**Tip for data extraction for meta-analysis - 14**



*Estimating a hazard ratio from time-to-event data*

Kathy Taylor

In this blog post I’m returning to extracting hazard ratios (HRs), but this time it’s about estimating HRs from time-to-event data (survival data). [Guyot et al](https://www.ncbi.nlm.nih.gov/pubmed/22297116) use image extraction software to extract the co-ordinates of Kaplan-Meier (K-M) curves, also known as survival curves. They apply an algorithm to reconstruct individual patient data, which they then re-analyse to estimate the HR. Guyot et al highlight other approaches, which use fewer data points from K-M curves including the methods of [Parmar et al](https://www.ncbi.nlm.nih.gov/pubmed/9921604) and [Williamson et al](https://www.ncbi.nlm.nih.gov/pubmed/12407676). [Tierney et al](https://www.ncbi.nlm.nih.gov/pubmed/17555582) revisit these methods and make them more accessible, by providing simpler notation, step-by-step instructions, equations, worked examples from a couple of published trials, and a very useful spreadsheet that does all the calculations. Note that there’s an updated paper in the pipeline.

I’m going to go through the paper by Tierney et al, add a bit more explanation, derive the equations, and run through data from a different trial. Note that, for consistency, I use the terms “survival” and “at risk” where Tierney et al use the term “event-free”.

Tierney et al start by highlighting the summary statistics that are required, for each trial, for meta-analysis:

HR hazard ratio

lnHR natural [logarithm](https://bit.ly/2ueyQem) of the hazard ratio

O-E difference between the observed and expected number of events in the intervention group

V variance of O-E

or

V\* variance of lnHR

Note that:

V and V\* are the reciprocal of each other i.e. V=1/V\* and V\*=1/V\*

O-E and V are also called the logrank O-E and logrank variance

They then show how to extract the above summary statistics when the following data are reported:

1. O and E or hazard rates for the intervention and control groups
2. O-E and V for the intervention group
3. HR and confidence intervals
4. HR and events in each treatment arm and a randomisation ratio of 1:1
5. HR and total events and a randomisation ratio of 1:1
6. HR, total events and the numbers randomised in each arm
7. p-value and events in each treatment arm and a randomisation ratio of 1:1
8. p-value and total events and a randomisation ratio of 1:1
9. p-value and total events and the number randomised to each arm
10. Kaplan-Meier curves
	1. Reported with information about follow-up
	2. Reported with numbers at risk

Their spreadsheet can be used for all of the above, although the underlying equations for 1 to 9 are straightforward. For 10a and 10b, the equations are more complicated and the inputs required for the spreadsheet include extracted curve data, and in order to estimate the numbers censored, either the reported maximum and minimum followup times (if these are not reported, Tierney et al offer advice on how these data may be estimated), or the reported numbers at risk. We say that a patient is censored if they leave the study before they’ve experienced the event of interest.

For 10a, the survival curve needs to be divided upon into a number of time intervals and the times and survival proportions extracted. These intervals should be chosen to give a good representation of the event rates over time, so when the event rate is high, you need to use closer intervals, and when the event rate is low, you can space out the intervals. You should also ensure that the minimum followup lies at the end of an interval (I’ll explain why in the next blog post). For 10b, only the survival proportions at the times of the reported numbers at risk need to be extracted.

I’m going to illustrate the use of the spreadsheet by working through an example based on the [FLOT4 trial](https://www.ncbi.nlm.nih.gov/pubmed/30982686). This was a trial of two different peri-operative chemotherapy regimes – fluorouracil plus leucovorin, oxaliplatin and docetax (FLOT group) and epirubicin, cisplatin, fluorouracil or capecitabine (ECF/ECX comparator group) in patients with gastric or gastro-oesophageal cancer.

The reported HR for overall survival is 0.77 (95% CI 0.63 to 0.94) and here are the K-M curves.:



Here are the extracted data (which I extracted using the software that I demonstrated in my [video post](https://bit.ly/2SLk0HF)) tabulated with the reported numbers of patients at risk:

Table. Data for the FLOT4 trial

|  |  |  |
| --- | --- | --- |
| **Time at start of interval (months)** | **Survival (event-free) %** | **Reported numbers at risk** |
| **FLOT** | **ECF/ECX** | **FLOT** | **ECF/ECX** |
| 0 | 100 | 100 | 356 | 360 |
| 2 | 99 | 99 |  |  |
| 4 | 98 | 97 |  |  |
| 6 | 93 | 91 |  |  |
| 8 | 91 | 90 |  |  |
| 10 | 87 | 83 |  |  |
| 12 | 84 | 80 | 297 | 287 |
| 14 | 80 | 75 |  |  |
| 16 | 78 | 73 |  |  |
| 18 | 76 | 69 |  |  |
| 21 | 72 | 63 |  |  |
| 24 | 69 | 58 | 231 | 202 |
| 27 | 65 | 55 |  |  |
| 30 | 61 | 54 |  |  |
| 33 | 60 | 51 |  |  |
| 36 | 57 | 49 | 140 | 126 |
| 39 | 55 | 47 |  |  |
| 42 | 54 | 46 |  |  |
| 45 | 53 | 45 |  |  |
| 48 | 50 | 44 | 87 | 83 |
| 54 | 49 | 40 |  |  |
| 60 | 45 | 36 | 39 | 33 |
| 66 | 43 | 35 |  |  |
| 72 | 43 | 32 | 5 | 9 |

The 1st worksheet of the spreadsheet calculator (Figure 1) is the summary input data screen. This shows the time-to-event data that was reported (in the white cells) for the FLOT4 trial.



Figure 1. Summary input screen

The 2nd worksheet (Figure 2) shows the extracted curve and followup data. The followup data was not reported and I estimated the minimum follow-up to be 15 months and the maximum follow-up to be 80 months. Note that using the data-extraction software that I demonstrated [previously](https://bit.ly/2SLk0HF) produces numbers to many decimal places but the times need to be inputted as integers. I also entered the survival curves as integers so that the calculated numbers in my worked examples in the next two posts match exactly the calculated numbers in the spreadsheet.

The figure in the right hand corner gives the estimated HR as 0.78 (the reported HR is 0.77). The accuracy of the calculated HR is pretty good but it could be improved by making the intervals smaller and extracting more data points.



Figure 2. Curve and followup data

The spreadsheet plots the extracted data in the next worksheet (Figure 3).



Figure 3. Plotted data corresponding with screen shot shown in Figure 2

The 4th worksheet (Figure 4) includes the numbers are risk and corresponding survival fractions.



Figure 4. Curve data and reported numbers at risk

For this case the calculated HR, shown in the upper right hand corner, is 0.79, which with the plotted curve (Figure 5) indicates the lower accuracy with less data.



Figure 5. Plotted data corresponding with screen shot shown in Figure 4

The output screen (Figure 6) provides the estimated HRs with their confidence intervals. The estimated HR using the survival curve and follow-up data is 0.78 (0.64 to 0.96) and the estimated HR using the survival curve and the numbers at risk is 0.79 (0.64 to 0.97). Both these estimates are very close to the actual HR of 0.77 (0.63 to 0.97).



Figure 6. Output screen

In my next two blog posts, I’m going to look more closely at the equations underlying these spreadsheet calculations. I will first deal with the case of estimating a HR from K-M curves reported with follow-up information (10a).

*Here’s a tip…*

*There are equations you can use to convert time-to-event data into a suitable form for meta-analysis and there’s a very useful spreadsheet available to do the calculations.*

**Dr Kathy Taylor teaches data extraction in Meta-analysis,** [**https://www.conted.ox.ac.uk/courses/meta-analysis</link**](https://www.conted.ox.ac.uk/courses/meta-analysis%3C/link)**> This is a short course that is also available as part of our MSc in Evidence-Based Health Care**

[**https://www.conted.ox.ac.uk/about/msc-in-evidence-based-health-care**](https://www.conted.ox.ac.uk/about/msc-in-evidence-based-health-care)**, MSc in Medical Statistics**

[**https://www.conted.ox.ac.uk/about/msc-in-ebhc-medical-statistics**](https://www.conted.ox.ac.uk/about/msc-in-ebhc-medical-statistics)**, and MSc in Systematic Reviews**

[**https://www.conted.ox.ac.uk/about/msc-in-ebhc-systematic-reviews**](https://www.conted.ox.ac.uk/about/msc-in-ebhc-systematic-reviews)

**Follow me on Twitter @dataextips for updates on my blog, related news, and to learn about other examples where people are making statistics more broadly accessible.**