

Chapter 349

Multi-Arm Non-Inferiority Tests for Vaccine Efficacy using Treatment vs. Control Hazard Ratios (Cox's Proportional Hazards Model)

Introduction

This module computes power and sample size for multiple non-inferiority tests for vaccine efficacy (VE) of treatment versus control survival curves (or logrank tests). VE is a traditional index of the protective efficacy of a vaccine. It is calculated as

$$VE_i = 1 - \frac{h_i}{h_c} = 1 - HR_i$$

where h_i and h_c are *hazard rates* of the disease being studied among those vaccinated and those not vaccinated. Hence, an analysis of vaccine effectiveness reduces to an analysis of the hazard ratio.

This multiple comparison procedure is 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 non-inferiority 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.

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.

Relative Vaccine Efficacy

Often, the goal of the study is to show that the attack rate of a new vaccine is no worse than that of the current standard vaccine. For example, the standard vaccine might have serious side effects, be expensive to produce, etc. In this case, the trial is conducted to show that the new vaccine is an attractive replacement for the standard vaccine. In this case, the control group does not receive a placebo. Rather, it receives the standard vaccine. In this case, the quantity of interest is called the *relative vaccine efficacy* (*rVE*). It is calculated as

$$rVE_i = 1 - \frac{h_i}{h_C} = 1 - HR_i$$

Technical Details

Suppose you have k treatment groups with hazard ratios HR_i of size N_i and one control group with size N_C . The total sample size is $N = N_1 + N_2 + \dots + N_k + N_C$.

Assuming that lower hazard rates are better, non-inferiority means that the treatment hazard rate is at most, only slightly higher than the control hazard rate. We may find it more convenient to state the hypotheses in terms of the hazard ratio, HR , rather than the Cox regression coefficient, b . Remembering that $b = \log(HR)$ and assuming that $HR_0 > 1$, non-inferiority requires that $HR < HR_0$. Here, HR_0 is the boundary of clinical insignificance or the non-inferiority boundary.

The statistical hypotheses that result in the conclusion of non-inferiority when the null hypothesis is rejected is

$$H_0: \log(HR) \geq \log(HR_0) \quad \text{vs.} \quad H_a: \log(HR) < \log(HR_0)$$

The k one-sided non-inferiority tests are

$$H_{0i}: VE_i \leq VE_0 \quad \text{vs.} \quad H_{1i}: VE_i > VE_0 \quad \text{for } i = 1, 2, \dots, k$$

Note that if lower hazard rates are better, as is usually the case when the studying disease prevention, $VE_0 < 1$.

For convenience, these hypotheses are collectively referred to as

$$H_0: VE \leq VE_0 \quad \text{vs.} \quad H_1: VE > VE_0$$

These hypotheses may be restated in terms of hazard ratios as

$$H_0: HR \geq HR_0 \quad \text{vs.} \quad H_1: HR < HR_0$$

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, $h(t|1)/h(t|0)$

Logrank Test

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.

The logrank test statistic is

$$L = \frac{\sum_{k=1}^K \left(I_k - \frac{Y_{1i}HR_0}{Y_{1i}HR_0 + Y_{2i}} \right)}{\left[\sum_{k=1}^K \left(\frac{Y_{1i}Y_{2i}HR_0}{(Y_{1i}HR_0 + Y_{2i})^2} \right) \right]^{-\frac{1}{2}}}$$

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 $(\log(HR) - \log(HR_0))\sqrt{P_1P_2dN}$ 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, $N_2 = N(P_2)$

Pe_{v1} is probability of the event of interest in the control group

Pe_{v2} is probability of the event of interest in the treatment group

d is the overall probability of an event, $d = Pe_{v1}P_1 + Pe_{v2}P_2$

HR is the observed hazard ratio

HR_0 is the non-inferiority boundary (limit) of the hazard ratio

Power Calculations

The power of this test is given by

$$\Phi \left((\log(HR_1) - \log(HR_0)) \sqrt{P_1 P_2 dN} - z_{1-\alpha} \right)$$

where HR_1 is the actual 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 $N = N_1 + N_2 + \dots + N_k + N_C$. 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 data may be analyzed using k separate regressions each producing a non-inferiority 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 formula 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, non-inferiority trial is being designed to compare the survivability associated with three doses of a test compound against the standard (control) therapy in patients with a specific type of disease.

Suppose the non-inferiority hazard ratio is set at 1.25. Hence its non-inferiority vaccine efficacy is -0.25. The researchers want to determine the sample size necessary to detect the situation when the vaccine efficacy of any of the new treatment increases to 0.0. This equates to a hazard ratio of 1.0 for all three treatments.

The researchers would like to study the influence of the value of *HR* on the sample size, so they would like to look at a range of possible values: 0.9 to 1. For planning purposes, they decide that the probability of an event is 0.75 in the control group and 0.5 in the three treatment groups. The researchers decide to use a 0.025 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
Power of Each Test	0.80
Overall Alpha	0.025
Bonferroni Adjustment	Standard Bonferroni
Group Allocation	Enter Group Allocation Pattern, solve for group sample sizes
Pev (Default Probability of an Event)	0.75
HR0 (Non-Inferiority Hazard Ratio)	1.25
Control Probability of an Event	0.75
Control Sample Size Allocation.....	1.732
Set A Number of Groups.....	3
Set A Hazard Ratio	0.9 0.95 1
Set A Probability of an Event	0.5
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 Cox Proportional Hazards Regression
 Hypotheses: H0: VE ≤ VE0 vs. H1: VE > VE0 or H0: HR ≥ HR0 vs. H1: HR < HR0
 Number of Groups: 4
 Bonferroni Adjustment: Standard Bonferroni (Divisor = 3)

Comparison	Power		Sample Size		Events Ei	Vaccine Efficacy		Hazard Ratio		Probability of an Event Pevi	Alpha	
	Target	Actual	Ni	Allocation		NI Limit VE0	Actual VEi	NI Limit HR0	Actual HRi		Overall	Bonferroni- Adjusted
Control			404	1.732	303.0					0.75		
vs A1	0.8	0.80142	233	1.000	116.5	-0.25	0.10	1.25	0.90	0.50	0.025	0.00833
vs A2	0.8	0.80142	233	1.000	116.5	-0.25	0.10	1.25	0.90	0.50	0.025	0.00833
vs A3	0.8	0.80142	233	1.000	116.5	-0.25	0.10	1.25	0.90	0.50	0.025	0.00833
Total			1103		652.5							
Control			577	1.732	432.8					0.75		
vs A1	0.8	0.80013	333	1.000	166.5	-0.25	0.05	1.25	0.95	0.50	0.025	0.00833
vs A2	0.8	0.80013	333	1.000	166.5	-0.25	0.05	1.25	0.95	0.50	0.025	0.00833
vs A3	0.8	0.80013	333	1.000	166.5	-0.25	0.05	1.25	0.95	0.50	0.025	0.00833
Total			1576		932.3							
Control			873	1.732	654.8					0.75		
vs A1	0.8	0.80034	504	1.000	252.0	-0.25	0.00	1.25	1.00	0.50	0.025	0.00833
vs A2	0.8	0.80034	504	1.000	252.0	-0.25	0.00	1.25	1.00	0.50	0.025	0.00833
vs A3	0.8	0.80034	504	1.000	252.0	-0.25	0.00	1.25	1.00	0.50	0.025	0.00833
Total			2385		1410.8							

- 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: Sample Size. The number of subjects in the ith group. The total sample size, N, is shown as the last row of the column.
- Allocation: The group sample size allocation pattern. 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. $Ei = Pevi \times Ni$.
- VE0: The non-inferiority bound on the vaccine efficacy determines whether to conclude that a treatment is non-inferior or inferior to the control group. Note that $VE0 = 1 - HR0$.
- VEi: The vaccine efficacy of the ith group at which the power is calculated. The formula is $VEi = 1 - HRi$.
- HR0: The non-inferiority hazard ratio boundary used to declare whether a treatment is non-inferior to the control.
- 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 (h_i) is non-inferior to the control group hazard rate (h_c), with a non-inferiority hazard ratio limit of 1.25 ($H_0: HR \geq 1.25$ versus $H_1: HR < 1.25$, $HR = h_i / h_c$), or, equivalently, whether the vaccine efficacy for each treatment group ($VE_i = 1 - h_i / h_c$) is non-inferior, with a non-inferiority vaccine efficacy limit of -0.25 ($H_0: VE \leq -0.25$ versus $H_1: VE > -0.25$). In this study, lower hazard rates are considered to be better. The non-inferiority hypotheses will be evaluated using 3 one-sided, two-sample, Bonferroni-adjusted, Cox's proportional hazards regression term Z-tests, with an overall (experiment-wise) Type I error rate (α) of 0.025. 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 0.75, 0.5, 0.5, and 0.5. To detect the treatment to control hazard ratios 0.9, 0.9, and 0.9 with at least 80% power for each test, the control group sample size needed will be 404 and the number of needed subjects for the treatment groups will be 233, 233, and 233 (totaling 1103 subjects overall). The corresponding total number of events is 652.5. The treatment vaccine efficacies that correspond to the detectable hazard ratios 0.9, 0.9, and 0.9 are 0.1, 0.1, and 0.1. These results assume that the hazard ratios are constant throughout the study.

Dropout-Inflated Sample Size

Group	Dropout Rate	Sample Size Ni	Dropout-Inflated Enrollment Sample Size Ni'	Expected Number of Dropouts Di
1	20%	404	505	101
2	20%	233	292	59
3	20%	233	292	59
4	20%	233	292	59
Total		1103	1381	278
1	20%	577	722	145
2	20%	333	417	84
3	20%	333	417	84
4	20%	333	417	84
Total		1576	1973	397
1	20%	873	1092	219
2	20%	504	630	126
3	20%	504	630	126
4	20%	504	630	126
Total		2385	2982	597

- Group Lists the group numbers.
- Dropout Rate The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
- Ni The evaluable sample size for each group at which power is computed (as entered by the user). If Ni subjects are evaluated out of the Ni' subjects that are enrolled in the study, the design will achieve the stated power.
- Ni' The number of subjects that should be enrolled in each group in order to obtain Ni evaluable subjects, based on the assumed dropout rate. Ni' is calculated by inflating Ni using the formula $Ni' = Ni / (1 - DR)$, with Ni' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lohknygina, Y. (2018) pages 32-33.)
- Di The expected number of dropouts in each group. $Di = Ni' - Ni$.

Dropout Summary Statements

Anticipating a 20% dropout rate, group sizes of 505, 292, 292, and 292 subjects should be enrolled to obtain final group sample sizes of 404, 233, 233, and 233 subjects.

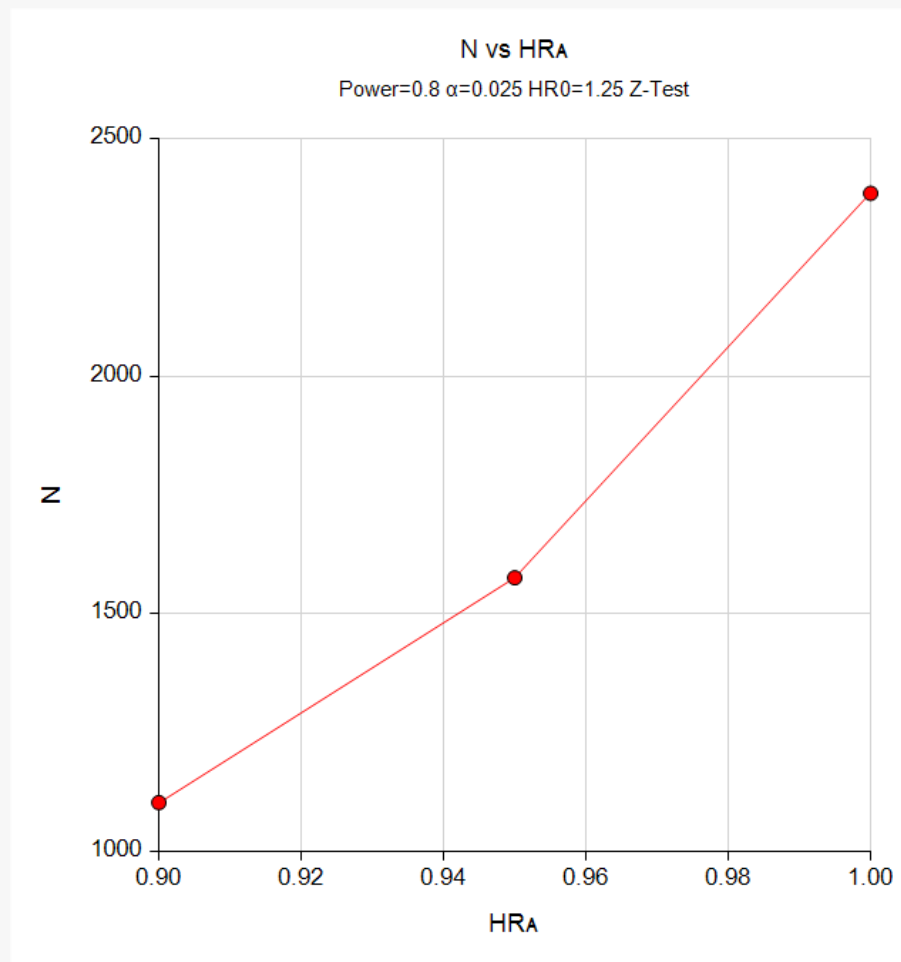
References

Chow, S.C., Shao, J., Wang, H., and Lohknygina, 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.
 Nauta, Jozef. 2020. Statistics in Clinical and Observational Vaccine Studies, 2nd Edition. Springer. Cham, Switzerland.

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 0.9 to 1.0.

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 (**Non-Inferiority Tests for Vaccine Efficacy Using the Hazard Ratio (Cox's Proportional Hazards Model)**) to produce the results for the following example.

Suppose the non-inferiority hazard ratio is set at 1.25. Hence its non-inferiority vaccine efficacy is -0.25. The researchers want to determine the sample size necessary to detect the situation when the vaccine efficacy of any of the new treatment increases to 0.0. This equates to a hazard ratio of 1.0 for all three treatments.

They decide that the probability of an event is 0.75 in the control group and 0.5 in the three treatment groups. The researchers decide to use a 0.00833 (0.025 / 3) significance level and a power of 0.8.

The sample sizes of all groups will be equal.

The **Non-Inferiority Tests for Vaccine Efficacy Using the Hazard Ratio (Cox's Proportional Hazards Model)** procedure is set up as follows

Design Tab

Solve For **Sample Size**
 Higher Hazards Are **Worse (H1: VE > VE0 or HR < HR0)**
 Power..... **0.8**
 Alpha..... **0.00833** (which is Alpha / k)
 Group Allocation **Equal (N1 = N2)**
 Pev1 (Event Probability in Group 1)..... **0.75**
 Pev2 (Event Probability in Group 2)..... **0.5**
 Vaccine Efficacy Input Type..... **Hazard Ratios**
 HR1 (Actual Hazard Ratio) **1**
 HR0 (Non-Inferiority Hazard Ratio)..... **1.25**

This set of options generates the following report.

Numeric Results

Solve For: [Sample Size](#)
 Higher Hazards Are: Worse
 Hypotheses: H0: VE ≥ VE0 vs. H1: VE < VE0

Power	Sample Size			Events		Vaccine Efficacy		Hazard Ratio		Event Prob		Alpha
	Cntl N1	Trt N2	Total N	Cntl E1	Trt E2	Act VE1	N.I. VE0	Act HR1	N.I. HR0	Cntl Pev1	Trt Pev2	
0.80009	673	673	1346	504.8	336.5	0	-0.25	1	1.25	0.75	0.5	0.00833

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 673 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
Power of Each Test	0.80
Overall Alpha	0.025
Bonferroni Adjustment	Standard Bonferroni
Group Allocation	Equal (Nc = N1 = N2 = ...)
Pev (Default Probability of an Event)	0.75
HR0 (Non-Inferiority Hazard Ratio)	1.25
Control Probability of an Event	0.75
Set A Number of Groups.....	3
Set A Hazard Ratio	1
Set A Probability of an Event	0.5
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 Cox Proportional Hazards Regression										
Hypotheses:		H0: VE ≤ VE0 vs. H1: VE > VE0 or H0: HR ≥ HR0 vs. H1: HR < HR0										
Number of Groups:		4										
Bonferroni Adjustment:		Standard Bonferroni (Divisor = 3)										
Comparison	Power		Sample Size Ni	Events Ei	Vaccine Efficacy		Hazard Ratio		Probability of an Event Pevi	Alpha		
	Target	Actual			NI Limit VE0	Actual VEi	NI Limit HR0	Actual HRi		Overall	Bonferroni-Adjusted	
Control			673	504.8					0.75			
vs A1	0.8	0.80013	673	336.5	-0.25	0	1.25	1	0.50	0.025	0.00833	
vs A2	0.8	0.80013	673	336.5	-0.25	0	1.25	1	0.50	0.025	0.00833	
vs A3	0.8	0.80013	673	336.5	-0.25	0	1.25	1	0.50	0.025	0.00833	
Total			2692	1514.3								

As you can see, the sample sizes are all 673. This matches the sample size found in the validation run above. The procedure is validated.