Chapter 695

Two-Group Survival Comparison Tests with Weights (Simulation)

Introduction

This procedure uses simulation to determine power or sample size for the following survival comparison tests:

- Logrank
- Gehan-Wilcoxon
- Tarone-Ware
- Peto-Peto
- Modified Peto-Peto
- Fleming-Harrington with flexible *p* and *q*

Survival rates (hazard rates, median survival times, proportion surviving, or mortality) can be piece-wise customized to specify proportional hazard scenarios or non-proportional hazard scenarios. The treatment group survival rates can be specified directly or based on hazard ratios (or mortality ratios).

Loss-to-follow-up and noncompliance proportions can be entered with piece-wise flexibility. Noncompliance survival rates can be specified directly or can be set to the opposite treatment group survival rate.

Accrual times and accrual patterns are completely flexible.

Four Different Effect Size Parameterizations

There are four closely related effect size parameterizations that are available in this procedure and documented in this chapter. The parameterization can be in terms of hazard rates, median survival times, proportions surviving a given time period, or mortality during a given time period. When median survival times, proportions surviving, or mortality are used, the values are converted to the corresponding hazard rates before the simulation process begins.

Hazard Rate Parameterization

In this case, the hazard rates for the control and treatment groups are specified directly.

Median Survival Time Parameterization

Here, the median survival time is specified. These are transformed to hazard rates using the relationship $h = \ln(2) / MST$.

Proportion Surviving Parameterization

In this case, the proportion surviving until a given time *T*0 is specified. These are transformed to hazard rates using the relationship $h = -\ln(S(T0)) / T0$. When separate proportions surviving are given for each time period, *T0* is taken to be the time period unit.

Mortality Parameterization

Here, the mortality until a given time *T*0 is specified. These are transformed to hazard rates using the relationship $h = -\ln(1 - M(T0)) / T0$. When separate mortalities are given for each time period, *T0* is taken to be the time period unit.

Technical Details

This section outlines the simulation procedure and the test statistic details.

Simulation Procedure

In this procedure, a large number of simulations are used to calculate power using the following steps:

1. The total sample size is divided into the control and treatment group sample sizes (N1 and N2) according to the Group Allocation specified.
2. Based on the specified survival rates, and noncompliance proportions and survival rates, a hazard rate function is generated for each subject of each group. The hazard rate function is used to generate a random survival time and the loss proportion is used to determine whether the simulated subject is right censored due to loss before the event.
3. A starting time for each subject is generated based on the accrual time and the accrual pattern.
4. For each sample (of N1 + N2 subjects), a test statistic is produced. Based on the test statistic, it is determined whether the null hypothesis is rejected for each sample.
5. The proportion of rejected null hypotheses is the estimated power.

Test Statistics

This section presents methods for testing that the survival curves, and thus the hazard rates, of two or more populations are equal. The specific hypothesis set that is being tested is

$$H_0: h_1(T) = h_2(T) \text{ for all t} \leq \tau$$

$$H_1: h_1(T) \neq h_2(T) \text{ for some t} \leq \tau$$

In words, the null hypothesis is that the hazard rates of the two populations are equal at all times less than the maximum observed time and the alternative hypothesis is that the two hazard rates differ at some time less than the observed maximum time.

The general form of the test statistic is

$$Z = \frac{\sum_{i=1}^{D} W(t_i)\left[d_{i1} - Y_{i1}\left(\frac{d_i}{Y_i}\right)\right]}{\sqrt{\sum_{i=1}^{D} W(t_i)^2 \frac{Y_{i1}}{Y_i}\left(1 - \frac{Y_{i1}}{Y_i}\right)\left(\frac{Y_i - d_i}{Y_i - 1}\right) d_i}}$$

where

D is the number of distinct event times

$W(t_i)$ is the weight function at time t_i

Y_{i1} is the number at risk in the Group 1 sample at time t_i

Y_i is the combined number at risk at time t_i

d_{i1} is the number of events in the Group 1 sample at time t_i

d_i is the combined number of events at time t_i

Details of the above formulas can be found in Klein and Moeschberger (1997), pages 191-202, and Andersen, Borgan, Gill, and Keiding (1992), pages 345-356.

Six different choices for the weight function, *W*(*T*), with the flexible *p* and *q* for the Fleming-Harrington weight function, result in a variety of tests that are available in this procedure. The most commonly used test is the Logrank test, which has equal weighting. The other tests shift the heaviest weighting to the beginning or end of the trial. This may be appropriate in some studies, but the use of one of these other weighting schemes should be designated before the data have been seen. Because of the different weighting patterns, they will often give quite different results.

The following table describes each of these tests:

Test	**Weight**	**Comments**
Logrank	1	This is the most commonly used test. It places equal weights across all times. This test has optimum power when the hazard rates of the *K* populations are proportional to each other.
Gehan-Wilcoxon	Y_i	Places weight on hazards at the beginning of the study.
Tarone-Ware	$\sqrt{Y_i}$	Places weight on hazards at the beginning of the study.
Peto-Peto	$\tilde{S}(t_i)$	Places weight on hazards at the beginning of the study.
Modified Peto-Peto	$\tilde{S}(t_i)Y_i/(Y_i + 1)$	Places weight on hazards at the beginning of the study.
Fleming-Harrington (1,0)	$\hat{S}(t_{i-1})$	Places weight on hazards at the beginning of the study.

Test	Weight	Comments
Fleming-Harrington (0.5,0.5)	$\sqrt{\hat{S}(t_{i-1})\left(1-\hat{S}(t_{i-1})\right)}$	Places weight on hazards in the middle of the study.
Fleming-Harrington (1,1)	$\hat{S}(t_{i-1})\left(1-\hat{S}(t_{i-1})\right)$	Places weight on hazards in the middle of the study.
Fleming-Harrington (0,1)	$1-\hat{S}(t_{i-1})$	Places weight on hazards at the end of the study.
Fleming-Harrington (0.5,2)	$\sqrt{\hat{S}(t_{i-1})}\left(1-\hat{S}(t_{i-1})\right)^2$	Places weight on hazards at the end of the study.

This table uses the following definitions:

$$\hat{S}(t)=\prod_{t_i\le t}\left(1-\frac{d_i}{Y_i}\right)$$

$$\tilde{S}(t)=\prod_{t_i\le t}\left(1-\frac{d_i}{Y_i+1}\right)$$

Example 1 – Calculating Sample Size

A clinical trial is to be conducted over a three-year period to compare the survival distribution of a new treatment to that of the current treatment. The hazard rate of the current treatment is 1.4. Although the researchers do not know the true hazard rate of the new treatment, they would like to determine the sample size needed to detect a difference in hazard rates if the hazard rate under the new treatment is 0.8. The desired power is 0.90. Testing will be done at the 0.05 significance level with a two-sided Gehan-Wilcoxon test. All enrollees are enlisted at the beginning of the study.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design 1 Tab

Solve For	**Sample Size**
Test Type	**Gehan-Wilcoxon**
Alternative Hypothesis	**H1: Hazard1 ≠ Hazard2**
Simulations	**10000**
Random Seed	**3901161** (for Reproducibility)
Power	**0.9**
Alpha	**0.05**
Group Allocation	**Equal (N1 = N2)**
Input Type	**Hazard Rate**
h1 (Hazard Rate of Control Group)	**1.4**
Treatment Group Parameter	**h2 (Hazard Rate)**
h2 (Hazard Rate of Treatment Group)	**0.8**

Design 2 Tab

Controls Lost	**0**
Treatments Lost	**0**
For Treatments Lost under H0, use	**Controls Lost**
Noncompliance Proportion (Control)	**0**
NCh1 (Noncompliance Hazard, Control)	**1.0**
At time of noncompliance, start NCh1 at	**Current time**
Noncompliance Proportion (Treatment)	**0**
NCh2 (Noncompliance Hazard, Treatment)	**1.0**
At time of noncompliance, start NCh2 at	**Current time**
For Treatment Noncomp. under H0, use	**Control Noncompliance**
Accrual Time (Integers Only)	**0**
Accrual Pattern	**Equal (Uniform)**
Total Time (Integers Only)	**3**

Output

Click the Calculate button to perform the calculations and generate the following output. The calculations will take a few moments to complete. The results will vary slightly due to simulation differences.

Numeric Results (Scenario 1)

Solve For:	Sample Size
Groups:	1 = Control, 2 = Treatment
Hypotheses:	H0: Hazard1 = Hazard2 vs. H1: Hazard1 ≠ Hazard2
Test Statistic:	Gehan-Wilcoxon Test
Simulations:	10000
Pool Size:	20000
Random Seed:	3901161

Power			Alpha				
	95% C.I. Limits		Value		95% C.I. Limits		
Value	Lower	Upper	Target	Actual	Lower	Upper	Beta
0.903	0.897	0.909	0.05	0.053	0.049	0.057	0.097

Sample Size		Hazard Ratio	Hazard Rate		Accrual	Time	
N1	N2	HR	h1	h2	Pattern	Accrual	Total
92	93	0.571	1.4	0.8	Equal	0	3

		Noncompliance*			
Proportion Lost*		Proportion		Hazard Rate	
Lost1	Lost2	NCP1	NCP2	NCh1	NCh2
0	0	0	0	1	1

* The reported values are during a single time period.

Power	The probability of rejecting a false null hypothesis. It is the total proportion of alternative hypothesis simulations for which H0 is rejected.
Power 95% LCL and UCL	The lower and upper confidence limits for the power estimate. The width of the interval is based on the number of simulations.
Target Alpha	The user-specified probability of rejecting a true null hypothesis.
Actual Alpha	The alpha level that was actually achieved by the experiment. It is the total proportion of the null hypothesis simulations for which H0 is rejected.
Alpha 95% LCL and UCL	The lower and upper confidence limits for the actual alpha estimate. The width of the interval is based on the number of simulations.
Beta	The probability of accepting a false null hypothesis. It is the total proportion of alternative hypothesis simulations for which H0 is not rejected.
N1 and N2	The sample sizes of each group.
HR	The hazard ratio at which power is computed. HR = h2/h1.
h1	The hazard rate of the control group. It is the hazard rate that is simulated for both groups under H0, and for group 1 under H1.
h2	The hazard rate of the treatment group. It is the hazard rate that is simulated for group 2 under H1.
Accrual Time	The time during which subjects are enlisted into the study. It is sometimes known as the enlistment period or recruitment period.
Accrual Pattern	Describes the distribution of accrual across the Accrual Time.
Total Time	The total length of the study. It is the sum of the accrual time and the follow-up time.
Lost1, Lost2	The proportion lost to follow-up and right censoring per period for the control and treatment groups, respectively.

NCP1, NCP2	The proportion noncompliant in each time period for the control and treatment groups, respectively.
NCh1, NCh2	The noncompliance hazard rates for the control and treatment groups, respectively.

Whole Study Averages (Scenario 1)

Cumulative Subject Time				Events			
H0		H1		H0		H1	
G1	G2	G1	G2	G1	G2	G1	G2
64.7	65.3	64.7	105.8	90.6	91.6	90.6	84.6

Whole Study Averages	H0 and H1 refer to the simulations under the null and alternative hypotheses, respectively.
Cumulative Subject Time	The average total time of subject involvement. It is the average sum of survival times, including event survival times and censored survival times.
G1, G2	The cumulative subject times for groups 1 and 2, respectively.
Events	The average number of events in the study before study termination.

Summary Statements (Scenario 1)

A parallel two-group design will be used to test whether the Group 1 hazard rate is different from the Group 2 hazard rate (H0: Hazard 1 = Hazard 2 versus H1: Hazard 1 ≠ Hazard 2). The comparison will be made using a two-sided Gehan-Wilcoxon test with a Type I error rate (α) of 0.05. The accrual time will be 0 (with accrual pattern 'Equal (Uniform)') and the total time (accrual plus follow-up) will be 3. To detect a hazard ratio (h2 / h1) of 0.571 (hazard rate 1 = 1.4, hazard rate 2 = 0.8) with 90% power, the number of needed subjects will be 92 in Group 1 and 93 in Group 2. These results are based on 10000 simulations of exponential survival times according to the specified null and alternative distribution parameters.

Detailed Input (Scenario 1)

Groups: 1 = Control, 2 = Treatment

						Noncompliance			
		Hazard Rate		Proportion Lost		Proportion		Hazard Rate	
Time Period	Accrual Pattern	h1	h2	Lost1	Lost2	NCP1	NCP2	NCh1	NCh2
1	100% Accrual	1.4	0.8	0	0	0	0	1	1
2	100% Accrual	1.4	0.8	0	0	0	0	1	1
3	100% Accrual	1.4	0.8	0	0	0	0	1	1

.
.
.

References

Klein, J.P., Moeschberger, M.L. 1997. Survival Analysis. Springer-Verlag. New York.
Piantadosi, S. 2005. Clinical Trials, A Methodologic Perspective, 2nd Ed. John Wiley & Sons, Inc. New Jersey.

Sample sizes of 92 and 93 are needed to achieve 90% power for the Gehan-Wilcoxon test. The additional portion of the output is not shown here since it used to compare multiple scenarios. Multiple scenarios occur when more than one value is entered for one or more of the parameters on the design tabs. In this example there is only one scenario.

Example 2 – Validation using Lakatos (1988)

Lakatos (1988), pages 231-234, presents an example that will be used to validate this procedure. In this example, a two-year trial is investigated. All subjects begin the trial together, so there is no accrual period. The hazard rates are 1.0 and 0.5 for the control and treatment groups, respectively. The yearly loss to follow-up is 3% per year in both groups. Noncompliance and drop-in rates are assumed to be 4% and 5% per year, respectively. The power is set to 90%. A two-sided Logrank test with alpha set to 0.05 is assumed. Equal allocation of the sample to both control and experiment groups is used. Lakatos obtains a total sample size of 139.

For reproducibility, we'll use a random seed of 5979259.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design 1 Tab

Solve For	**Sample Size**
Test Type	**Logrank**
Alternative Hypothesis	**H1: Hazard1 ≠ Hazard2**
Simulations	**10000**
Random Seed	**5979259** (for Reproducibility)
Power	**0.9**
Alpha	**0.05**
Group Allocation	**Equal (N1 = N2)**
Input Type	**Hazard Rate**
h1 (Hazard Rate of Control Group)	**1.0**
Treatment Group Parameter	**h2 (Hazard Rate)**
h2 (Hazard Rate of Treatment Group)	**0.5**

Design 2 Tab

Controls Lost	**0.03**
Treatments Lost	**0.03**
For Treatments Lost under H0, use	**Controls Lost**
Noncompliance Proportion (Control)	**0.05**
NCh1 (Noncompliance Hazard, Control)	**0.5**
At time of noncompliance, start NCh1 at	**Current time**
Noncompliance Proportion (Treatment)	**0.04**
NCh2 (Noncompliance Hazard, Treatment)	**1.0**
At time of noncompliance, start NCh2 at	**Current time**
For Treatment Noncomp. under H0, use	**Control Noncompliance**
Accrual Time (Integers Only)	**0**
Accrual Pattern	**Equal (Uniform)**
Total Time (Integers Only)	**2**

Output

Click the Calculate button to perform the calculations and generate the following output. The calculations will take a few moments to complete. The results will vary slightly due to simulation differences.

Numeric Results (Scenario 1)

Solve For:	Sample Size
Groups:	1 = Control, 2 = Treatment
Hypotheses:	H0: Hazard1 = Hazard2 vs. H1: Hazard1 ≠ Hazard2
Test Statistic:	Logrank Test
Simulations:	10000
Pool Size:	20000
Random Seed:	5979259

Power			Alpha				
	95% C.I. Limits		Value		95% C.I. Limits		
Value	Lower	Upper	Target	Actual	Lower	Upper	Beta
0.906	0.9	0.912	0.05	0.053	0.048	0.057	0.094

Sample Size		Hazard Ratio	Hazard Rate		Accrual	Time	
N1	N2	HR	h1	h2	Pattern	Accrual	Total
69	70	0.5	1	0.5	Equal	0	2

		Noncompliance*			
Proportion Lost*		Proportion		Hazard Rate	
Lost1	Lost2	NCP1	NCP2	NCh1	NCh2
0.03	0.03	0.05	0.04	0.5	1

* The reported values are during a single time period.

Whole Study Averages (Scenario 1)

Cumulative Subject Time				Events			
H0		H1		H0		H1	
G1	G2	G1	G2	G1	G2	G1	G2
59.2	59.9	59	85.6	57.7	58.6	57.8	43.9

Detailed Input (Scenario 1)

Groups: 1 = Control, 2 = Treatment

Time Period	Accrual Pattern	Hazard Rate		Proportion Lost		Noncompliance			
						Proportion		Hazard Rate	
		h1	h2	Lost1	Lost2	NCP1	NCP2	NCh1	NCh2
1	100% Accrual	1	0.5	0.03	0.03	0.05	0.04	0.5	1
2	100% Accrual	1	0.5	0.03	0.03	0.05	0.04	0.5	1

The total sample size is 69 + 70 = 139, which matches the published result of Lakatos (1988). The total sample size of each run with a random seed may vary slightly due to simulation differences.