Sample Size Determination with SAS® Studio

William Coar, Axio Research

ABSTRACT

Experiments are designed to answer specific questions. In order for the results to have a reasonable amount of certainty and statistical validity, a sufficient number of observations are required. To achieve this, the concepts of type 1 error, power, and sample size are introduced in to the experimental design.

Real world constraints such as budget and feasibility are equally as important. Thus, statisticians often re-evaluate sample size under varying sets of assumptions, sometimes on the fly. SAS® Studio provides a variety of tasks associated with determining sample sizes using a point-and-click approach to enter assumptions, yet it also provides the ability to save the underlying SAS code so that it can be either refined or used at a later time.

The purpose of this Hands-on-Workshop is to introduce some of the features of SAS Studio for sample size determination. A number of examples will be introduced, including tests associated with proportions, means, and survival analysis. Each exercise will start with a research question, proposed methodology, and list of requirements needed for estimating the sample size. The attendees will then have the opportunity to work through the exercise using SAS Studio, and allow for discussion of adaptations that may be necessary for its use in their everyday programming environment.

INTRODUCTION

The ICH Guidelines for Good Clinical Practice discuss and define standards for the development and registration of investigational products. Section 2.5 indicates that *clinical trials should be scientifically sound, and described in a clear, detailed protocol*. The protocol is a document that describes the objective(s), design, methodology, statistical considerations, and organization of a trial [1]. But what does it mean to be scientifically sound as it relates to statistical considerations? While a complete answer would require a lengthy discussion, we focus on statistical considerations as they relate to the number of participants, or sample size, of a designed experiment.

The Statistical Principals for Clinical Trials (ICH E9) provide more direct guidance associated with sample size determination in clinical trials [2]. Section 3.5 of E9 states "*The number of subjects in a clinical trial should always be large enough to provide a reliable answer to the questions addressed.*" This statement is so simple and true, yet the process of determining a sample size is often far more complex. There are typically multiple questions to be addressed, and "large enough" introduces the reality of finite resources (time, money, patients, etc.).

This hands-on workshop is not specific to clinical trials though the examples are from a clinical trial setting. A clinical trial is a designed experiment. The general principles behind determining a sample size for a clinical trial extrapolate to any well designed experiment. Any well designed experiment should have enough observations to provide reliable answers.

HYPOTHESIS TESTING

Experiments are designed to answer research questions of interest, or hypotheses. To help answer the research question, the experiment must provide consideration for:

- Null and Alternative hypotheses
- A primary variable
- A test statistic
- Type 1 error (probability of erroneously rejecting the null hypothesis)
- Type 2 error (probability of erroneously failing to reject the null hypothesis).

Note that we often refer to statistical power rather than Type 2 error. Power is defined to be 1-(Type 2 error), and is typically expressed as a percentage. For a given alternative hypothesis, power is the probability that the null hypothesis is rejected when that alternative hypothesis is in fact true. Methods for sample size determination should be consistent with statistical methods used in the planned analysis, or justification should be provided when this is not the case. Critical assumptions will be made with respect to effect size and variance.

SAS STUDIO NAVIGATION

In SAS Studio, select **Task and Utilities-> Power and Sample Size**.

EXAMPLE 1: ONE SAMPLE PROPORTION

A new formulation of a therapeutic intervention is being considered. To be considered consistent with the existing formulation, the proportion of responders needs to be at least 25%. Historical data suggests the true response ranges from 30%-40%.

The research hypothesis is that the new formation has a success rate of at least 25%. The statistical hypotheses are:

$$H_o: \pi_T \leq 0.25$$
$$H_a: \pi_T > 0.25$$

where π_T is the true unknown proportion of responders. Since interest is only in values greater than 0.25, the hypothesis is one-sided (upper tailed test). It is anticipated that an approximate z-test can be used for the analysis. For calculation the variance, the observed proportion will be used rather than the hypothesized value of 0.25.

A sample size is needed for an experiment to show the response rate is at least 25%, with a one sided alpha=0.025, power=0.80.

Table 1: Design parameters for one sample proportion

Parameter	Value
Test for:	Equality
Solve for:	Total Sample Size
Test:	Approximate Z
Number of sides	One (upper)
Variance	Based on sample
Proportion values	.30 .35 .40
Null proportion	.25
Alpha	.05
Power	.80 .90
Plots:	Sample size by effect size: Min=0.30 Max=0.40

In SAS Studio, select **Task and Utilities-> Power and Sample Size->Tests for Proportions.**

In the center window you will see tabs for Tests for Proportions. Within this, you will see tabs for: **Settings**, **Code/Results**, **Split**.

- Click on the Settings tab to enter our design inputs.

- The Code/Results tab is used to see the underlying Proc Power code as well as the results.

- The Split tab can be selected to view the Settings and Code/Results tab simultaneously.

Within the Settings tab, you will find tabs for **Properties**, **Plots**, and **Information**.

- The Properties tab is used to enter design assumptions.

- The Plot tab can be selected to indicate specific plots of interest.

- The Information tab gives an overview of the Task selected.

After entering information from Table 1 into the Properties and Plots tab, we run the procedure. As good programmers should, we first check the log for various errors, warnings, or notes. In the absence of concerning messages, the Results tab is selected to view the following results:

Fixed Scenario Elements	
Method	Normal approximation
Variance Estimate	Sample Varianc
Number of Sides	U
Null Proportion	0.25
Alpha	0.05

Computed N Total				
Index	Proportion	Nominal Power	Actual Power	N Total
1	0.30	0.8	0.800	520
2	0.30	0.9	0.900	720
3	0.35	0.8	0.801	141
4	0.35	0.9	0.900	195
5	0.40	0.8	0.800	66
6	0.40	0.9	0.902	92

A quick look at the table of Computed N Total shows that sample sizes decrease as the true proportion increases. This makes sense in that we need fewer patients as the true proportion of responders increases in comparison to the threshold of 0.25. Additionally, an increased sample size is needed to achieve higher power.
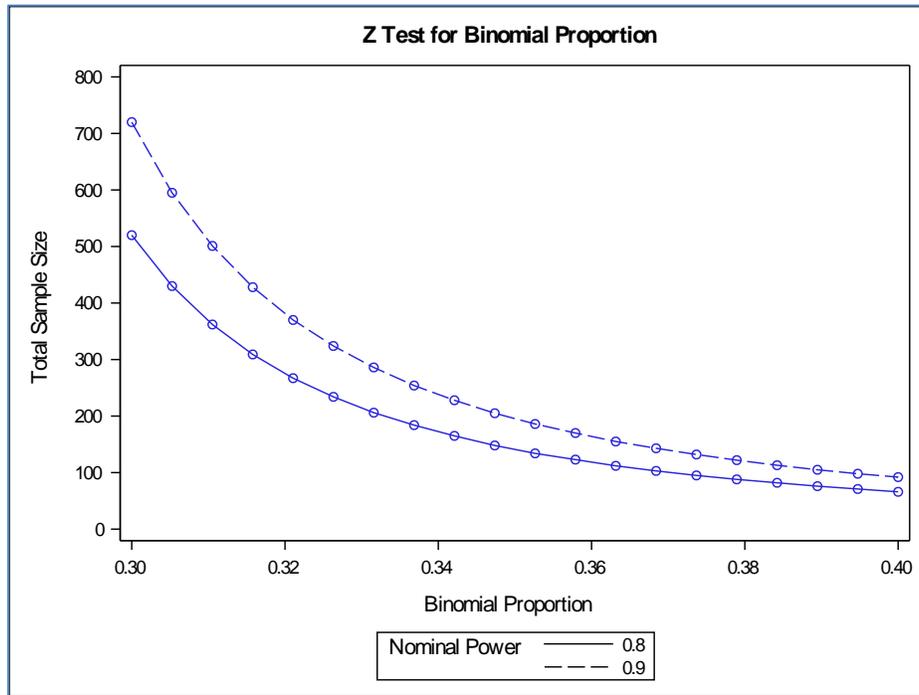
The Results tab provides a summary of the input design parameters and the sample sizes at the requested sets of assumptions. We further see from the plot that we can easily obtain sample sizes for other assumed proportions, not just the 0.30, 0.35, and 0.40 as originally requested.

Under the assumption that the true proportion of responses is around 0.35 (or 35%), then 141 patients are needed to achieve approximately 80% power using apha=0.05 (see Index=3 in the table above).

A sample size of 141 is selected. With this information, the study team wishes to understand the power associated with showing there is no more than a 10% increase in adverse reactions with the new formulation. Information from doctors suggests an increase may be tolerable as long as it does not exceed 10%.

What is the power to show there is no more than a 10% increase in event rates, assuming the true event rate is 78%? This translates into statistical hypotheses of:

$$H_o: \pi_T \geq 0.88$$
$$H_a: \pi_T < 0.88$$

In this case, interest is in showing the event rate for the new formulation is less than 0.88, suggesting a one-sided lower tailed test is appropriate.

Table 2: Updated design parameters for one sample proportion

| Parameter	Value
Test for:	Equality
Solve for:	Power
Test:	Approximate Z
Number of sides	One sided (lower)
Variance	Based on sample
Proportion values	.78
Null proportion	.88
Alpha	.05
Total Sample Size	141
Plots:	Power by effect size: Min=0.70 Max=0.90

With n=141, we find there is almost 90% power to show the event rate does not exceed 88%.

Fixed Scenario Elements	
Method	Normal approximation
Variance Estimate	Sample Variance
Number of Sides	L
Null Proportion	0.88
Alpha	0.05
Binomial Proportion	0.78
Total Sample Size	141

Computed Power
Power
0.889

The requested plot allows us to evaluate the power at proportion values other than 0.78. We see the power decreases as the assumed proportion (true event rate) gets closer to the assumption in the null hypothesis.

In reviewing the log, there is a warning that the effect is in the opposite direction of the alternative hypothesis. This is because our alternative hypothesis suggests the proportion is 0.88, yet we specified a max=0.90 for plotting.

Z Test for Binomial Proportion
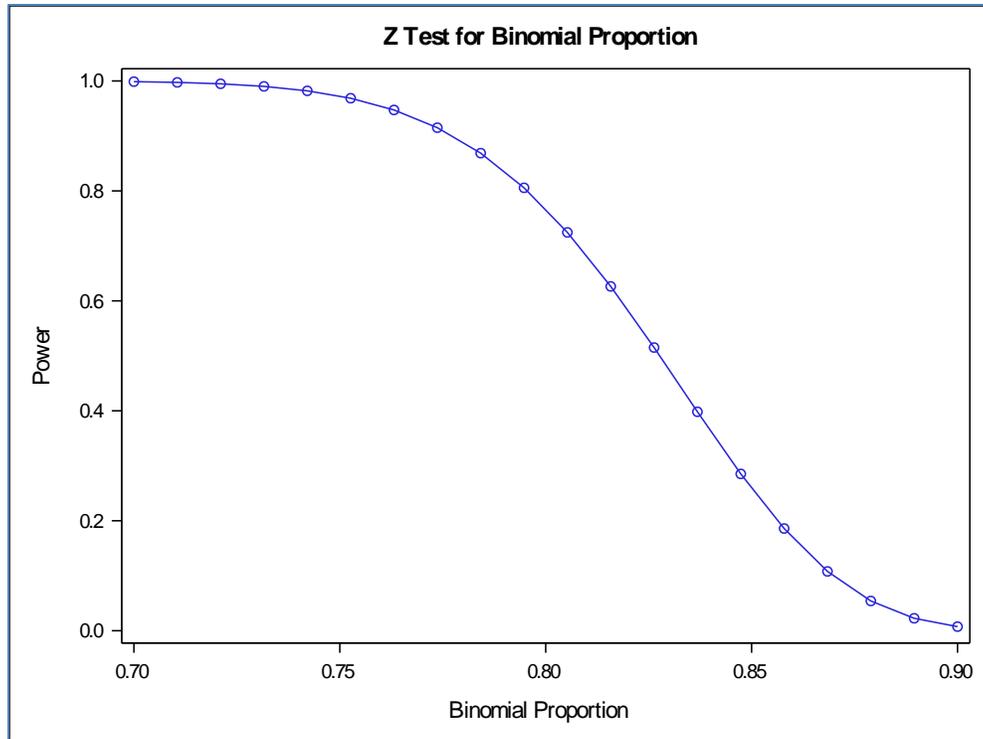
Helpful hints:

1. You may need to enter a particular test in order to change some of the settings. In this example, we need to enter the approximate z-test first in order to be able to determine a total sample size.

2. The Code window can be used to see the underlying Proc Power syntax. This can be saved, or cut/paste into a separate SAS program and further modified to accommodate specific designs that may not be pre-programmed into SAS Studio.

3. Click on one of the tabs to download results into html, pdf, or rtf files.

EXAMPLE 2: TWO SAMPLE MEANS

An intervention is being considered to increase exercise capacity in patients with a certain medical condition. Exercise capacity is measured by the distance a patient can walk in 6 minutes, often referred to as the 6 Minute Walk Test (6MWT). On average, patients with the medical condition can walk approximately 360 meters in 6 minutes. The new intervention hopes to increase this to 400 meters. A study is designed to demonstrate the new intervention increases distance of the 6MWT.

What is the power to show there is an increase in the distance of the 6MWT, assuming the distance increases to 400m? If $\mu_{D=}$ μ_T-μ_C, then positive values for the mean difference suggest an increased walk distance. Our hypotheses are:

$$H_o: \mu_D \leq 0$$
$$H_a: \mu_D > 0$$

The standard deviation of the 6MWT distance is assumed to be 30m for each group.

In SAS Studio, select **Task and Utilities-> Power and Sample Size->t Tests.**

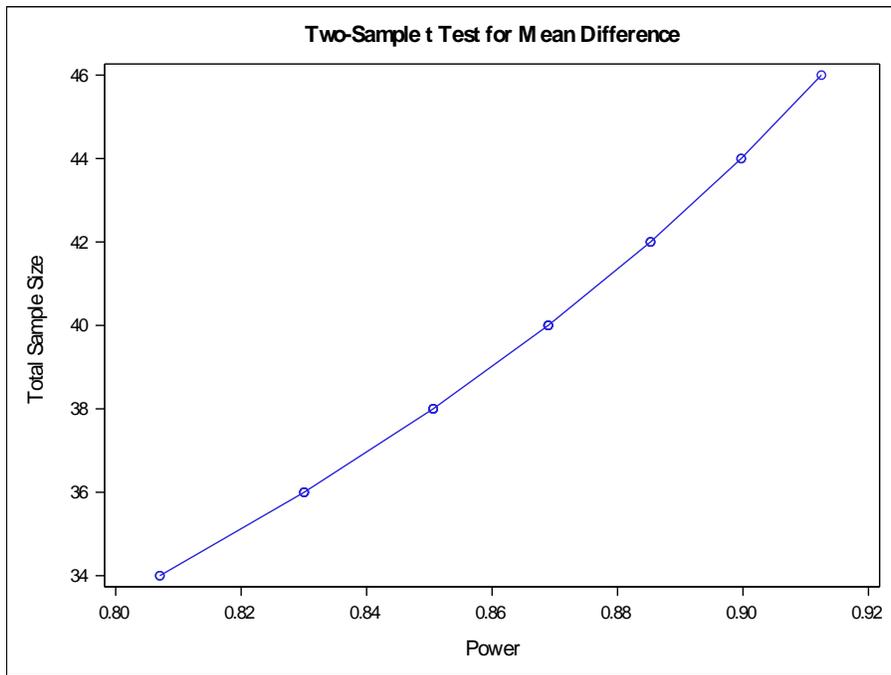**Table 3: Design parameters for difference in 2 means**

Parameter	Value
Type of Test	Two-sample test
Solve for:	Total Sample Size
Analysis details	Normal distribution Two-sided test Pooled t test
Significance level	0.05
Mean	Group means
Group 1	360
Group 2	400
Standard Deviation	40
Power	.9
Sample Size	Equal group sizes
Plots:	Power by Sample Size: Min=0.80 Max=0.90

Note we specified a two-sided alternative with alpha=0.05. In this design, this is equivalent to testing a one sized alternative of alpha=0.025 due to symmetry in the Normal distribution. Under the assumptions provided, we find n=46 patients are need to for the experiment.

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Number of Sides	2
Alpha	0.05
Group 1 Mean	360
Group 2 Mean	400
Standard Deviation	40
Nominal Power	0.9
Null Difference	0
Group 1 Weight	1
Group 2 Weight	1

Computed N Total	
Actual Power	N Total
0.912	46

We use the requested plot to assess sample size for other powers of potential interest.

Two-Sample t Test for Mean Difference

The following are changes to input design parameters that yield the same (or similar) results:

Table 4: Updated design parameters for difference in means

Parameter	Value
Mean	Difference between means
Mean difference	40

We see that power and sample size determination for difference between group means is very straightforward.

Helpful hints:

1. The above design could have specified a one-sided hypothesis test at 0.025. It is important to enter the consistent values for Group 1 and Group 2. If these are not consistent (i.e., one-sided lower tailed test is specified, but Group 1>Group 2), SAS Studio may not let you even proceed.

EXAMPLE 3: SURVIVAL ANALYSIS

An experimental treatment for cancer is expected to increase a patient's life expectancy. To show this, the company needs to show the median time from treatment to death is increased if patients receive the new experimental treatment. Alternatively, the company

needs to demonstrate the hazard rate decreases (with respect to Standard of Care) if patients receive the new experimental treatment.

EXPONENTIAL DISTRIBUTION

Under an assumption that event times follow an exponential distribution which defines a constant risk over time, we see the following relationships, which are helpful when determining the design properties for determining a sample size.

- Hazard rate can be determined simply from knowing the survival at a fixed point in time

- Let p is the proportion with event at some time T, then (1-p) is the survival rate at time T

With these definitions, we find:

$$hazard\ rate\ \lambda = \frac{-\ln(1-p)}{T}$$

$$Median\ survival = \frac{-\ln(.5)}{\lambda}$$

From this we see that decreased hazard rate directly translates to an increased median survival.

For a detailed review of survival analysis, see [3].

SURVIVAL ANALYSIS

Returning to our example above, additional information is needed either about the hazard rate or median survival. We learn that after 18 months of treatment, the expected survival rate for the control group is only 40% whereas the survival rate for those receiving experimental treatment is 55%. This implies that at 18 months, the corresponding event rates are expected to be 60% and 45%.

Using the above relationships we find:

Table 5: Relationships in exponential survival

Hazard rate for Control Group, λ_C	-ln(1-.6)/18=0.0509
Hazard rate for Treatment Group, λ_T	0.0332
Median survival for Control	13.6 months
Median survival for Treatment	20.9 months
Hazard ratio (λ_T / λ_C)	0.6525

Not all rows of Table 2 are needed for the design inputs in SAS Studio. They are provided to demonstrate the relationships. Underlying assumptions obtained may make certain parameters easily obtainable.

To construct a statistical hypothesis, we consider a ratio of hazard rates (λ_T / λ_C) which we refer to as the hazard ratio. If the experimental treatment does in fact decrease the hazard rate, the hazard ratio is less than one (1). If the experimental treatment is the same as Standard of Care, we expect the hazard ratio to be equal to 1. Thus, the hypothesis is:

$$H_o: {}^{\lambda_T}/_{\lambda_C} \geq 1$$

$$H_a: {}^{\lambda_T}/_{\lambda_C} < 1$$

The Log-rank test is commonly used to test the above hypothesis.

In SAS Studio, select **Task and Utilities-> Power and Sample Size->Survival Rank Tests**.

Using this information, we find the following inputs for SAS Studio:

Table 6: Design parameters for survival analysis

| Parameter	Value
Solve for	Number of events
Test	Log rank
Number of sides	Two
Alpha	.05
Survival distribution	Group survival hazards
Group 1	0.0509
Group 2	0.0332
Accrual time	Accrual and follow-up times
Accrual	24 months
Follow-up	18 months
Power	0.9
Plots	Total events by power (min=0.8, max=0.9)

The Results tab provides a summary of the input design parameters and the numbers of events at the requested sets of assumptions.

Fixed Scenario Elements	
Method	Lakatos normal approximation
Form of Survival Curve 1	Exponential
Form of Survival Curve 2	Exponential
Number of Sides	2
Accrual Time	24
Follow-up Time	18
Alpha	0.05
Group 1 Survival Exponential Hazard	0.0509
Group 2 Survival Exponential Hazard	0.0332
Nominal Power	0.9
Number of Time Sub-Intervals	12
Group 1 Loss Exponential Hazard	0
Group 2 Loss Exponential Hazard	0

Fixed Scenario Elements	
Group 1 Weight	1
Group 2 Weight	1

Computed Ceiling Event Total		
Fractional Event Total	Actual Power	Ceiling Event Total
230.327839	0.901	231

We further see from the plot that we can easily obtain event totals for other assumed values for power ranging from .8 to .9.

Log-Rank Test for Two Survival Curves
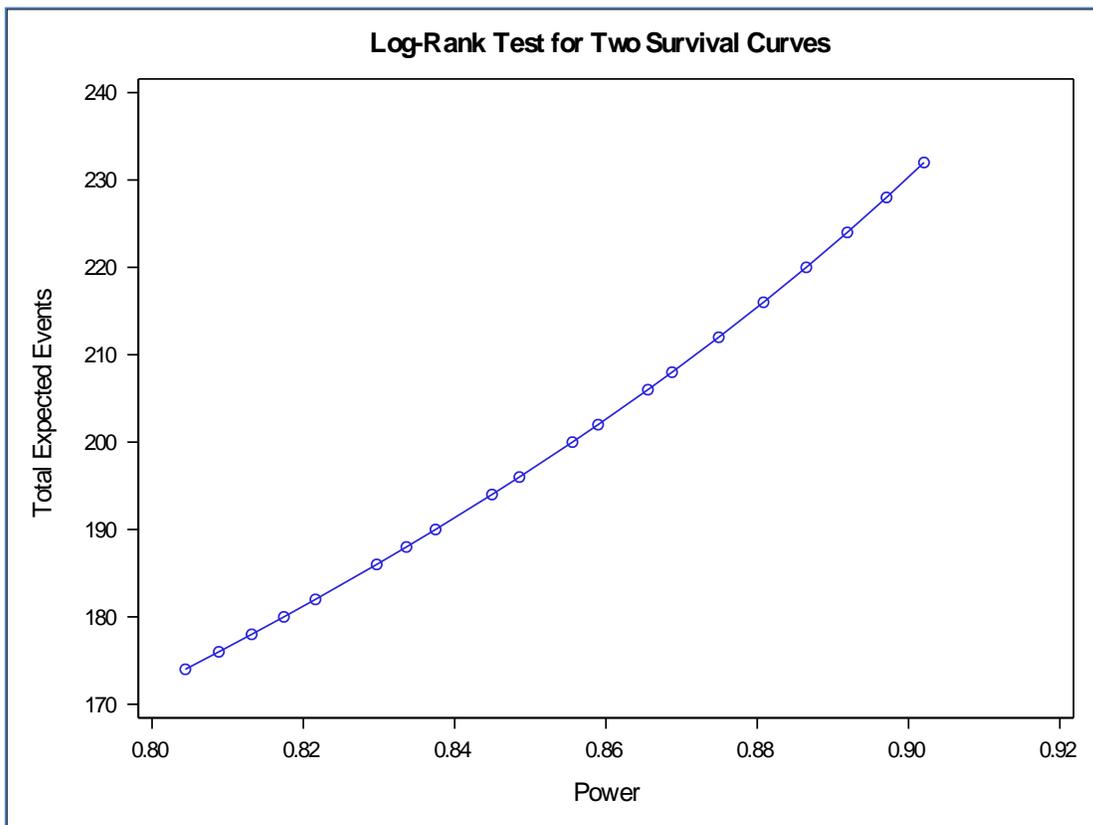
We see that under the specified assumptions, a total of 231 events are needed.

While this is important from a statistical perspective, operationally we need to determine how many patients to enroll such that 231 events can be obtained. If we have a long follow-up time to observe events, we can enroll fewer patients (i.e., have a smaller sample size). Similarly, if we can have rapid enrollment (i.e., smaller accrual period) of a larger number of

patients, we may need a shorter follow-up time. Thus, assumptions for the accrual and follow-up period should be consistent with what is planned operationally.

To determine the sample size needed, change the input parameter to solve for total sample size.

Table 7: Updated parameters for survival analysis

Parameter	Value
Solve for	Total sample size
Sample size	Equal group sizes

With the specified accrual and follow-up time assumptions, approximately 332 patients are needed to obtain the required 231 events.

The following are changes to input design parameters to use the median survival times instead of the hazards that yield the same (or similar) results:

Table 8: Alternative design parameters for survival analysis

Parameter	Value
Solve for	Number of events
Test	Log rank
Number of sides	Two
Alpha	.05
Survival distribution	Group median survival times
Group 1	13.6
Group 2	20.9
Accrual time	Accrual and total times
Accrual	24 months
Total time	42 months
Power	0.9

With this, we see the relationships in Table 2 allow different specifications for design inputs, and all yield consistent results. Differences are due to precision of input parameters.

Helpful hints:

1. Power is strictly based on the number of events. It doesn't matter what duration for accrual is entered when determining the number of events needed.

2. Generally requires two executions – one to determine the number of events while the second is to determine the number of patients needed to achieve the necessary events.

3. Depending on the number of decimal places entered and various parameters used, it is possible to get minor differences in event totals or sample sizes. The differences are simply due to precision.

CONCLUSION

SAS Studio provides the ability to quickly determine sample sizes for many applications. Only a few were discussed here, but we hope this hands-on workshop encourages the use of SAS Studio in your own applications. The user interface is easy to manage, results can be obtained quickly, and the code can be saved elsewhere for future use.

REFERENCES

[1]   ICH GCP Guidelines
[2]   E9 Statistical Principles for Clinical Trials
[3]   Hosmer, D.W., and Lemeshow, S., 1999, *Applied Survival Analysis*, John Wiley & Sons.

ACKNOWLEDGMENTS

RECOMMENDED READING

- *Biostatistics by Example Using SAS Studio*

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Bill Coar
Axio Research
2601 4th Ave, Ste 200
Seattle, WA, USA 98121
Work Phone: 970-219-8564
Email: williamc@axioresearch.com, bill.coar@gmail.com