**Solutions**

Advanced Health Economic Modelling

Instructions for computer lab survival analysis

In this computer practical you will use:

* The **four Excel files**, **R project** and an R markdown file provided on CANVAS (Make sure to save everything in one folder for the remainder of the practical. We recommend downloading the entire folder from Canvas and store this on a location place on your computer (not your download folder!)
* **Two scientific papers** (link provided on Canvas):
  + *de Bono et al., 2010* - Prednisone plus cabazitaxel or mitoxantrone for metastatic castration-resistant prostate cancer progressing after docetaxel treatment: a randomised open-label trial. The Lancet, Volume 376, Issue 9747, 1147 - 1154
  + *Bahl et al., 2013* - Impact of cabazitaxel on 2-year survival and palliation of tumour-related pain in men with metastatic castration-resistant prostate cancer treated in the TROPIC trial, Annals of Oncology, Volume 24, Issue 9, 2402-2408.
* **R and R studio** (please make sure you have this software downloaded)
  + University computers: this can be done through the software centre
  + Personal computer: use the instructions on these websites – it is freely available.
    - R: <https://rstudio-education.github.io/hopr/starting.html> (Explains what R and R studio do)
    - R studio: <https://posit.co/download/rstudio-desktop/#download> (download free desktop version)

Start by opening Excel on your computer. Once you have opened Excel, please follow the set-up step below. This fixes one of the most common issues experienced by students in this exercise.

## Excel English vs European number set up.

Windows:

1. In Excel go to File -> Options -> Advanced
2. Untick “Use system separators” and then please ensure that the decimals is ‘ .’ (i.e. a full stop) and thousands is ‘ ,’ (i.e. a comma) as per the English numerical system. It should look like this:

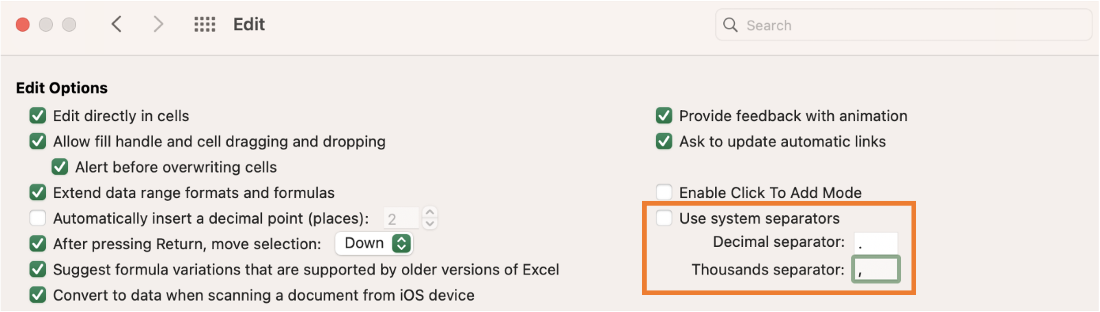
Graphical user interface, text, application

Description automatically generated

1. And press OK. Make sure this gets applied to all Excel files that are required for this practical.

Mac:

1. In Excel got to Excel -> Preferences -> Edit
2. Untick “Use system separators” and then please ensure that the decimals is ‘ .’ (i.e. a full stop) and thousands is ‘ ,’ (i.e. a comma) as per the English numerical system. It should look like this:



Now you are ready to start with the questions!

**Q0:**

**Open both the de Bono et al. (2010) and the Bahl et al. (2013) papers. Compare Figure 2 from de Bono et al. (2010) with Figure 1 from Bahl et al. (2013). What are the main differences and which paper should we use for better overall survival (OS) estimates?**

|  |
| --- |
| **Bahl et al. (2013) study provides OS results after 2 years of follow up. Even though the Kaplan Meier curves look similar, number at risk values after month 12 are higher in this study in comparison to those values in de Bono et al. (2010). We prefer the most updated results, therefore OS results from Bahl et al. (2013) will be used.** |
|  |

In order to estimate transitions of the patients, you would normally need the raw patient-level data on overall survival (OS) and progression free survival (PFS) from the trial, and, if not all patients have progressed or died, then fit parametric survival curves (e.g., exponential, Weibull, log-normal, log-logistic etc.) through them. The type of output you will get from such a patient-level survival analysis is demonstrated in Table 3.2 from the Briggs et al. book[[1]](#footnote-1).

Since we do not have the patient-level data, but only the published KM curves, we need to create pseudo-patient-level data (i.e., create data that will lead to the same KM curve as published) to be able to use parametric survival modelling. For this purpose, we will use the published OS (Overall Survival) curves in the updated Bahl et al 2013 paper[[2]](#footnote-2) and PFS (Progression Free Survival) curves in the original Bono et al 2010 paper[[3]](#footnote-3) and follow the method described in the Hoyle & Henley paper[[4]](#footnote-4).

The relevant X and Y points from the corresponding Kaplan Meier curves are already extracted for you in the relevant Excel files using the [WebPlotDigitizer4.1](https://apps.automeris.io/wpd/) software.

**Q1:**

Open the *XY* sheet in the *Hoyle\_and\_Henley\_Bahl\_OS\_Mito.xls* file. Check if these values are in line with the graph of the OS Kaplan-Meier curve of mitoxantrone from the Bahl 2013 paper. Next, open the “*Number events & censored*” sheet and fill in the white cells in column *B* with these “*X*” values (i.e. 0, 1.5, 3, …)andfill the white cells in column *E* with the “*Y*” values (i.e. 1, 0.98, 0.95, … ). Finally, fill in the white cells in column *H* with the corresponding “*number at risk*” values given under the OS KM curve in Bahl 2013 paper (i.e. 377, 299, 195, …)

**Can you recognize the calculations in columns I, J, K, L, M and N from the Hoyle and Henley 2011 article?**

|  |
| --- |
| **The calculations in these columns give the estimates of number of death and number of censored patients in corresponding time intervals as given in equations 1b, 2b and 3b, assuming constant censoring within each interval.** |
|  |

**Q2:**

Check the numbers under empirical survival probability S(t) (column *E)*. At the point where the value is 0.500, we observe the median survival time (in column B). However, column E never becomes exactly 0.50. So, find the time at which the survival is slightly above 0.5 and the next time where survival is just below 0.5. Take the average of these times and check if this value corresponds to the median of the *OS KM* curve given in the Bahl et al. 2013 paper in Figure 1 and to the reported median in the de Bono 2010 paper.

|  |
| --- |
| **Column E= 12.75 months= (12+13.5)/2**  **The Bono et al. (2012) paper: 12.7 months**  **Ruler & Bahl et al. (2013): Similar around 12.7 months** |

**Q3:**

Now open the “*R data*” sheet. What do these number correspond to? Can you say how many patients died and were censored between time points 0 and 1.5? How about between time points 15 and 16.5?

|  |
| --- |
| **0-1.5 --> 11 died, 3 censored.**  **15-16.5 --> 27 died, 3 censored.** |

**Q4:**

**Follow the steps below:**

* Make sure you watch the instruction video (see Canvas).
* Download all the files needed for this computer practical and save them in the same folder
* Open the R project. (Double click on the files ending with .Rproj)
* A screenshot of a computer

  Description automatically generatedR studio will be opened - check that you see the folder name on the top right
* Open the R markdown file script. (file ending with .Rmd – “*PC\_lab\_R.Rmd*”, right bottom of R studio)
* R markdown scripts combine text with chunks of R code. The R code chunks are in grey. You can run a piece of R code by clicking on the “green play button” on the top right corner of each chunk.

A screen shot of a computer

AI-generated content may be incorrect.

* Note: R will read the data from the Excel files. Therefore, it is important that the file name exactly match with the file names in the script. So please don’t rename the Excel files.
* Now run the chunks of code of Part 1
* This part will extract the survival parameters. You will get output that looks like the following table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *OS for mitoxandrone* | | | | |
|  | exponential | weibull | lognormal | loglogistic |
| AIC |  |  |  |  |
| intercept |  |  |  |  |
| log(scale) |  |  |  |  |

**Please have a look at these AIC values. Which value is the highest? Which values is the lowest? What does the AIC value mean? And how should you interpret these values?**

|  |
| --- |
| **For the OS of Mitoxandrone the highest AIC is exponential, the lowest the Weibull. The AIC value is a measure of prediction error. It is a measure of how well the predicted data fits the observed data. A low AIC means that the predicted data closely follows the observed data. This lowest AIC does not say anything about the absolute quality of the data, it is only a quality relative to the other models. In addition, this is only about the observed data. The AIC value says nothing about the extrapolated data. Therefore, the AIC value should be used with caution in selected the best survival model.** |

Next, repeat the steps you did in **Q1** for the OS of Caba and the PFS of both Mito and Caba. You will use the following Excel files correspondingly:

1. PFS of Cabazitaxel *(Hoyle\_and\_Henley\_ PFS\_Caba.xlsm)*.
2. PFS of Mitoxantrone *(Hoyle\_and\_Henley\_ PFS\_Mito.xlsm)*
3. OS of Cabazitaxel *(Hoyle\_and\_Henley\_Bahl\_OS\_Caba.xlsm)*

**Q5: After you filled-in the “***Number events & censored”-***sheet in all the Excel files, you can now use the R code from Part II. This code will repeat the exercise of Q4 for all the dataset at once.**

**In addition, the R code exports the results in an excel sheet. This file will show-up in your output folder. Open the folder “output\_parametric\_survival\_model\_parameters.xlsx**

**Q6:**

Open the “*extrapolation*” sheet of the “*Cabazitaxel model template.xlsm”* file, and paste the corresponding AIC, intercept and logscale parameters for each distribution to the corresponding cells (*AD31:AS33*). Make sure you save this file. You need this in the follow-up assignment.

**Q7:**

For each of the 4 curves (PFS caba and mito, OS caba and mito), which parametric model appears to have the best statistical fit, based on AIC? And the worst?

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | |  | Best | Worst | | PFS caba | weibull | exponential | | PFS mito | weibull | exponential | | OS caba | loglogistic | exponential | | OS mito | weibull | exponential | |

**Part III: Parametric Survival Formula of the code**

The survival model parameters can be used the estimate survival probabilities at specific timepoints. The parametric formulations for exponential, Weibull, lognormal and loglogistic distributions are as follows.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ***Exponential*** | ***Weibull*** | ***lognormal*** | ***loglogistic*** |
| Formula | Lambda (λ) =  exp(-intercept) | Lambda =  exp(-intercept/scale)  Gamma (γ) = 1/scale | Lambda = intercept  Gamma = scale | Lambda =  exp(-intercept/scale)  Gamma= 1/scale |

**Q8: For the OS of mitoxantrone, based on the formulae above and the parameters obtained from R, please calculate the S(t) for t = 16 weeks for each of the 4 distributions** (recall that *scale= exp(logscale) = elogscale* and the *intercept* and *logscale* values are based on a monthly time scale, *ɸ(x)* is the standard cumulative normal distribution function, and its values can be found through [z-tables](https://www.math.arizona.edu/~rsims/ma464/standardnormaltable.pdf)). **You can use R as your calculator.**

|  |
| --- |
| **Exp:**  **Lambda = e^(-intercept) = e^(-2.7365) = 0.064796804**  **S(16 weeks) = S((16/52)\*12 months) = e^(-lambda\*(16/52)\*12)) = 78.7%**  **Weibull:**  **Scale = e^(logscale) = 0.614673774**  **Gamma = 1/scale = 1.626879235**  **Lambda = e^(-intercept/scale) = e^(-2.7515/0.614673774) = 0.011374263**  **S(16 weeks) = S((16/52)\*12 months) = e^(-lambda\*((16/52)\*12)^gamma) = 90.9%**  **Logn:**  **Lambda = intercept = 2.4421**  **Gamma = scale = e^(-0.2315)**  **S(16 weeks) = S((16/52)\*12 months) =1- *ɸ* ((ln((16/52)\*12) - lambda)/gamma)=1- *ɸ(-1.4317)=1-0.07636 = 92.4%***  **Logl:**  **Scale = e^(logscale) = 0.444**  **Gamma = 1/scale = 2.252**  **Lambda = e^(-intercept/scale) = e^(-2.4736/0.444) = 0.003807**  **S(16 weeks) = S((16/52)\*12 months) = 1/(1 + lambda \* t ^ gamma)) = 93.3%** |

**## Part IV Organizing Results for Reporting - Cholesky Matrix**

**Make sure you run the last chunk of R code. This will estimate and organize the Cholesky Matrix. The R code also exports an Excel file for you callend “output\_cholensky.xlsx”. Open this file and** fill in the related Cholesky decomposition of the covariance matrices in cells *AC5:AO22* on the “*extrapolation*” sheet of the “*Cabazitaxel model template.xlsm”*. You will not use the Cholesky matrix outputs this week, but they will be necessary for the probabilistic sensitivity analysis part later in this course.

Check the cell formulae in the same sheet in columns D to S, under row 15. What are these formulae? Compare the results you found in the previous question for OS results in week 16, with the corresponding Excel results you have in front of you.

|  |
| --- |
| **The formulae are the same as the ones given above, but now in excel language instead of mathematical notation. The results you calculated for week 16 (Q6) should be more or less the same** |

1. Briggs A, Schulper M, Claxton K. Chapter 4. Making decision models probabilistic. In: Decision Modelling for Health Economic Evaluation. Oxford, UK: Oxford University Press; 2007: 77-120. [↑](#footnote-ref-1)
2. Bahl, A., et al. "Impact of cabazitaxel on 2-year survival and palliation of tumour-related pain in men with metastatic castration-resistant prostate cancer treated in the TROPIC trial." Annals of Oncology 24.9 (2013): 2402-2408. [↑](#footnote-ref-2)
3. de Bono JS, Oudard S, Ozguroglu M, et al. Prednisone plus cabazitaxel or mitoxantrone for metastatic castration-resistant prostate cancer progressing after docetaxel treatment: a randomised open-label trial. Lancet 2010;376:1147-54. [↑](#footnote-ref-3)
4. Hoyle and Henley BMC Medical Research Methodology 2011, 11:139 [↑](#footnote-ref-4)