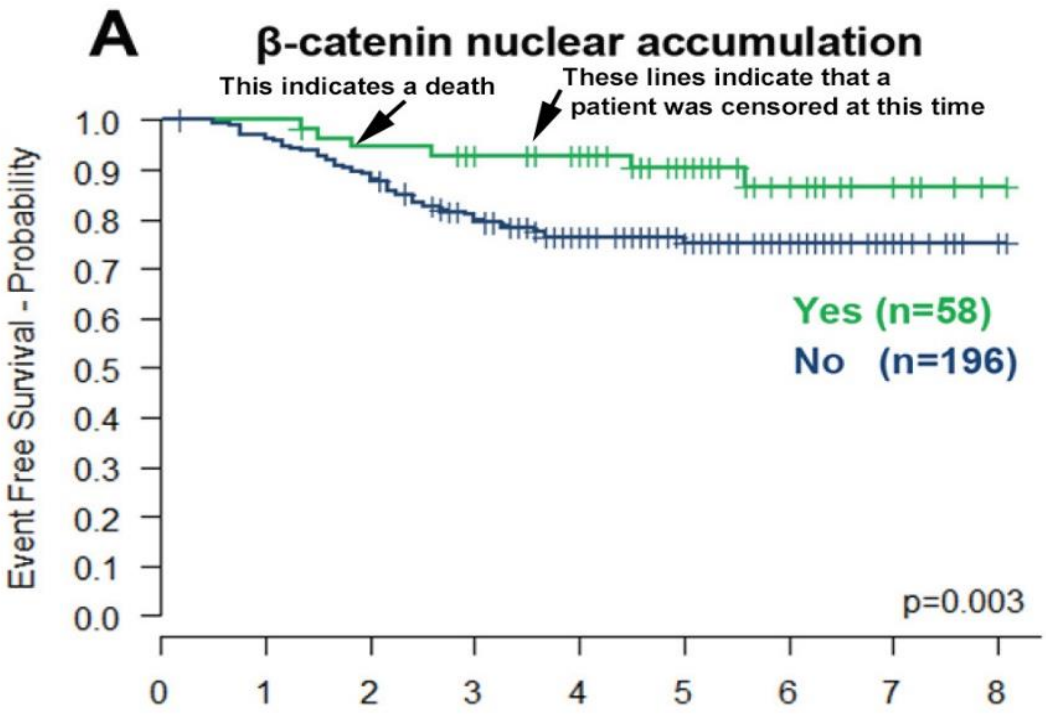
Kaplan-Meier Curve Data Extraction Method

This adapts the method described by Tierney et al., (Trials 2007; **8**:16). This method suggests printing curves and drawing lines to estimate event times (either death, disease recurrence or the patient being lost to the trial). This can be done more easily and accurately using a web-based tool such as WebPlotDigitiser. The Tierney paper provides an Excel file into which figures are entered and calculations performed, but the method described herein extracts data in a form that can be read by the *survival* function in R. Below is shown a worked example from the Clifford et al., (2015) paper looking at b-catenin nuclear localisation.

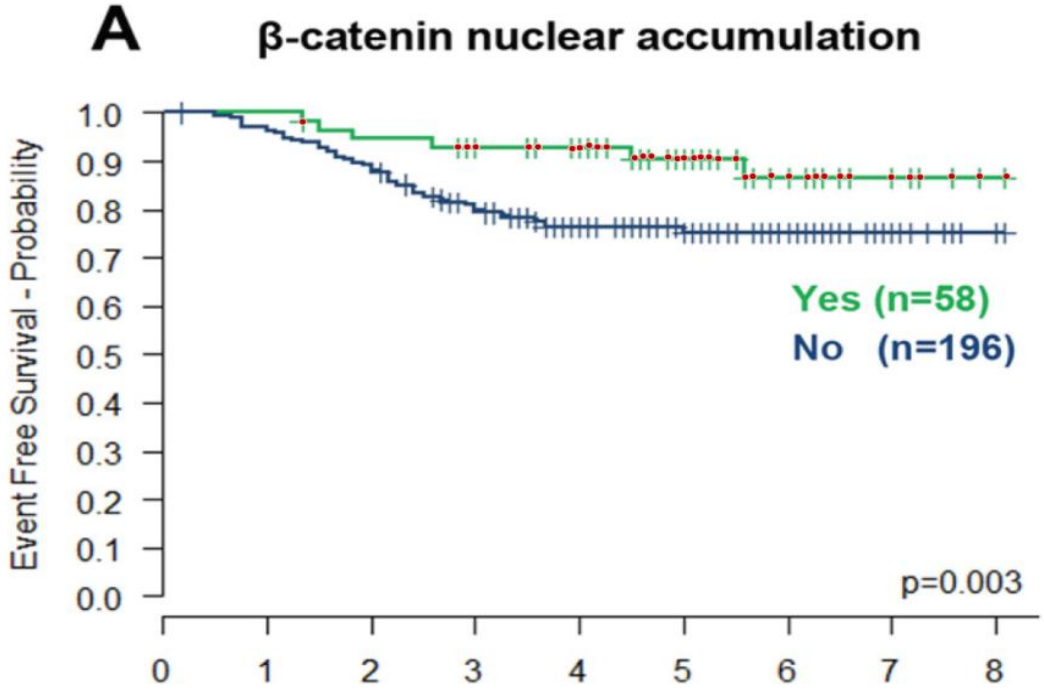
1.- Screen grab the relevant curve. Maximise the size of the screen area to give the best possible resolution. An annotated curve from the Clifford paper is shown.



2.- Upload the image (use File-> Load Image) as a 2D (X-Y) plot to the WebPlotDigitiser and follow the calibration instructions (use Align X-Y axes).

3.- Start plotting dots on the first survival line to give values for the time when patients were censored. In the attached image, dots on censored patients in the b-catenin group have been placed.

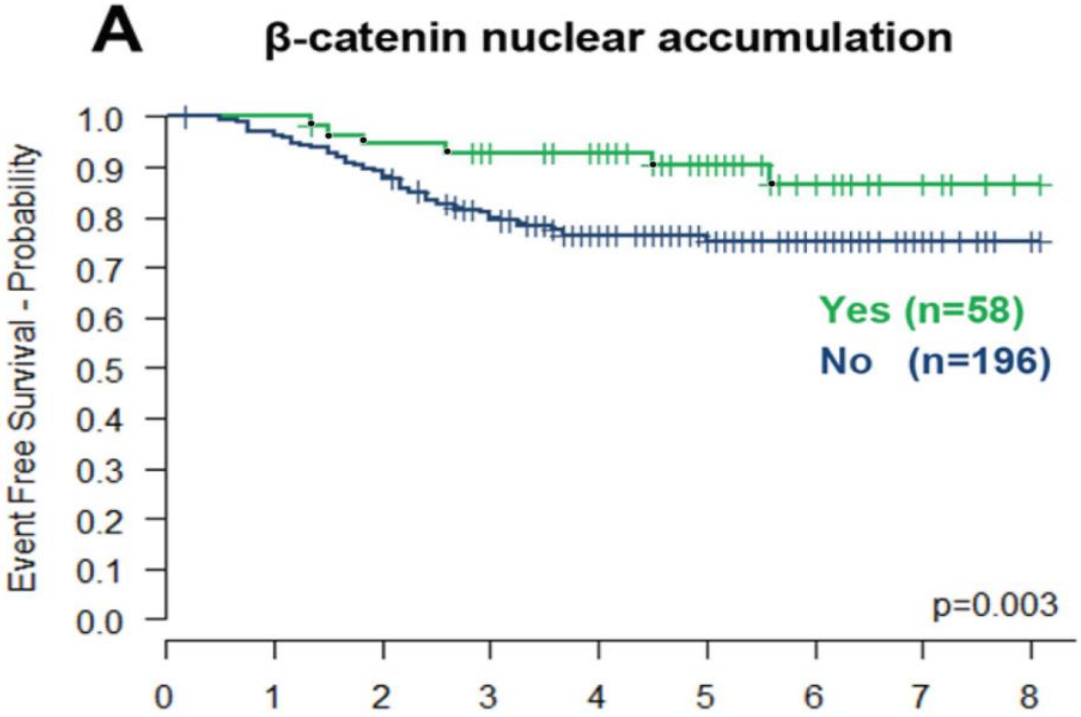
If you select “view data” you can see the time (in years) when these events occurred in the first column.



4.- Begin to populate a survival table as follows. This should be done in Excel, but a Word table is shown. As these points are censored, no event has occurred at these timepoints, and a “0” is entered. This group has been called one, the intervention group. Continue until all the dates for censored patients are entered. Note that data can be exported as a comma separated value (CSV file) .

Subject	Event	Time	Group
1	0	1.3	1
2	0	2.8	1
3	0	2.9	1

5.- Repeat for the deaths. I used Datasets->Add dataset->Add single dataset and set the colour to black. In this was you download all data in a .csv file.



5.- This stage introduces additional complexity, as each drop in the survival curve does not necessarily correspond to a single death or relapse. However, based on the paper, it is confirmed that six patients experienced relapse, which aligns precisely with the number of drops observed, thereby validating the accuracy of the event data. These events are recorded as follows:

Subject	Event	Time	Group
38	1	1.3	1
39	1	1.5	1
40	1	1.8	1

6.- The cohort initially comprised 58 patients, of whom 37 experienced events and an additional 6 relapsed, resulting in a total of 43 patients lost during the study period. Consequently, the remaining 15 patients were alive at the conclusion of the study, approximately 8.2 years post-enrolment. These outcomes are recorded as follows:

Subject	Event	Time	Group
44	0	8.2	1

45	0	8.2	1
46 etc	0	8.2	1

7.- For the β-catenin negative group, repeat the same analytical steps, ensuring the group is designated as "0" in the Excel dataset. Subsequently, re-plot the data in R, calculate the 5-year survival rates, and compare the results with those reported in the referenced paper. This process allows for iterative refinement of the extracted data, as needed. Additionally, using Excel's *autofill* and *autoincrement* functions is recommended to efficiently populate columns, such as those detailing patient numbers, thereby reducing manual input time.

References

Clifford, S. C., Lannering, B., Schwalbe, E. C., Hicks, D., O'Toole, K., Nicholson, S. L., Goschzik, T., et al. (2015). 'SIOP-Europe PNET Group. Biomarker-driven stratification of disease-risk in non-metastatic medulloblastoma: Results from the multi-center HIT-SIOP-PNET4 clinical trial', *Oncotarget*, 6(36), pp:38827-39. doi: 10.18632/onco target.5149.

Tierney, J. F., Stewart, L. A., Ghersi, D., Burdett, S. and Sydes, M. R. (2007). 'Practical methods for incorporating summary time-to-event data into meta-analysis'. *Trials*, 8(16). doi: 10.1186/1745-6215-8-16.