

## Instructions for X-Tile

X-Tile is best used in the Google Chrome browser.

Drag and drop a tab- or comma-delimited text (.txt) file onto the window. It will turn green.

Once loaded, your variables will be displayed on the left of the screen.

Click a variable to display a list of its values (click again to hide)

Right-click on a variable to make it a censor, time, marker, or training/validation marker

There are a few limitations on variables:

The time and marker variables must be numeric

The training/validation marker must include the terms: "test" or "training", and "validation"

The censor variable will be encoded as follows:

Uncensored = "uncensored", "dead", "died", "d", "recur", "recurred", or 1

Censored = anything else (except blanks)

You may use your own training/validation variable OR

X-tile will generate a Training/Validation variable (the last one on the list).

This variable is created by:

- 1) creating two separate lists of "censored" and "uncensored" variables
- 2) sorting the lists by the "time" variable
- 3) the first row of each list is placed in the "training" set, the second is placed in the "validation" set, followed by "training" then "validation." This is repeated for both lists.

Note that only patients with a recorded "time", "censor", and "marker" value are used.

That means the training/validation cohorts will be different depending upon the chosen marker.

Once a censor, time, and marker variable have been selected, an Options and Create Plot button will appear.

### Options:

Analysis Type:

Survival Analysis – standard survival analysis.

Means Test – finds cutpoints by analyzing the expression on another variable. This is NOT a survival analysis. Additional details are described below.

X-Tile Display:

Hi vs. Low – for a two-cutpoint (3-population) X-Tile plot, Log rank Chi-Square values are based on the high and low populations (the middle population is discarded)

Hi/Mid/Low – all three populations are accounted for.

Note that this only changes the X-Tile display. Both Hi/Low and Hi/Mid/Low Log rank Chi-Square values are shown in the stats area.

Min Pop Size:

The smallest population size (percent) that you consider to be a valid

This is predominantly used in calculating the corrected P-value (see below), as well as determining the maximum Log rank Chi-Square values

Max Chi:

Any Log rank Chi-Square value above this will be displayed at maximal intensity red or green

Time Cutoff:

The time-point above which all patients are censored

Initially the value is set at the maximum value of your time censor (so no cutoff)

You will need to reduce the number based on the type of time variable you are using (e.g., if your time variable is in months and you wish a 10-year cutoff, enter 120).

Show Uncorrected P-values:

Check this box if you wish to see uncorrected p-values. Warning: these p-values are not valid when analyzing multiple, different cutpoints.

### **X-Tile plots without a training/validation cohort:**

Click the “Generate Plot”

Plot generation will take some time depending upon the size your cohort.

The screen will contain a histogram and an X-Tile plot

Move your cursor over the plots to select different cutpoints

Cutpoints are then represented on the Kaplan-Meier plot and the stats display

If you click on the histogram or the X-Tile plot, the screen freezes. Click again to unfreeze.

To view a different marker, right-click on it and set it to “marker”. Then press the “Generate Plot” button

### **Stats Display:**

The size (n), % size, # events, overall survival, and min and max marker values of each population (Low, Middle, and High) are displayed.

The Log rank Chi-Square for the current cursor position is displayed, as well as for the best 3-population and 2-population cutpoints. Clicking on hyperlink text will bring you to that position in all windows.

A corrected p-value is displayed for any cutpoint that consists of a high and low population (without a middle one). Calculation of this corrected p-value is performed using the procedure outlined in: Dangers of Using "Optimal" Cutpoints in the Evaluation of Prognostic Factors Altman et al. JNCI v86 #11, 1994 pg. 829-835.

A Cox continuous univariate analysis of the marker. Note this is based on the data as a continuous variable and is not dependent upon the current cutpoint (so it doesn't change when moving the cursor around).

### **X-Tile plots WITH a training/validation cohort:**

When a training/validation cohort is set (along with a marker, censor, and time variable), the “Generate Training Set Plot” button will appear. Click it to display the plots.

When you have identified a cutpoint that you want to test (either by clicking on the X-Tile or histogram, OR by clicking the Best 3 pop or Best 2 pop hypertext in the stats display), the “Check Cutpoint on Validation Set” button will appear. Click it to show the validation cohort.

X-Tile will automatically cut the validation cohort at the same place you identified in the training cohort. Note that the Kaplan-Meier plot and stats display will show the p-value of that cutpoint on the validation cohort.

You may press the “Redisplay Training Set Plot” button to go back to the training set, BUT selecting cutpoints until you find one that validates is not statistically valid.

### **Saving X-Tile Plots:**

Click on the “Create Image” button.

When it appears, click on the “Download File” button.

The file will be downloaded and available at the bottom of your browser window.

The plots are saved in the vector-based .svg format and can be edited using Adobe Illustrator or the like. They can also be displayed in the browser window.

### **Means Test:**

Rather than finding an optimal survival cutpoint, this option allows you to define a cutpoint on one marker based on the expression of another marker. To use this feature, select a “marker” and a “dependent marker” variable. Then run the test. The analysis produces a Fisher F-statistic and associated p-values.

Questions? Please contact me at [robert.camp@yale.edu](mailto:robert.camp@yale.edu).