Chapter 343

Multi-Arm Tests for Treatment and Control Survival Curves using Cox's Proportional Hazards Model

Introduction

This module computes power and sample size for multiple comparisons of treatment survival curves versus a control survival curve based on the results in Machin, Campbell, Tan, and Tan (2018). In this design, there are *k* treatment groups and one control group. A survival curve is measured in each group. A total of *k* hypothesis tests are anticipated, each comparing a treatment group with the common control group using a simple z-test based on a Cox proportional hazards regression coefficient.

The formulation for testing the significance of a Cox regression coefficient is identical to the standard logrank test. Thus, the power and sample size formulas for one analysis also work for the other. The Cox Regression model has the added benefit over the exponential model that it does not assume that the hazard rates are constant, but only that they are proportional. That is, that the hazard ratio remains constant throughout the experiment, even if the hazard rates vary.

A Bonferroni adjustment of the type I error rate may be optionally made because several comparisons are being tested using the same data. Making a multiplicity adjustment is usually recommended, but not always. In fact, Saville (1990) advocates not applying it and Machin, Campbell, Tan, and Tan (2018) include omitting it as a possibility.

Background

Whether you want to test several doses of a single treatment or several types of treatments, good research practice requires that each treatment be compared with a control. For example, a popular three-arm design consists of three groups: control, treatment A, and treatment B. Two tests are run: treatment A versus control and treatment B versus the same control. This avoids having to obtain a second control group for treatment B. Besides the obvious efficiency in subjects, it may be easier to recruit subjects if their chances of receiving the new treatment are better than 50-50.

Technical Details

Cox's Proportional Hazards Regression

Cox's proportional hazards regression is widely used for survival data. The regression model is

$$h(t|z) = h(t|0)\exp(bz)$$

where

b is the regression coefficient which is equal to $\log[h(t|1)/h(t|0)] = \log(HR)$

z is a binary indicator variable of treatment group

t is elapsed time

h(*t*|*z*) is the hazard rate at time t, given covariate z

HR is the hazard ratio, $h(t|1)/h(t|0)$

The two-sided, statistical hypothesis testing survival equality is a test of whether *b* is zero. This hypothesis is stated as

$$H_0: b = 0 \quad \text{vs.} \quad H_1: b \neq 0$$

Test Statistic

It can be shown that the test of *b* based on the partial likelihood method of Cox (1972) coincides with the common logrank test statistic shown next.

Logrank Test

The logrank test statistic is

$$L = \frac{\sum_{k=1}^{K}\left(I_k - \frac{Y_{1i}}{Y_{1i}+Y_{2i}}\right)}{\left[\sum_{k=1}^{K}\left(\frac{Y_{1i}Y_{2i}}{(Y_{1i}+Y_{2i})^2}\right)\right]^{-1/2}}$$

where *K* is the number of deaths, Y_{ij} is the number of subjects at risk just prior to the j^{th} observed event in the i^{th} group, and I_k is a binary variable indicating whether the k^{th} event is from group 1 or not.

The distribution of *L* is approximately normal with mean $b\sqrt{P_1P_2dN}$ and unit variance, where

P_1 is the proportion of *N* that is in the control group

P_2 is the proportion of *N* that is in the treatment group

N is the total sample size

N_1 is the sample size from the control group, $N_1 = N(P_1)$

N_2 is the sample size from the treatment group, $N_2 = N(P_2)$

Pev_1 is probability of the event of interest in the control group

Pev_2 is probability of the event of interest in the treatment group

d is the overall probability of an event, $d = Pev_1P_1 + Pev_2P_2$

b is the Cox regression coefficient, $b = \log(HR)$

Power Calculations

The power of the statistical test of *b* is given by

$$\Phi\left(b\sqrt{P_1P_2dN} - z_{1-\alpha/2}\right)$$

or equivalently

$$\Phi\left(\log(HR_1)\sqrt{P_1P_2dN} - z_{1-\alpha/2}\right)$$

where HR_1 is the actual assumed value of the hazard ratio under the alternative hypothesis.

Testing Multiple Treatment Groups versus a Single Control Group

Suppose you have *k* treatment groups with samples of size N_i and one control group with a sample of size N_C. The total sample size is $N = N_1 + N_2 + \ldots + N_k + N_C$. The response for each subject is their survival time until they either exhibit the event of interest or they are censored from the study.

A Cox proportional hazards regression model is fit to the data in which one of the independent variables is a binary variable that is zero if the subject is from the control group or one if they are from the i^{th} treatment group. Suppose that the regression coefficient associated with this independent variable is called b_i. As pointed out above, it turns out that

$$b_i = \log(HR_i)$$

where HR_i is the hazard ratio comparing the treatment and control groups. If HR_i = 1, there is no difference between the groups.

The hypotheses for a two-sided statistical test are

$$H_0\!: b_i = 0 \quad \text{vs.} \quad H_1\!: b_i \neq 0$$

Similar results are available for the two one-sided tests.

Hence, the data may be analyzed using *k* separate regressions each producing a test of the hazard ratio comparing a treatment group to the common control group.

The power for each of the *k* tests can be computed using the formulas given above.

Multiplicity Adjustment

Because *k* z-tests between treatment groups and the control group are run when analyzing the results of this study, many statisticians recommend that the Bonferroni adjustment be applied. This adjustment is easy to apply: the value of alpha that is used in the test is found by dividing the original alpha by the number of tests. For example, if the original alpha is set at 0.05 and the number of treatment (not including the control) groups is five, the individual tests will be conducted using an alpha of 0.01.

The main criticism of this procedure is that if there are many tests, the value of alpha becomes very small. To mitigate against this complaint, some statisticians recommend separating the treatment groups into those that are of primary interest and those that are of secondary interest. The Bonferroni adjustment is made by the using the number of primary treatments rather than the total number of treatments.

There are some who advocate ignoring the adjustment entirely in the case of randomized clinical trials. See for example Saville (1990) and the discussion in chapter 14 of Machin, Campbell, Tan, and Tan (2018).

Size of the Control Group

Because the control group is used over and over, some advocate increasing the number of subjects in this group. The standard adjustment is to include $\sqrt{k}$ subjects in the control group for each subject in one of the treatment groups. See Machin, Campbell, Tan, and Tan (2018, pages 231-232). Note that often, the treatment groups all have the same size.

Example 1 – Finding the Sample Size

A parallel-group clinical trial is being designed to compare the survivability induced by three doses of a test compound against the standard (control) therapy in patients with a specific type of disease.

The proportion surviving one-year after the current treatment is 0.50 ($h_c = 0.693$). The researchers want to determine the sample size necessary to detect the situation when the proportion surviving one-year after the new treatment is 0.75 ($h_t = 0.288$). Hence, they want to compute the power when

$$HR = 0.288/0.693 \ = 0.4156$$

The researchers would like to study the influence of *HR* on the sample size, so they would like to look at a range of possible values for 0.3 to 0.5.

For planning purposes, they decide that the probability of an event is 0.50 in the control group and 0.25 in the three treatment groups. The researchers decide to use a two-sided test at the 0.05 significance level and a power of 0.8.

Following standard procedure, the control group multiplier will be set to $\sqrt{k} = \sqrt{3} = 1.732$ since the control group is used for three comparisons in this design.

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.

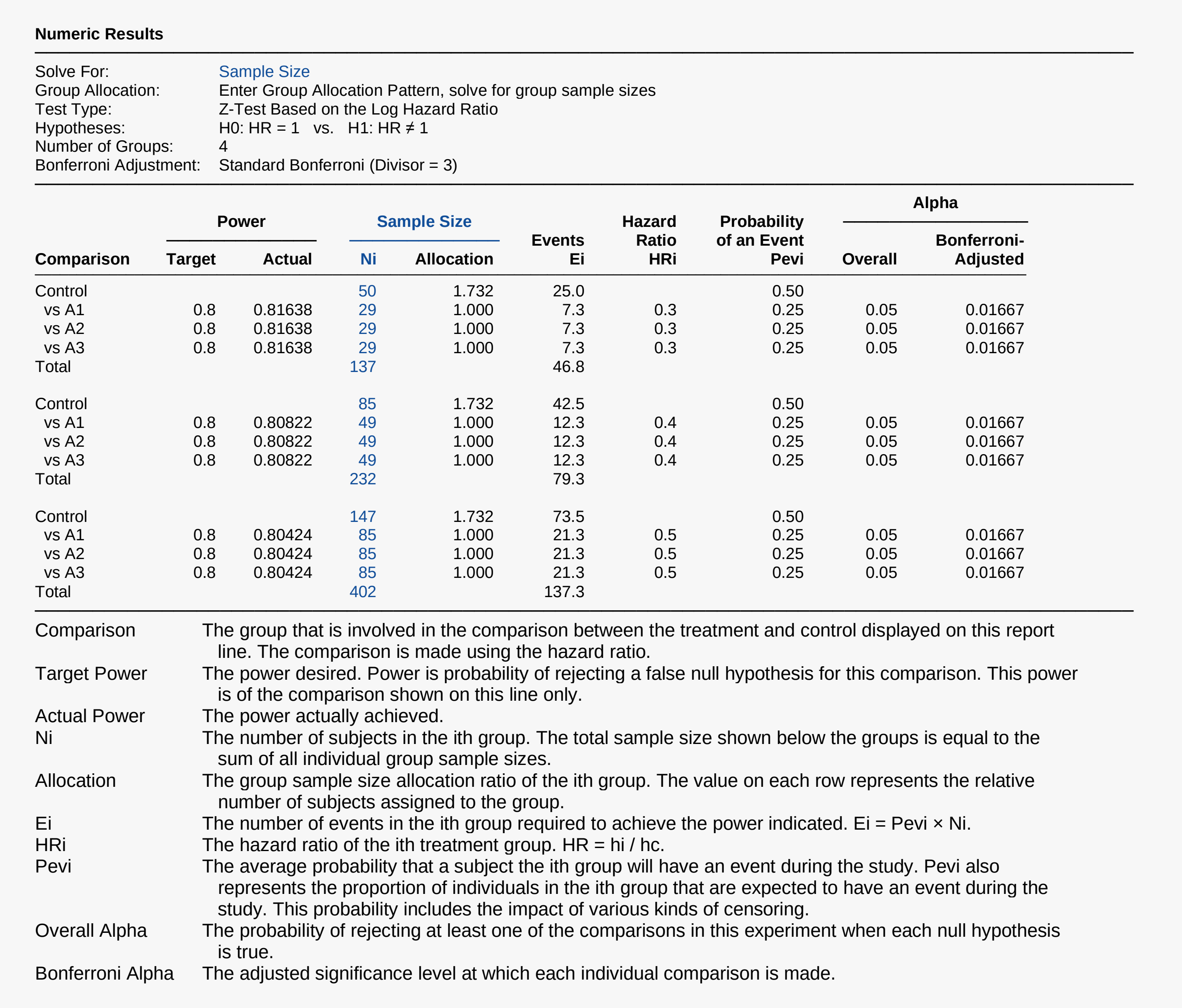
Design Tab

Solve For	**Sample Size**
Alternative Hypothesis	**Two-Sided (H1: HR ≠ 1)**
Power of Each Test	**0.80**
Overall Alpha	**0.05**
Bonferroni Adjustment	**Standard Bonferroni**
Group Allocation	**Enter Group Allocation Pattern, solve for group sample sizes**
Pev (Default Probability of an Event)	**0.5**
Control Probability of an Event	**0.5**
Control Sample Size Allocation	**1.732**
Set A Number of Groups	**3**
Set A Hazard Ratio	**0.3 0.4 0.5**
Set A Probability of an Event	**0.25**
Set A Sample Size Allocation	**1**
Set B Number of Groups	**0**
Set C Number of Groups	**0**
Set D Number of Groups	**0**
More	**Unchecked**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Reports



```
Numeric Results
─────────────────────────────────────────────────────────────────────────────────────────────────────────
Solve For:               Sample Size
Group Allocation:        Enter Group Allocation Pattern, solve for group sample sizes
Test Type:               Z-Test Based on the Log Hazard Ratio
Hypotheses:              H0: HR = 1   vs.   H1: HR ≠ 1
Number of Groups:        4
Bonferroni Adjustment:   Standard Bonferroni (Divisor = 3)
─────────────────────────────────────────────────────────────────────────────────────────────────────────
                                                                                           Alpha
                          Power            Sample Size                Hazard  Probability  ───────────────────────
                    ───────────────    ──────────────────    Events   Ratio   of an Event              Bonferroni-
Comparison          Target   Actual      Ni   Allocation         Ei     HRi          Pevi  Overall        Adjusted
─────────────────────────────────────────────────────────────────────────────────────────────
Control                                  50        1.732       25.0                  0.50
  vs A1                0.8  0.81638      29        1.000        7.3     0.3          0.25     0.05         0.01667
  vs A2                0.8  0.81638      29        1.000        7.3     0.3          0.25     0.05         0.01667
  vs A3                0.8  0.81638      29        1.000        7.3     0.3          0.25     0.05         0.01667
Total                                   137                    46.8

Control                                  85        1.732       42.5                  0.50
  vs A1                0.8  0.80822      49        1.000       12.3     0.4          0.25     0.05         0.01667
  vs A2                0.8  0.80822      49        1.000       12.3     0.4          0.25     0.05         0.01667
  vs A3                0.8  0.80822      49        1.000       12.3     0.4          0.25     0.05         0.01667
Total                                   232                    79.3

Control                                 147        1.732       73.5                  0.50
  vs A1                0.8  0.80424      85        1.000       21.3     0.5          0.25     0.05         0.01667
  vs A2                0.8  0.80424      85        1.000       21.3     0.5          0.25     0.05         0.01667
  vs A3                0.8  0.80424      85        1.000       21.3     0.5          0.25     0.05         0.01667
Total                                   402                   137.3
─────────────────────────────────────────────────────────────────────────────────────────────────────────
Comparison        The group that is involved in the comparison between the treatment and control displayed on this report
                    line. The comparison is made using the hazard ratio.
Target Power      The power desired. Power is probability of rejecting a false null hypothesis for this comparison. This power
                    is of the comparison shown on this line only.
Actual Power      The power actually achieved.
Ni                The number of subjects in the ith group. The total sample size shown below the groups is equal to the
                    sum of all individual group sample sizes.
Allocation        The group sample size allocation ratio of the ith group. The value on each row represents the relative
                    number of subjects assigned to the group.
Ei                The number of events in the ith group required to achieve the power indicated. Ei = Pevi × Ni.
HRi               The hazard ratio of the ith treatment group. HR = hi / hc.
Pevi              The average probability that a subject the ith group will have an event during the study. Pevi also
                    represents the proportion of individuals in the ith group that are expected to have an event during the
                    study. This probability includes the impact of various kinds of censoring.
Overall Alpha     The probability of rejecting at least one of the comparisons in this experiment when each null hypothesis
                    is true.
Bonferroni Alpha  The adjusted significance level at which each individual comparison is made.
```

Summary Statements

A parallel, 4-group design (with one control group and 3 treatment groups) will be used to test whether the hazard rate for each treatment group is different from the control group hazard rate (H0: HR = 1 versus H1: HR ≠ 1, HR = treatment hazard rate i / control hazard rate). The hypotheses will be evaluated using 3 two-sided, two-sample, Bonferroni-adjusted, Cox's proportion hazards regression term Z-tests, with an overall (experiment-wise) Type I error rate (α) of 0.05. (As a note, these Cox's proportional hazards regression term Z-tests are equivalent to the common logrank test.) It is anticipated that the proportions of subjects in each group that will have an event during the course of the study (beginning with the control group) will be 0.5, 0.25, 0.25, and 0.25. To detect the treatment to control hazard ratios 0.3, 0.3, and 0.3 with at least 80% power for each test, the control group sample size needed will be 50 and the number of needed subjects for the treatment groups will be 29, 29, and 29 (totaling 137 subjects overall). The corresponding total number of events is 46.8. These results assume that the hazard ratios are constant throughout the study.

References

Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. 2018. Sample Size Calculations in Clinical Research, 3rd Edition. Chapman & Hall/CRC. Boca Raton, FL. Pages 86-88.

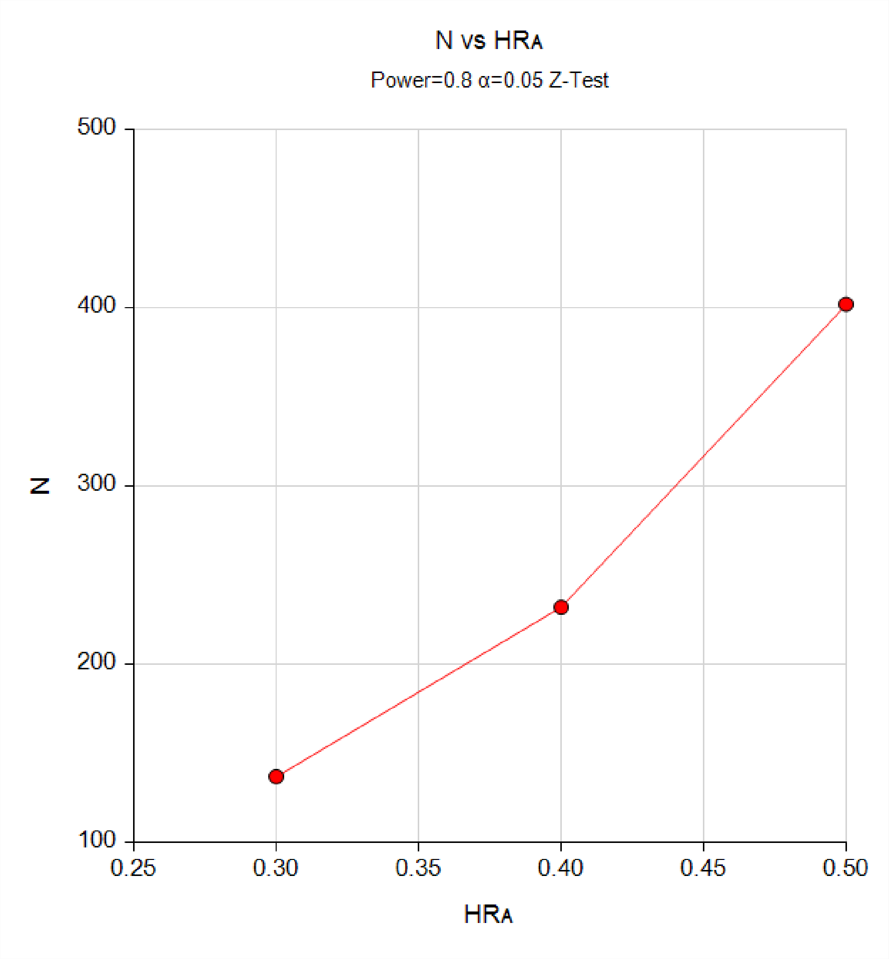
Machin, D., Campbell, M.J., Tan, S.B, and Tan, S.H. 2018. Sample Sizes for Clinical, Laboratory, and Epidemiology Studies, 4th Edition. Wiley Blackwell.

Schoenfeld, David A. 1983. 'Sample Size Formula for the Proportional-Hazards Regression Model', Biometrics, Volume 39, Pages 499-503.

This report shows the numeric results of this power study. Notice that the results are shown in blocks of three rows at a time. Each block represents a single design.

Plots Section

Plots



This plot gives a visual presentation to the results in the Numeric Report. We can quickly see the impact on the sample size of changing the hazard ratio from 0.3 to 0.5.

Example 2 – Validation using a Previously Validated Procedure

We could not find a validation result in the statistical literature, so we will use a previously validated **PASS** procedure (**Tests for Two Survival Curves Using Cox's Proportional Hazards Model**) to produce the results for the following example.

A parallel-group clinical trial is being designed to compare the survivability induced by three doses of a test compound against the standard (control) therapy in patients with a specific type of disease.

The proportion surviving one-year after the current treatment is 0.50 ($h_c = 0.693$). The researchers want to determine the sample size necessary to detect the situation when the proportion surviving one-year after the new treatment is 0.75 ($h_t = 0.288$). Hence, they want to compute the power when

$$HR = 0.288/0.693 \; = 0.4156$$

For planning purposes, they decide that the probability of an event is 0.50 in the control group and 0.25 in the three treatment groups. The researcher decides to use a two-sided test at the 0.05 significance level and a power of 0.8. Since a Bonferroni adjustment is made, the significance level is reduced to 0.05 / 3 = 0.01667.

The sample sizes of all groups will be equal.

The **Tests for Two Survival Curves Using Cox's Proportional Hazards Model** procedure is set up as follows.

```
Design Tab
_____________________________________________________________
Solve For .......................................Sample Size
Alternative Hypothesis ..........................Ha: HR ≠ 1
Power............................................0.8
Alpha............................................0.01667 (which is Alpha / k)
Group Allocation ................................Equal (N1 = N2)
Pev1 (Probability of a Control Event)............0.5
Pev2 (Probability of a Treatment Event) .........0.25
HR1 (Actual Hazard Ratio to Detect) .............0.4156
```

This set of options generates the following report.

```
Numeric Results
──────────────────────────────────────────────────────────────────────────────────────────
Solve For:     Sample Size
Groups:        1 = Control, 2 = Treatment
Hypotheses:    H0: HR = 1   vs.   Ha: HR ≠ 1
──────────────────────────────────────────────────────────────────────────────────────────
                                           Number of                    Probability
              Sample Size      Percent       Events         Hazard      of an Event
           ────────────────    Group 1  ──────────────────   Ratio    ──────────────
Power       N     N1     N2        %N1     E     E1     E2     HR1    Pev1    Pev2    Alpha
───────────────────────────────────────────────────────────────────────────────────────
0.80359   146     73     73         50  54.8   36.5   18.3  0.4156     0.5    0.25  0.01667
──────────────────────────────────────────────────────────────────────────────────────────
```

In order to maintain a power of 80% for all three groups, it is apparent that the groups will all need to have a sample size of 73 per group. This table contains the validation values. We will now run these values through the current procedure and compare the results with these values.

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 Tab
Solve For .....................................................Sample Size
Alternative Hypothesis ..................................Two-Sided (H1: HR ≠ 1)
Power of Each Test ......................................0.80
Overall Alpha ................................................0.05
Bonferroni Adjustment ..................................Standard Bonferroni
Group Allocation ...........................................Equal (Nc = N1 = N2 = ...)
Pev (Default Probability of an Event) ............0.5
Control Probability of an Event ......................0.5
Set A Number of Groups................................3
Set A Hazard Ratio .......................................0.4156
Set A Probability of an Event .........................0.25
Set B Number of Groups................................0
Set C Number of Groups ...............................0
Set D Number of Groups ...............................0
More.............................................................Unchecked
```

Output

Click the Calculate button to perform the calculations and generate the following output.

```
Numeric Results
──────────────────────────────────────────────────────────────────────────────────────────
Solve For:               Sample Size
Group Allocation:        Equal (Nc = N1 = N2 = ...)
Test Type:               Z-Test Based on the Log Hazard Ratio
Hypotheses:              H0: HR = 1   vs.   H1: HR ≠ 1
Number of Groups:        4
Bonferroni Adjustment:   Standard Bonferroni (Divisor = 3)
──────────────────────────────────────────────────────────────────────────────────────────
                                                                                Alpha
                          Power            Sample             Hazard  Probability  ──────────────────────
                    ──────────────────       Size   Events     Ratio  of an Event            Bonferroni-
Comparison          Target      Actual         Ni       Ei       HRi         Pevi  Overall       Adjusted
──────────────────────────────────────────────────────────────────────────────────────────
Control                                        73     36.5                   0.50
  vs A1                0.8     0.80357         73     18.3    0.4156         0.25     0.05        0.01667
  vs A2                0.8     0.80357         73     18.3    0.4156         0.25     0.05        0.01667
  vs A3                0.8     0.80357         73     18.3    0.4156         0.25     0.05        0.01667
Total                                         292     91.3
──────────────────────────────────────────────────────────────────────────────────────────
```

As you can see, the sample sizes are all 73. This matches the sample size found in the validation run above. The procedure is validated.