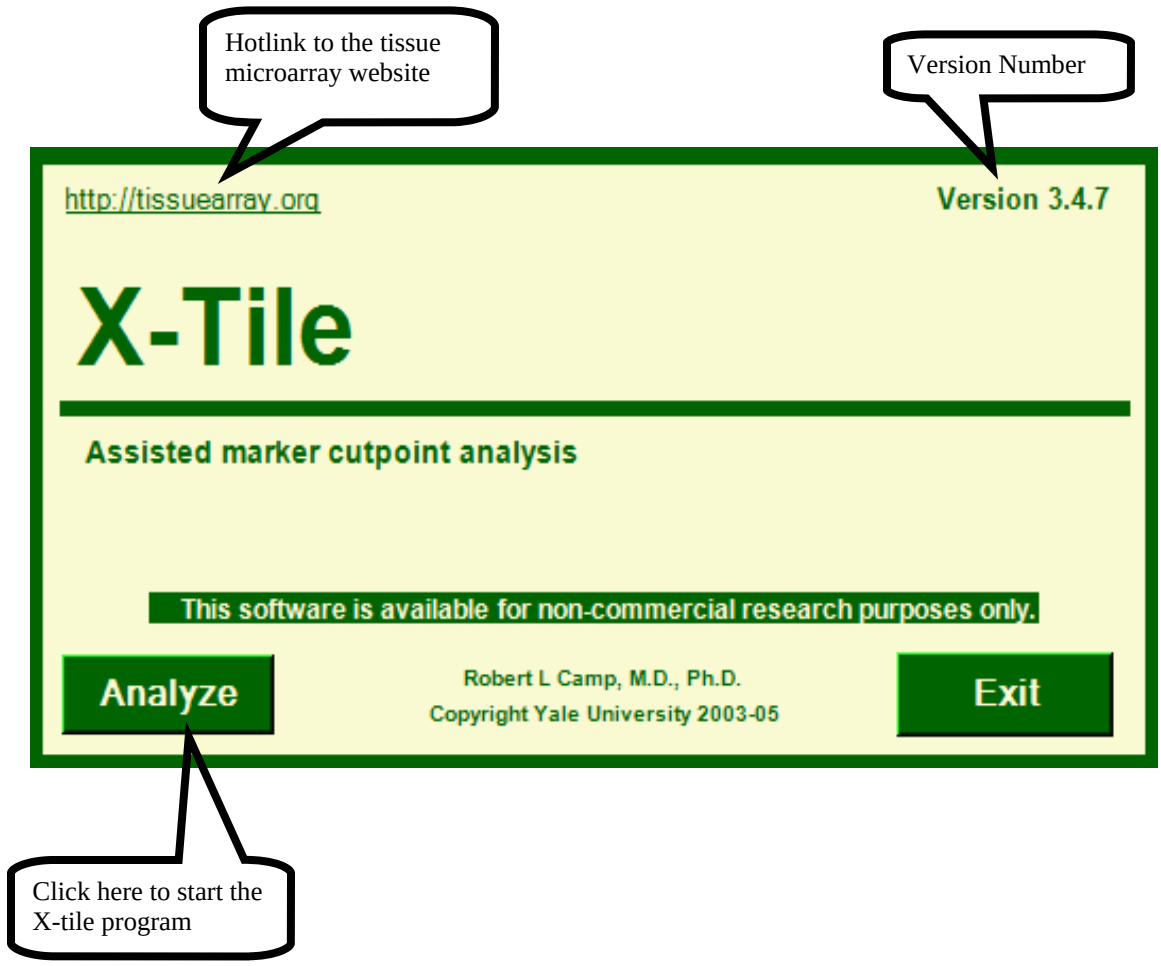
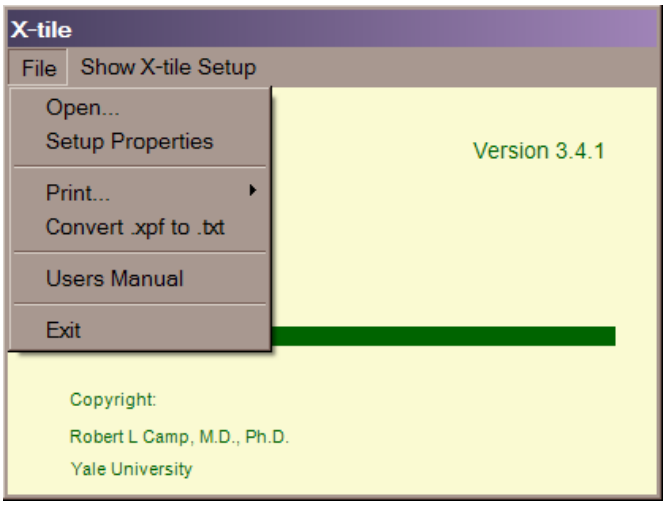
X-TILE USER'S MANUAL

FOR USE WITH X-TILE 3.5.0 AND LATER VERSIONS



MAIN MENU OPTIONS



Open…

Lets you open tab-delimited text files (.txt) (see page 3) or previously created X-tile data files (.xpf).

Setup Properties

Lets you set up how X-tile runs and displays its analyses (see page 4).

Print…

Allows you to print copies of your X-tile plots or save them to a file.

Convert .xpf to .txt

Converts an X-tile data file (.xpf) into a tab-delimited text file (.txt)

Users Manual

Displays the user's manual in .pdf format

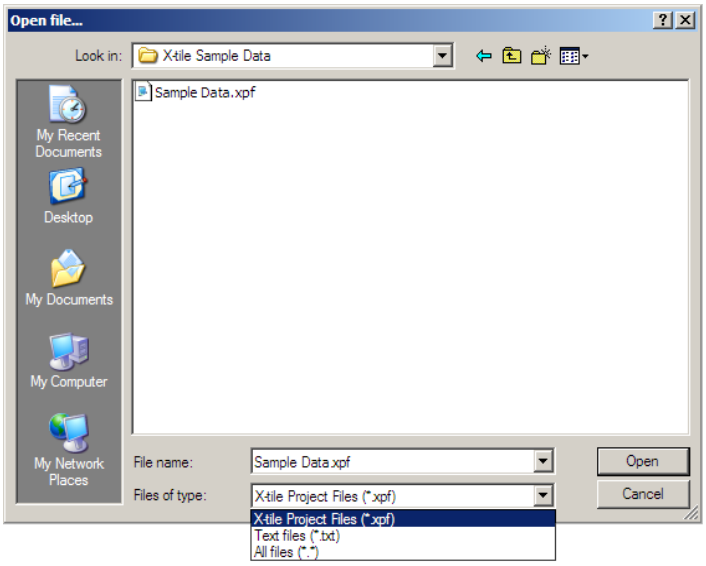
Exit

Exits the program

Show X-tile Setup

Toggles the setup screen on and off (see page 5).

FILE > OPEN FILES...



You can open any tab-delimited text file containing data. Once opened, X-tile creates an new folder containing a ".xpf" file which contains your data in xml format, and a ".set" file which contains information on how you want X-tile to generate your plots.

Once the ".xpf" file is created, you can directly access it when you restart X-tile.

X-tile ignores data cells that are blank, contain only a period ("."), or are "-1".

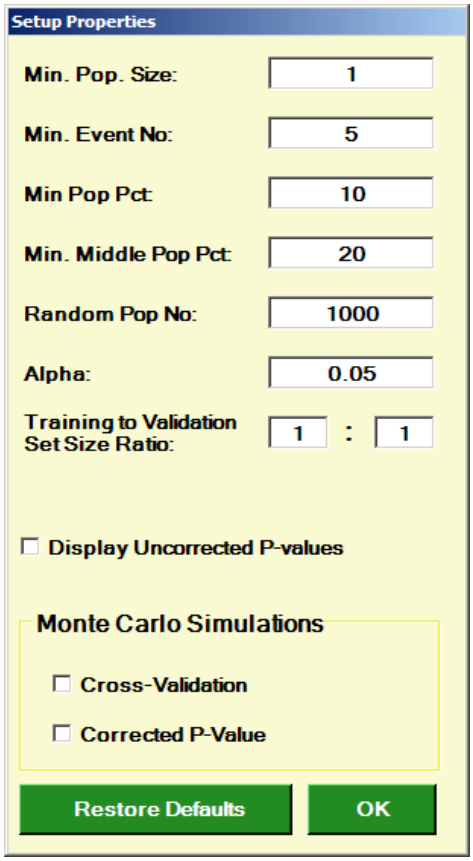
Once your data is converted to an ".xpf" file, you can convert it back to a tab-delimited text file using the main menu option. This is useful if you want to record any test/validation set splits that X-tile creates for you (see page 7).

X-tile comes with a sample data set in both ".xpf" and ".txt" formats. These are located in the "c:\program files\x-tile\" subdirectory.

NOTE that X-tile must be able to read and write files to/from the location of your data files (e.g. it cannot work with data that is on a CD or a networked computer for which you do not have full read/write privileges).

FILE > SETUP PROPERTIES

Allows you to define how subsequent X-tile analyses will be setup. To access this window, a dataset must be already loaded.



Min Pop Size:
The minimum number of patients that you consider to be a clinically valid sub-population.

Min Event No:
The minimum number of events (e.g. deaths) required within a sub-population for you to consider it to be a statistically valid sub-population.

Min. Pop. Pct:
The minimum percent of the total patient cohort that you consider to be a clinically valid sub-population.

Min Middle Pop Pct:
The minimum percent of the total patient cohort that you consider to be a clinically valid **middle** sub-population.

Random Pop No:
The number of random populations you want to use to develop Monte Carlo p-value simulations.

Alpha:
The cutoff you consider to be "statistically significant." Play no role in statistical analyses per se. Usually set at 0.05.

Training to Validation Set Size Ratio:
Allows you to alter the ratio of cases in the training set vs. the validation set when using X-tile to create such subsets. Initially set a 1:1.

Display Uncorrected P-values
Check to display uncorrected ("traditional") p-values.

Monte Carlo Simulations

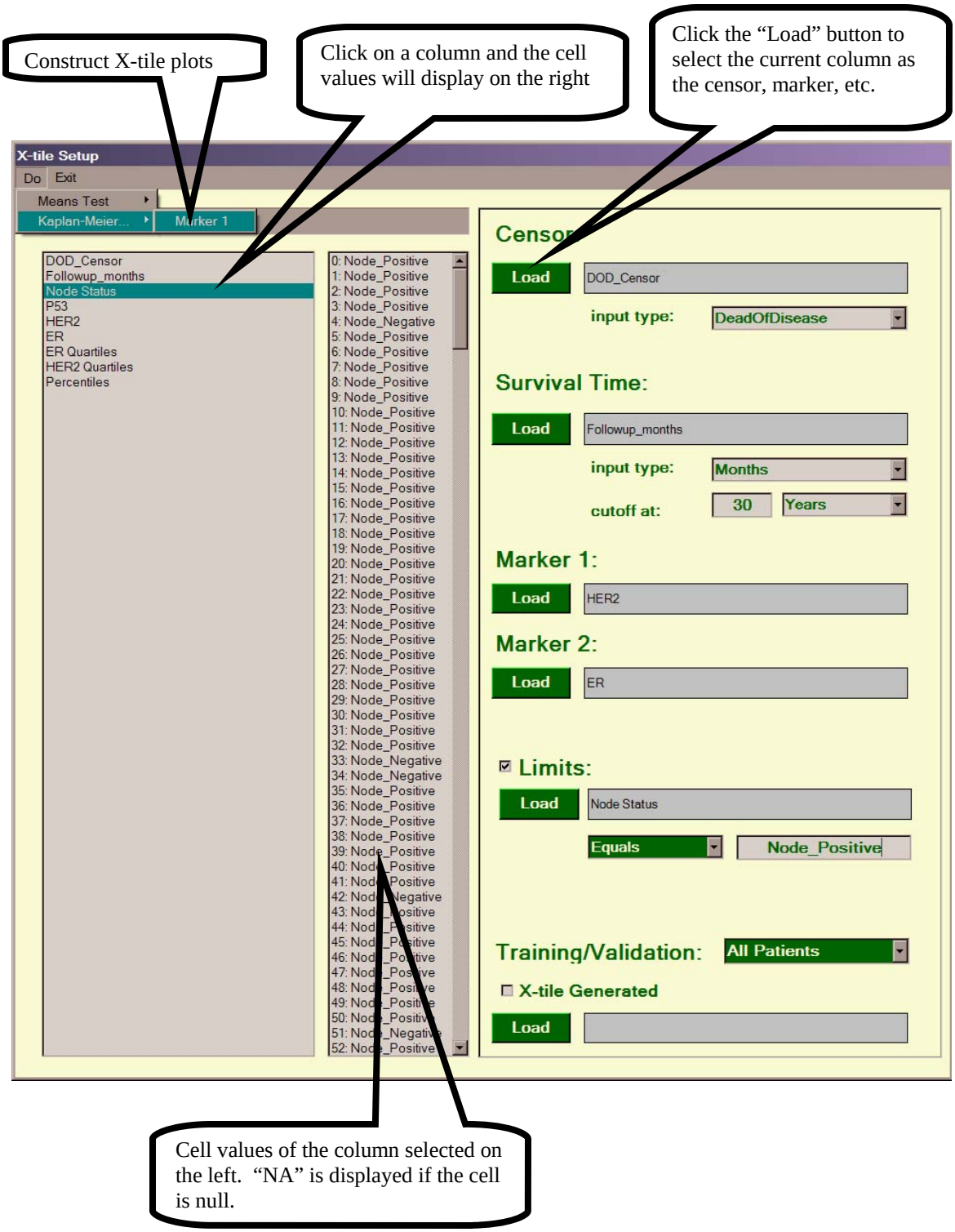
To avoid the problem of multiple cut-point selection (see page 14), X-tile can produce corrected p-values using several Monte Carlo simulations:

Cross-Validation takes your dataset, randomly splits it into two halves, finds the optimal cut-point of one half, and then divides the other half according to this cut-point. Then, it finds the optimal cut-point of the second half and similarly divides the first. Finally, X-tile performs a survival analysis of the entire dataset based on these optimal cut-points. Because the initial division into halves is random, the results obtained would vary every time a cross validation was performed. To overcome this problem, X-tile performs the cross-validation hundreds of times (based on the **Random Pop No**) and averages the results. Note that altering the "Training to Validation Set Size Ratio" does not affect the cross-validation algorithm (see reference 5).

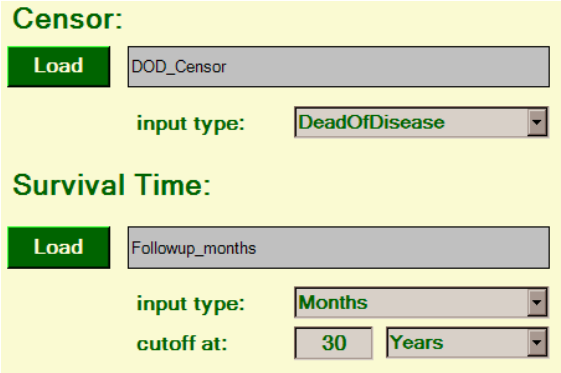
Corrected P-Value finds the optimal cut-point for your cohort and then compares the survival difference to that obtained from hundreds to thousands of random populations (based on the **Random Pop No**).

Restore Defaults restores the default values for the setup properties.

X-TILE SETUP



X-TILE SETUP — CENSOR AND SURVIVAL TIMES



Censor

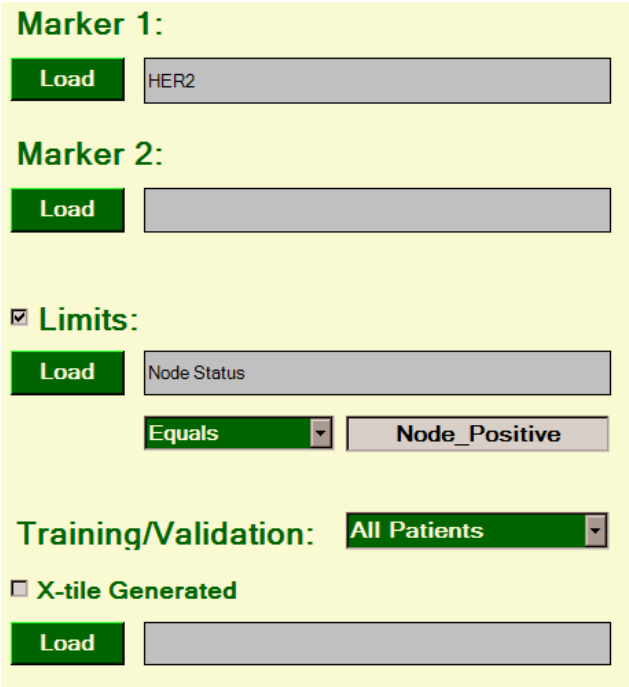
X-tile accepts censor data in several forms, and assigns cases as "censored", "uncensored", or "unknown", depending upon the selected "input type."

Cell Value	Alive / Dead	Dead of Disease	Dead with Disease	Recurrence
0	censored	censored	censored	censored
1	uncensored	uncensored	uncensored	uncensored
N or No	censored	censored	censored	censored
Y or Yes	uncensored	uncensored	uncensored	uncensored
A or Alive	censored	censored	censored	unknown
D or Dead	uncensored	unknown	unknown	unknown
AWD	censored	censored	censored	uncensored
ANED	censored	censored	censored	censored
DWD	uncensored	censored	uncensored	uncensored
DOD	uncensored	uncensored	uncensored	uncensored
DNED	uncensored	censored	censored	censored

Survival Time

The survival time must be numeric. Set the input type to match the inputted data and the cut-off to the desired value. Patient survival times will be truncated if they extend beyond the cutoff time. Patients with events occurring after the cutoff time will be considered "censored."

X-TILE SETUP — MARKER DATA



Marker 1 / Marker 2

Kaplan Meier analyses use "Marker 1" to divide the data and ignore "Marker 2."

Means Testing compares the distribution of "Marker 1" to "Marker 2"

X-tile can analyze continuous as well as categorical / nominal data; however, the values must be numeric (e.g. 0,1,2,3).

Limits

You can limit the analysis of your dataset based on the values in one of the data columns. Select the checkbox to turn the limits on, and deselect it to turn them off. Select "equals" for non-numeric columns, or "greater than", "less than", or "between" for numeric columns.

Training/Validation

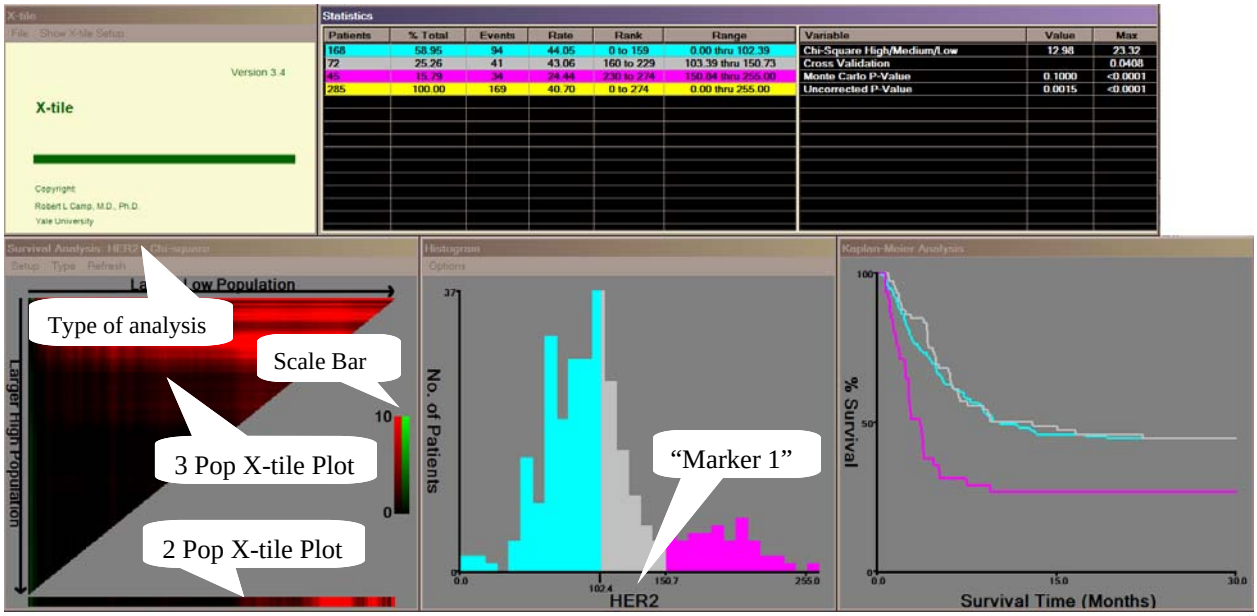
Check the "X-tile Generated" checkbox if you want X-tile to generate a randomized "training" and "validation" cohorts. When selected, X-tile will always generate the same random sets. Sets are normalized so that their base survival curves are similar. When X-tile generates "training" and "validation" sets, the results are stored in a column called "x-tile train/valid." The "training" and "validation" patients may be different depending upon what marker is being analyzed (e.g. because of missing values, etc.).

Alternatively, if you have already selected your "training" and "validation" sets, deselect the "X-tile Generated" checkbox, and a "Load" button and box will appear. Select the column with your "training/validation" data. Data cells must state specifically "training" or "validation" for that particular patient to be included in the appropriate set.

You can choose to view the X-tile plot of the "training" or "validation" cohorts, or for the entire population, "all patients." **For rigorous statistical analysis, view the "training" cohort first, decide upon an appropriate cut-point and apply that cut-point to the "validation" cohort. If you have applied the cut-point to the "validation" cohort and found that the cut-point is not significant , you SHOULD NOT return to the "training" set to define a new cut-point, because of the problem of multiple cut-point selection (see page 14).**

Instead of using "training" and "validation" cohorts, you can assess statistical significance by analyzing the entire cohort with the aforementioned Monte Carlo simulations (see pg. 4).

X-TILE CONSOLE — SURVIVAL ANALYSIS

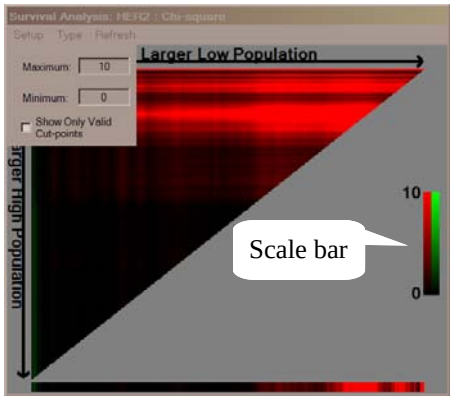


Using X-tile plots: Run the cursor over the X-tile plot (lower left), to divide the cohort. The histogram and Kaplan-Meier curves will adjust depending upon the divisions defined by the cursor position. The triangular X-tile plot allows you to divide the cohort into "High", "Middle", and "Low" subsets (3-population). The rectangular X-tile plot below the triangular plot, allows you to divide the cohort into "High" and "Low" subsets (2-population).

Finding the "optimal" cut-point. When the cursor is on the X-tile plot, press the left mouse button and the cursor will move to the point on the plot with the highest value for the currently displayed X-tile plot type. Once there the plot will lock (a lock icon will appear in the lower right-hand corner of the X-tile window). You can find the optimal cut-point on either the 2 or 3-population X-tile plot. The right mouse button toggles between the "locked" and "unlocked" states.

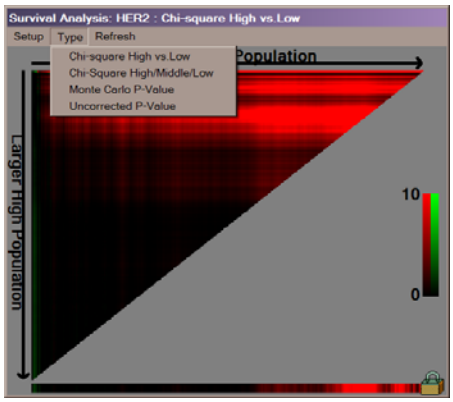
Construction of X-tile plots: X-tile plots are created by dividing marker data into three populations: low, middle, and high (i.e. two divisions). All possible divisions of the marker data are assessed. Associations can be calculated at each division by a variety of standard statistical tests, including the log-rank test for survival, and means tests for associations between other marker data. The data is represented graphically in a right-triangular grid where each point (pixel) represents the data from a given set of divisions. The vertical axis represents all possible "high" populations, with the size of the high population increasing from top to bottom. Similarly, the horizontal axis represents all possible "low" populations, with the size of the low population increasing from left to right. Data along the hypotenuse represent divisions in which all cases belong to either the high or low population. Data points away from the hypotenuse define an additional "middle" population, which increases in size with greater distances from the hypotenuse. Coloration of the plot represents the strength of the association at each division, ranging from low (dark, black) to high (green or red). Indirect associations between marker expression and survival (e.g. high expression connotes poorer survival) are colored red, whereas direct associations are colored green.

X-TILE PLOTS



Setup Menu

The popup setup menu allows you to change the maximum and minimum for X-tile plots. This will alter the coloration of the plot as indicated on the scale bar. These values are specific for each X-tile plot type, and X-tile will remember the values and apply them to subsequent X-tile plots of the same type. The "Show Only Valid Cut-points" checkbox will limit the X-tile plot to only those populations that are valid based on your setup properties (see page 4). Invalid cut-points will appear as background grey.



OE = Observed Events
EE = Expected Events

Chi-Square High vs. Low

$(OE_{low}-EE_{low})^2 / EE_{low} + (OE_{high}-EE_{high})^2 / EE_{high}$

Chi-Square High / Middle / Low

$(OE_{low}-EE_{low})^2 / EE_{low} + (OE_{middle}-EE_{middle})^2 / EE_{middle} + (OE_{high}-EE_{high})^2 / EE_{high}$

The number of expected events is progressively calculated at each time-point using a standard log-rank calculation.

Type Menu

Chi-Square High vs. Low

Shows the Kaplan-Meier log rank chi-square value of the high vs. the low sub-population. In the triangular 3 population plot, the middle population is disregarded (see box left).

Chi-Square High/Middle/Low

Shows the Kaplan-Meier log-rank chi-square value of a high / middle / low sub-setting. Nothing will be displayed in the rectangular two population X-tile plot. This plot is good for defining cut-points for U-shaped distributions (e.g. where both the high and low subpopulations do better/worse than the middle population) (see box left).

Monte Carlo P-value

Shows a corrected 2 and 3 population p-value that is adjusted to avoid the problem of multiple cut point analysis.

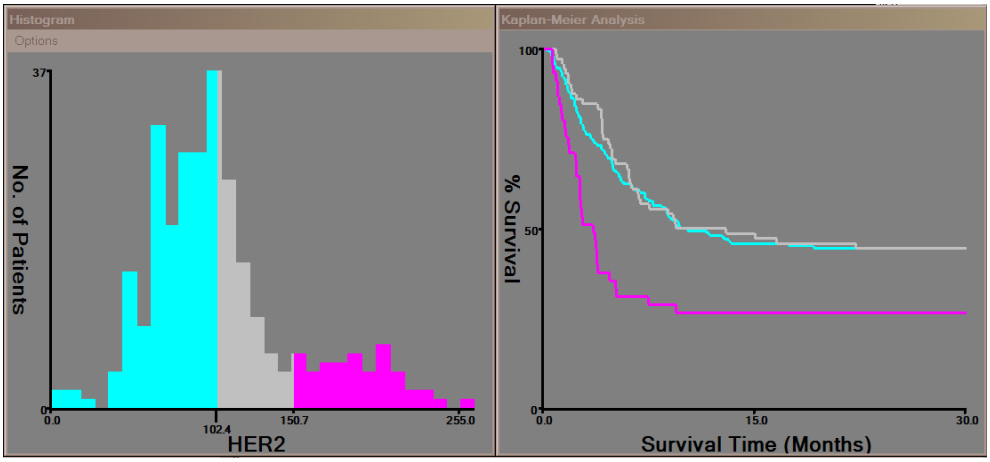
Uncorrected P-value

Shows uncorrected p-values. Be aware of the problems of multiple cut point analysis (page 14). You will only see this plot if you have selected the checkbox on the setup properties menu (see page 4).

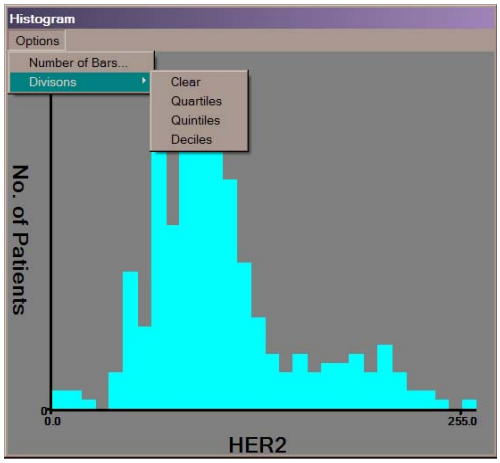
Note that p-value calculations are always based on the three population chi-square value for the triangular X-tile plot and the two population chi-square value for the rectangular X-tile plot.

Refresh causes the plot to be redrawn.

HISTOGRAM AND KAPLAN-MEIER PLOTS



Moving the cursor on the X-tile causes the histogram and Kaplan-Meier plots to display the corresponding populations. Moving the cursor over the histogram will cause the Kaplan-Meier plot to update accordingly (but has no effect on the X-tile plot). The left mouse button will create a cut point at the current position on the histogram. You can create as many cut points as you want and the corresponding sub-populations will be shown on the Kaplan-Meier curve.



Number of bars alters the number of bars that divide the data.

Divisions

Clear removes the existing divisions

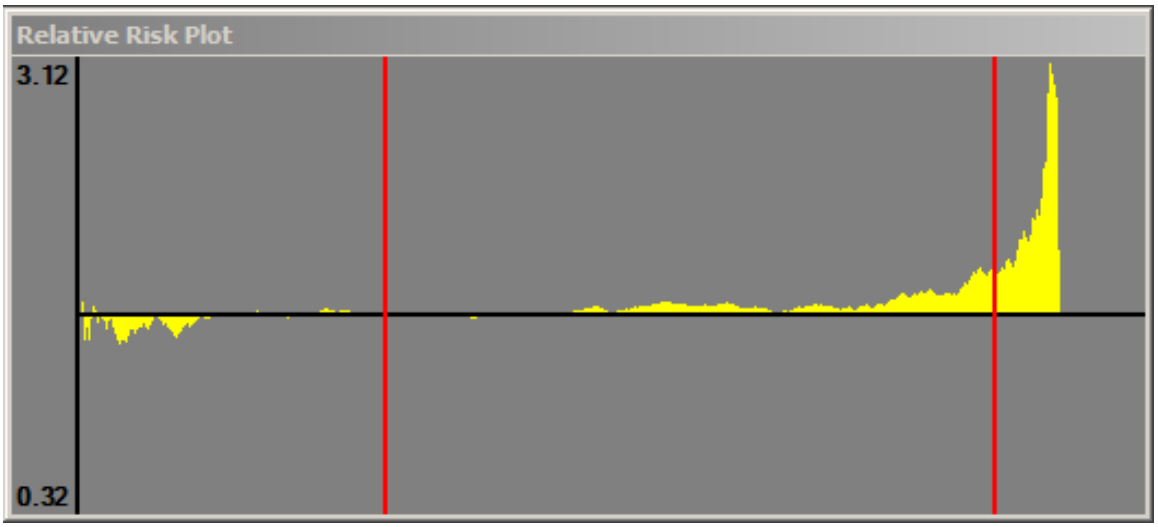
Quartiles divides the population into quarters

Quintiles divides the population into fifths

Deciles divides the population into 10 subsets

Note that in some datasets where there are several cases with the same marker value, the divisions option requires that all cases with the same marker value be placed in the same subset. Consequently, the divisions may not always result in equally sized subsets.

CONTINUOUS RELATIVE RISK PLOT



This window plots the relative risks for all cutpoints from low to high (left to right, x-axis). Relative risks (RR) are calculated as (event in high population / event risk in low population). A RR of 2 means the high population is two times more likely to have an event (e.g. die of their disease) than the low population. A RR of 0.5 means the low population is two times more likely than the high population to experience the event. An RR of 1.0 means the is no difference in the event rate between high and low populations. The red bars indicate current cutpoints displayed on the X-tile plot and/or histogram. Note that the Y-axis of the graph is log (base 2) rather than linear, and normalizes the relative risks around 1.0. Thus a RR of 0.5 and a RR of 2 are equidistant from RR = 1 (the x-axis). The relative risks are also stored in a file (ending in _RR.txt) in the X-tile subdirectory containing your data. X-tile does test for the statistical significance of relative risk, which is generally assessed using a Cox proportional hazards model (available in many commercial statistics programs). In general, the relative risk plot is a graphic representation of the 2-population X-tile plot.

STATISTICS

The **left panel** shows data about the sub-populations in the X-tile, Histogram, and Kaplan Meier plots. The coloration of each line is the same as the color of the population identified in one of the plots. The last line provides a summary of all of the patients in the cohort. When the cursor is over the X-tile plot defines a cut-point where one or more of the sub-populations fails to fulfill the setup properties (see page 4), the line color will dim.

The **right panel** shows statistical information based on the cut-points defined. These values will change depending upon where the cursor is (over the 3-population X-tile plot, the 2-population X-tile plot, or the histogram.

Chi-Square High/Middle/Low
Provides the 3-population log-rank chi-square value

Chi-Square High/Low
Provides the 2-population log-rank chi-square value

Cross Validation
Provides the overall significance of a cross-validation analysis (see pg. 4)

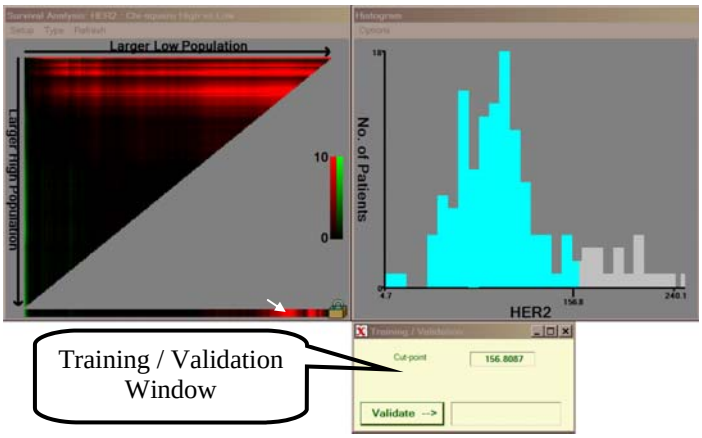
Monte Carlo P-Value
Provides a corrected p-value based by comparing to random populations (see pg. 4)

Uncorrected P-Value
Provides an uncorrected p-value using a traditional look-up table. This value will only show up if you have checked the "Display Uncorrected P-Values" box on the setup properties window (see pg. 4). **Note that this value will vastly overstate the significance of your data unless you have cut your data once and only once.**

Miller-Siegmund P-Value
Provides a corrected p-value based on the model proposed by R Miller and D. Siegmund (see reference 6). This value is only displayed for two sub-population cuts.

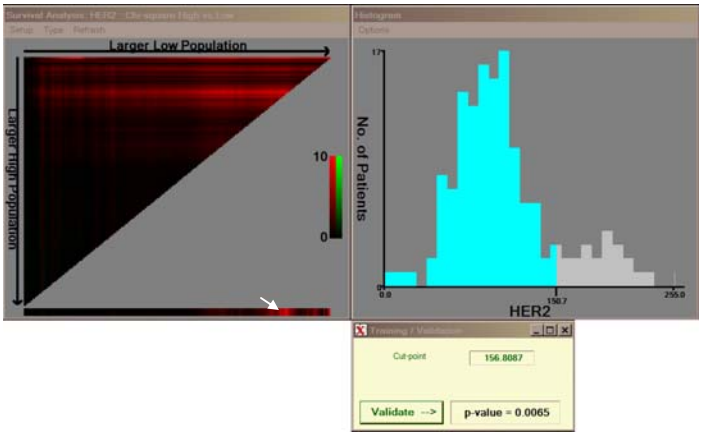
The "Max" column provides the maximum value for a variable for the **valid** population cut-points the X-tile plot. You may find that the "value" at non-valid cut-points (i.e. when the line color is dimmed) exceeds the "Max" value.

TRAINING SET / VALIDATION SET P-VALUE DETERMINATION



Training Set

On the X-tile setup window (see page 5) select the "Training Set" and run an X-tile Kaplan-Meier analysis (Do->Kaplan-Meier->Marker 1). On the X-tile plot select the most appropriate cut-point(s). You can allow X-tile to show you what it thinks is the appropriate cut-point (using the left mouse button) (see page 8) or select one yourself (using the right mouse button). Once selected, the **Training/Validation Window** will appear. This shows the cut-point you have selected. Press the "Validate" button to validate the cut-point.

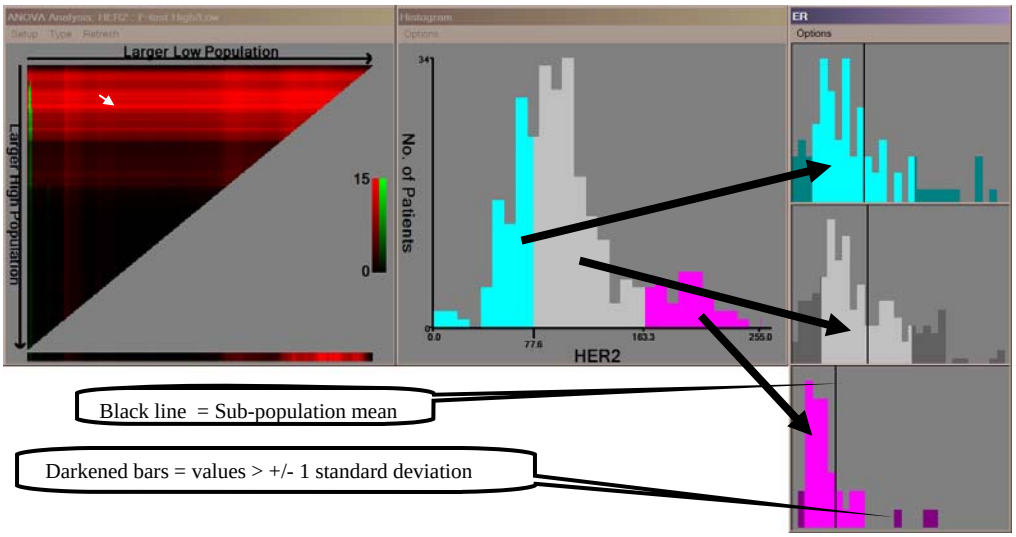


Validation Set

Once you have pressed the "Validate" button, X-tile will switch display the Validation Cohort and move the cursor to the cut-point you selected on the Training Set. X-tile will display the p-value for the selected cut-point to the right of the validate button.

Note that if your cut-point is not significant, you should NOT return to the training set to select another cut-point. Doing so will invalidate the significance of the p-value (see page 14).

X-TILE: MEANS TESTING



X-tile can also be used to perform means tests. X-tile calculates a standard Analysis of Variance (ANOVA) F-test score for two or three populations. Calculations of Monte-Carlo and Cross-Validation P-Values are similar to those for survival analyses (see page 4). As with survival analyses, an uncorrected p-value will vastly overestimate the statistical significance of your p-value if you test more the one cut-point(s) (see pg. 14). The windows for means testing are slightly different. Note the absence of a Kaplan-Meier curve and the addition of a window with three panels. X-tile uses the values of Marker 1 (see pg. 5), in the case above HER2, to cut the cohort. The values of Marker 2 (ER above) are then collected for each subpopulation, and a means test performed. In the case above, High HER2 expression (the magenta sub-population) shows low level ER expression. The mean of each sub-population is represented by a black line on each sub-panel. The dark bars represent values greater than one standard deviation from the mean.

Statistics

Pati...	% Total	Mean	Std Dev	Rank	Range
107	33.75	82.79	54.51	0 to 101	0.00 to 81.97
152	47.95	88.81	50.42	102 to 247	82.80 to 133.81
58	18.30	60.92	48.89	248 to 305	134.29 to 255.00
317	100.00	81.68	52.41	0 to 305	0.00 to 255.00

Variable	Value	Max
Monte Carlo P-Value	0.0600	<0.0001
Uncorrected P-Value	0.0023	<0.0001
F-test High/Medium/Low	6.1807	13.3336
F-test High/Low	9.3924	20.8799
Cross Valid. High vs Low	NA	0.5843
Cross Valid. High/Middle/Low	NA	<0.0001

The statistics window shows information similar to that in a survival analysis (see pg. 11). The left panel shows information about the different sub-populations. The mean and standard deviation values refer to the mean of Marker 2 when cut at a point along the Marker 1 values. Means testing is performed by standard Analysis of Variance (ANOVA) calculation on both 2-population (high vs. low) and 3-population (high / middle /low) cut-points. The F-test statistics represents the ratio of the variation (sum-of-squares) between the subsets to (divided by) the variation within the subsets. Statistical significance can be assessed by either training set / validation set analysis (see pg. 12) or Monte Carlo simulations (see pg. 4).

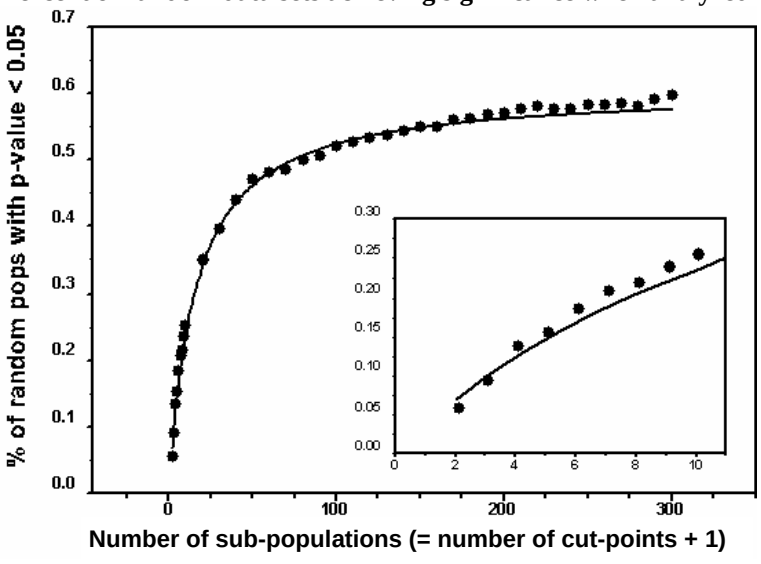
The statistics panel also provides you with the results of a **linear regression** model which provides a p-value based on a continuous assessment of the data.

OPTIMAL CUTPOINT ANALYSIS

The inherent problems of "optimal cutpoint analysis" or "the minimum p-value method" are well established (1-3), particularly in regards to parsing continuous data. In this method, marker data is divided, usually into two sub-populations (e.g. high vs. low expressers) based on a single cut-point. This cut-point is then varied until a statistically significant p-value is obtained. The problem with this approach lies in the fact that every cut-off analyzed represents an additional opportunity to achieve statistical significance by random chance. Traditional statistical tests (e.g. the log-rank test for survival) are dependent upon the use of one and only one cutoff. Analysis of multiple cut-points can dramatically increase the likelihood of finding an aberrantly low p-value (see figure below). Statisticians have noted that these discrepancies most likely explain why the significance of many prognostic markers cannot be confirmed in subsequent studies (4). The algorithms in X-tile, as described below, correct for this discrepancy.

The use of nominally acquired data (as is found in most traditional marker studies) may limit the number of potential cut-points, but does not eliminate this problem. Traditionally, marker studies are performed by manually scoring the intensity of immunohistochemically-stained tumor samples on a nominal scale (e.g. 0 to 3+). This scale still provides several ways to divide the data into "high" and "low" subsets (i.e. 0 vs. 3+, 0 and 1+ vs. 2 and 3+, 0 vs. the rest, 1 vs. the rest). Since there is no *a priori* reason to choose one over another cut-point, a researcher will frequently examine all possible divisions of the data, until one is found to be statistically significant. If this fails to provide a significant p-value, the data may be re-analyzed, either by averaging the results of an additional independent observer, or by a re-interpretation of where the cutoffs between 0, 1+, 2+ and 3+ should actually be. Either case further amplifies the number of possible cutoffs. By not correcting for the number of cutoffs, the researcher will report an incorrectly low p-value. The difference between the calculated and actual p-value when multiple cut-points are analyzed is not trivial (see figure below). Note the hyperbolic amplification of this difference upon analysis of 1-9 cut-points. Similarly continuous data (e.g. cohorts with >50 divisions) show a substantial difference between calculated and actual p-values; although increasing the number of divisions has far less impact on this difference.

Percent of random data sets achieving significance when analyzed with different point scales. A hypothetical cohort of 300 patients with 50% survival was used to determine the chance that a significant p-value (< 0.05) could be achieved when sequentially looking at multiple cut-points. A detail of assessing 1 to 9 cut-points is shown in the inset. As expected, looking once at a single cut-point dividing the cohort into 2 sub-populations (e.g. High vs. Low), 5% of random data sets achieve a p-value of 0.05 (inset). In this case, the requirement of examining a single cutoff is fulfilled, and a standard statistical analysis would provide an accurate p-value. In contrast, looking at 3 possible cut-points resulting in 4 sub-populations (e.g. 0, 1+, 2+, 3+) 13% of random data sets achieve a p-value of 0.05 (inset). This is because there are 3 ways to achieve statistical significance (0 vs 1-3, 0-1 vs 2-3, and 0-2 vs 3). Looking at all of the cut-points in a cohort of 300 patients (i.e. completely continuous data), the number of random data sets achieving a p-value of 0.05 approaches 60%. Note that the error in p-values increases in a hyperbolic fashion such that nominally scored data (e.g. < 10-point scale) can exhibit large errors with subtle alterations of the data (e.g. if the results of two scoring sessions are averaged).



SELECTED REFERENCES

(1) Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. J Natl Cancer Inst 1994;86:829-35.

(*2*) Altman DG. Suboptimal analysis using 'optimal' cutpoints. Br J Cancer 1998;78:556-7.

(*3*) Cantor AB, Shuster JJ. Re: Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. J Natl Cancer Inst 1994;86:1798-9.

(4) Hilsenbeck SG, Clark GM, McGuire WL. Why do so many prognostic factors fail to pan out? Breast Cancer Res Treat 1992;22:197-206.

(5) Faraggi D, Simon, R. A simulation study of cross-validation for selecting an optimal cutpoint in univariate survival analysis. Stat. in Med. 1996;15:2203-2213.

(6) Miller R, Siegmund, D. Maximally selected chi square statistics. Biometrics 1982;38:1011-1016.