 1)$ = effective number of events on the research arm during time interval ($t - 2, t - 1$)

$C_{ri}(t)$ = effective number of patients censored on the research arm during time interval ($t - 1, t$)

$C_{ci}(t)$ = effective number of patients censored on the control arm during time interval ($t - 1, t$)

$C_{ri}(t - 1)$ = effective number of patients censored on the research arm during time interval ($t - 2, t - 1$)

$S_{ri}(t_s)$ = event-free probability on the research arm at the start of time interval ($t - 1, t$)

$S_{ri}(t_e)$ = event-free probability on the research arm at the end of time interval ($t - 1, t$)

$F_{min}$ = minimum follow-up

$F_{max}$ = maximum follow-up

Estimation of the numbers event-free at the start of a time interval:

$$R_{ri}(t_s) = R_{ri}(t - 1) - D_{ri}(t - 1) - C_{ri}(t - 1) \qquad (AP17)$$

Estimation of the numbers censored during a time interval

if $t_s \geq F_{min}$ and $F_{min} \leq t_e \leq F_{max}$

$$C_{ri}(t) = R_{ri}(t_s) \left\{ \frac{1}{2} \frac{(t_e - t_s)}{(F_{max} - t_s)} \right\} \text{ (assuming censoring at constant rate)}$$
$$(AP18)$$

Estimation of the numbers at risk during a time interval, adjusted for censoring

$$R_{ri}(t) = R_{ri}(t_s) - C_{ri}(t) \qquad (AP19)$$

Estimation of the number of events during a time interval

$$D_{ri}(t) = \left[ R_{ri}(t) \times \left( \frac{S_{ri}(t_s) - S_{ri}(t_e)}{S_{ri}(t_s)} \right) \right] \qquad (AP20)$$

Note that equations 17–20 are also are used for the control arm.

Estimation of the HR and *V* for a time interval from a Kaplan-Meier curve

$$\ln(HR_i(t)) = \ln\left( \frac{D_{ri}(t)/R_{ri}(t)}{D_{ci}(t)/R_{ci}(t)} \right) \qquad (AP21)$$

$$\text{var}[\ln(HR_i(t))] = \frac{1}{D_{ri}(t)} - \frac{1}{R_{ri}(t)} + \frac{1}{D_{ci}(t)} - \frac{1}{R_{ci}(t)}$$
$$(AP22)$$

*3. Generating the HR and V from published Kaplan-Meier curves and the numbers at risk*

For equations 23–26, and following the notation of [2], for time interval *i*:

*j* = treatment group (where 1 = the control arm and 2= the research arm)

$t_{i-1}$ = time at the start of the current interval

$t_{i-1}$= time at the start of the prior interval

$n_{j,i}$ = number at risk at end of interval $[t_{i-1}, t_i)$ in group *j*

$n_{j,i-1}$ = number at risk at start of interval $[t_{i-1}, t_i)$ in group *j*

$n^*_{j,i}$ = number at risk during interval $[t_{i-1}, t_i)$ in group *j*

$d^*_{j,i}$ = number of events during interval $[t_{i-1}, t_i)$ in group *j*

$c^*_{j,i}$ = number censored during interval $[t_{i-1}, t_i)$ in group *j*

$s^*_{j,i}$ = event-free probability at end of interval $[t_{i-1}, t_i)$ in group *j*

$s^*_{j,i-1}$ = event-free probability at start of interval $[t_{i-1}, t_i)$ in group *j*

$e^*_{j,i}$ = logrank expected events during interval $[t_{i-1}, t_i)$ in group j = 2 (the research arm)

Estimation of the numbers at risk during a time interval from a Kaplan-Meier curve

$$n^*_{j,i} = \frac{(n_{j,i-1} + n_{j,i})s^*_{j,i-1}}{(s^*_{j,i-1} + s^*_{j,i})} \qquad \text{(AP23)}$$

Estimation of the number of events during a time interval from a Kaplan-Meier curve

$$d^*_{j,i} = \frac{(n_{j,i-1} + n_{j,i})(s^*_{j,i-1} - s^*_{j,i})}{(s^*_{j,i-1} + s^*_{j,i})} \qquad \text{(AP24)}$$

Estimation of the numbers censored during a time interval from a Kaplan-Meier curve

$$c^*_{j,i} = \frac{2(n_{j,i-1}s^*_{j,i} - n_{j,i}s^*_{j,i-1})}{(s^*_{j,i-1} + s^*_{j,i})} \qquad \text{(AP25)}$$

Estimation of the number of logrank expected events during a time interval from a Kaplan-Meier curve

$$e^*_{2,i} = (d^*_{2,i} + d^*_{1,i}) * \frac{n^*_{2,i}}{n^*_{2,i} + n^*_{1,i}} \qquad \text{(AP26)}$$

## Appendix 2: Estimating or educated 'guesstimating' minimum and maximum follow-up

When the minimum and maximum follow-up are not explicitly reported, it may be possible to estimate them for a particular trial, provided that some indicators of extent of follow-up are provided. In descending order of preference, the following are some strategies that we have employed to estimate the minimum and maximum follow-up:

### For minimum follow-up, if the trial report presents
1. Censoring tick marks on Kaplan-Meier curve

➢ Assume first tick mark indicates the point of minimum follow-up

2. Median follow-up and accrual period

➢ Assume minimum follow-up = median follow-up minus half the accrual period

3. Date of analysis and accrual period, could assume

➢ Assume minimum follow-up = date of analysis minus final date of accrual

4. Date of submission and accrual period

➢ Assume estimated date of analysis = date of submission minus 6 months

➢ Assume minimum follow-up = estimated date of analysis minus final date of accrual

### For maximum follow-up, if the trial report presents
1. Censoring tick marks on Kaplan-Meier curve

➢ Assume last tick mark indicates the point of maximum follow-up

2. Median follow-up and accrual period

➢ Assume maximum follow-up = median follow-up plus half the accrual period

3. Date of analysis and accrual period, could assume

➢ Assume maximum follow-up = date of analysis minus first date of accrual

4. Date of submission and accrual period, could assume

➢ Assume estimated date of analysis = date of submission minus 6 months

Assume maximum follow-up = estimated date of analysis minus first date of accrual

## Additional material

### Additional file 1

*HR calculations spreadsheet. Spreadsheet to facilitate the estimation of hazard ratios from published summary statistics or data extracted from Kaplan-Meier curves.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1745-6215-8-16-S1.xls]

## Acknowledgements

## References

1.  Parmar MKB, Torri V, Stewart L: **Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints.** *Statistics in Medicine* 1998, **17:**2815-34.
2.  Williamson PR, Tudur Smith C, Hutton JL, Marson AG: **Aggregate data meta-analysis with time-to-event outcomes.** *Statistics in Medicine* 2002, **21:**3337-51.
3.  Altman DG, De Stavola BL, Love SB, Stepniewska KA: **Review of survival analyses published in cancer.** *British Journal of Cancer* 1995, **72:**511-8.
4.  Pocock SJ, Clayton TC, Altman DG: **Survival plots of time-to-event ouctomes in clinical trials.** *Lancet* 2002, **359:**1686-9.
5.  Mangioni C, Bolis G, Pecorelli S, Bragman K, Epis A, Favalli G, Gambino A, Landoni F, Presti M, Torri W, Vassena L, Zanaboni F, Marsoni S: **Randomized trial in advanced ovarian cancer comparing cisplatin and carboplatin.** *Journal of the National Cancer Institute* 1989, **81:**1461-71.
6.  International Collaboration of Trialists on behalf of the Medical Research Council Advanced Bladder Cancer Working Party, EORTC Genito-urinary Group Australian Bladder Cancer Study Group, National Cancer Institute of Canada Clinical Trials Group, Finnbladder, Norwegian Bladder Cancer Study Group and Club Urologico Espanol de Tratamiento Oncologico (CUETO) group: **Neoadjuvant cisplatin, methotrexate, and vinblastine chemotherapy for muscle-invasive bladder cancer: a randomised controlled trial.** *Lancet* 1999, **354:**533-40.
7.  Yusuf S, Peto R, Lewis JA, Collins R, Sleight P: **Beta blockade during and after myocardial infarction: an overview of the randomized trials.** *Progress in Cardiovascular Diseases* 1985, **27:**335-71.
8.  DerSimonian R, Laird N: **Meta-analysis in clinical trials.** *Controlled Clinical Trials* 1986, **7:**177-88.
9.  Tudur C, Williamson PR, Khan S, Best L: **The value of the aggregate data approach in meta-analysis with time-to-event outcomes.** *Journal of the Royal Statistical Society A* 2001, **164:**357-70.
10. Tierney JF, Burdett S, Stewart LA: **Feasibility and reliability of using hazard ratios in meta-analyses of published time-to-event data.** *preparation* .
11. Dickersin K: **The existence of publication bias and risk factors for its occurrence.** *Journal of the American Medical Association* 1990, **263:**1385-9.
12. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR: **Publication bias in clinical research.** *Lancet* 1991, **337:**867-72.
13. Dickersin K, Min Y-I, Meinert CL: **Factors influencing publication of research results.** *Journal of the American Medical Association* 1992, **267:**374-8.
14. Chan A-W, Hróbjartsson A, Haarh MT, Gøtzche PC, Altman DG: **Empirical evidence for selective reporting of outcomes in randomized trials.** *Journal of the American Medical Association* 2004, **291:**2457-65.
15. Schulz KF, Grimes DA, Altman DG, Hayes RJ: **Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology.** *BMJ* 1996, **312:**742-4.
16. Tierney JF, Stewart LA: **Investigating patient exclusion bias in meta-analysis.** *International Journal of Epidemiology* 2005, **34(1):**79-87.
17. Hollis S, Campbell F: **What is meant by intention-to-treat analysis? Survey of published randomised controlled trials.** *BMJ* 1999, **319:**670-4.