

Paper 195-29**"Proc Power in SAS 9.1"**

Debbie Bauer, MS, Sanofi-Synthelabo,
Russell Lavery, Contractor, About Consulting, Chadds Ford, PA

ABSTRACT

Features were added to SAS® V9.1 STAT that will allow many users to do all their power and sample size calculations in SAS. Two free STAT Procs (Proc Power and Proc GLMPower) can allow companies to save money by canceling their licenses for software packages that only do sample size calculations. Not only are these Procs full featured, but the ODS capabilities in SAS make it easy to take the results into a finished document.

This paper will quickly review the logic of power calculations and give examples of SAS syntax typically used in a pharmaceutical setting. Finally, it will present a table that compares the features of SAS to several power calculating packages. It is hoped that this table can be used to determine if SAS V9.1 can meet all of the reader's needs for power and sample size calculations. **NOTE: This paper was created using a pre-production of V9.1 and users should, before canceling licenses for other products, check that the production version of 9.1 contains the features they need.**

INTRODUCTION

SAS has had sample size capabilities as part of SAS STAT since Version 7.0. It has been part of SAS Analyst and not generally accessed through SAS code. The sample size windows in Analyst are easy to use, and the output is attractive, but SAS lacked the ability to calculate sample sizes/power for several tests commonly used by pharmaceutical companies. This forced many companies to write their own macros to calculate sample size or to license software to perform these calculations. SAS V9.1, has changed this. The two new Procs are so full featured, and easy to use, that SAS can now become the sole sample size/power package for many companies.

The paper is organized into the following parts:

- | | |
|------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|
| 1) An introduction to power | 2) An imaginary four-section research project |
| 3) Section 1: Examine "what if situations" for the project | 4) Section 2: A two sample mean |
| 5) Section 3: Fisher's exact test | 6) Section 4: A survival analysis |
| 7) The versatility of V9.1 power calculations | 8) A feature comparison table among (SASV9.1, SAS V8.2, Nquery , UnifyPow, StatXact 5.0, Sampsize 2.0) |

1) AN INTRODUCTION TO POWER

The power of a hypothesis test is the probability of rejecting the null hypothesis when the null hypothesis is false. Power generally increases with sample size. We want a test to be sufficiently powered to detect a meaningful difference, but an increase in power means an increase in the cost of conducting the experiment. Not enrolling enough patients, may lead an "under-powered" test where a significant difference may not be detected. Enrolling too many patients, would lead to an "over-powered" test, which would waste time and money. To conduct an adequately powered test, an experimenter would need to know the following:

- 1) A clinically meaningful effect size,
- 2) The standard deviation of the response variable,
- 3) The total number of patients that can be enrolled

2) AN IMAGINARY FOUR-SECTION RESEARCH PROJECT- OVERVIEW

The first step in the drug/disease research process is to do a literature search to learn about previous studies conducted and apply some experience to what is found. Imagine we are studying a new anti-depressant compound and studies in depression have found: standard deviation in the change from baseline in the HAM-D score is about 8.5 points, a clinically meaningful effect is considered to be 3.5 and studies typically have 5% of their subjects leave the study prior to an efficacy assessment. Additionally, we learned that some medications for the treatment of depression have caused dizziness or nausea in 20% of the patients. We would like to investigate an experimental drug in the treatment of depression with a fewer incidence of these adverse events.

Generally, balanced studies (studies with the same number of patients in the treatment and placebo/control groups) require the minimal number of patients. Studies with equal numbers of patients in treatment and placebo/control groups minimize study cost and provide higher power-per-dollar than unbalanced studies. Often cost or risk differences between treatment/control may preclude use of balanced designs. There are ethical constraints in assigning patients to treatment types. In some indications, it may be unethical or undesirable to assign patients to placebo/control. For example, unmedicated, severely depressed patients who are at higher risk to commit suicide. To reduce the risk to patients in this study, we decide to have a larger, more costly study by deliberately "unbalancing" the sample. The researcher decides to conduct a study with 3 treatment arms and have one fifth of the

subjects as placebo/controls and to have two fifths of the subjects get the experimental drug (which we expect to be an effective treatment for depression) and another two fifths get a standard treatment for depression. The imaginary research project will have three sections requiring performance of power calculations, and achieving the goals of those three sections requires some compromising on sample size.

The first section of the study involves making judgments based on information found in the literature search. Parameters passed to the SAS Procs are based on these judgments.

The second section investigates the primary objective of the study: effectiveness. It will be a two part test of the ability of the drug to reduce the average score on the HAM-D scale. The primary endpoint will be the change from baseline in HAM-D total score. The difference between the experimental drug and placebo will also be evaluated. The study will primarily be powered to detect a meaningful difference in these scores.

The third section of the study investigates “gentleness” of the compound. It will compare the rate of adverse events for this drug against the rates of the standard comparator. A drug could be marketable even if it is less effective than the competition. Patients are often willing to select a less effective drug, if that drug has a lower incidence of adverse events. This will be investigated using Fisher’s Exact Test. The researcher determines meaningful differences in the proportion of patients experiencing several adverse events and checks the power of Fisher’s test to detect these differences using the sample size calculated for the primary endpoint.

In the fourth section, the study investigates speed of action. It will study the average time to a sustained 50% reduction in the HAM-D total score (i.e. sustained responders). A fast acting drug would be more marketable and preferable to patients suffering from depression than a slower acting drug. This will be investigated with a log-rank test comparing the survival curves. The researcher will check the power of the test, for the log-rank test, using the sample size calculated for the primary endpoint.

3) SECTION 1: EXAMINE “WHAT IF SITUATIONS” FOR THE PROJECT

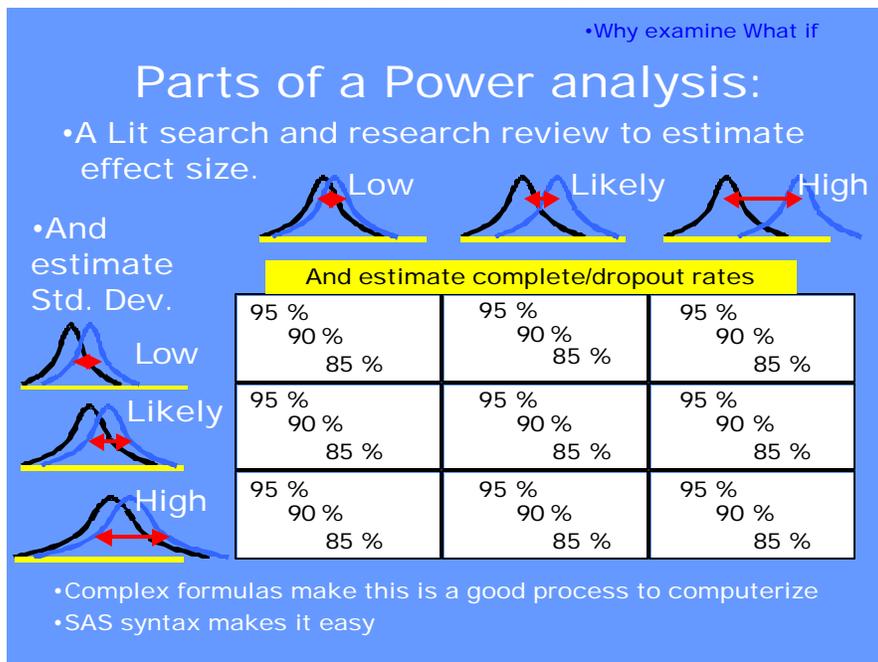


Figure 1

In this step, the statistician establishes several reasonable values (usually Low, Likely, High) for parameters that will be passed to SAS Procs and decides if the study should be balanced or not. The parameter estimates (effect sizes, standard deviations and drop out rates) found in the literature search are results of past studies, with different subjects and under different conditions. There is no assurance that the proposed study will have similar numbers. The researcher will always want to investigate the sensitivity of the study power to the accuracy of these numbers. Figure 1 illustrates how just three “reasonable parameter values” multiply into a great number of situations for which power must be calculated. Researchers avoid manually calculating power because the calculations are complex. Easy to use software, like SAS, is a great help.

4) SECTION 2: TWO SAMPLE INDEPENDENT MEAN

The syntax below shows how easy it is to have SAS do a power calculation for a test of two means.

```
proc power;
  twosamplemeans
  meandiff= 3 3.5 4
  stdev=8 to 9 by .5
  groupweights=(1 2)
  power=0.8
  ntotal=.;
  plot y=power min=0.5 max=0.99;
run;
```

The fact that ntotal (total sample size) is set to . is the instruction to SAS that other numbers should be used to calculate total sample size. The power option instructs SAS to calculate the total number of subjects required for an 80 percent chance to detect an effect. The meandiff option instructs SAS to do the calculations for three different estimates of effect (3, 3.5 and 4). The stdev option instructs SAS to perform the calculation for standard deviations of 8, 8.5 and 9. To limit subjects on placebo, we use the groupweights statement to have SAS allocate twice as many subjects to the experimental drug. The syntax above produces the plot shown in Figure 2.

Output

| Mean dif | Std Dev | Power | N total |
|------------|------------|--------------|------------|
| 3 | 8 | 0.803 | 255 |
| 3 | 8.5 | 0.803 | 288 |
| 3 | 9 | 0.801 | 321 |
| 3.5 | 8 | 0.805 | 189 |
| 3.5 | 8.5 | 0.805 | 213 |
| 3.5 | 9 | 0.803 | 237 |
| 4 | 8 | 0.802 | 144 |
| 4 | 8.5 | 0.801 | 162 |
| 4 | 9 | 0.805 | 183 |

The table above shows (assuming a mean difference of 3.5 with a standard deviation of 8.5) that 213 patients (142 on treatment, and 71 on placebo) are needed to provide 80% power. Assuming 5% drop-out, approximately, 375 total patients should be enrolled. (75 on placebo and 150 each on the experimental treatment and standard comparator)

SAS is versatile and the following code modifications, power=. Ntotal=250, and plot x=n min=150 max=400, produces the "reverse plot" that is shown in Figure 3.

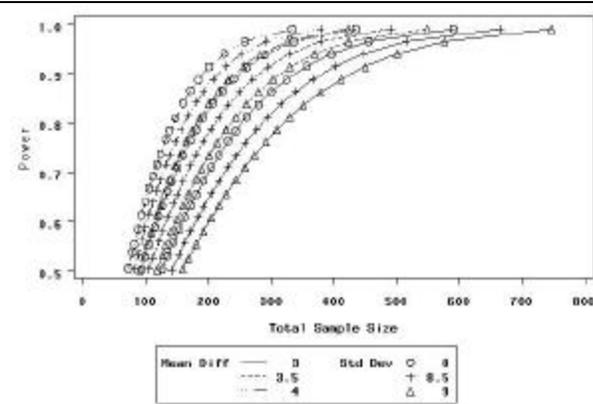


Figure 2

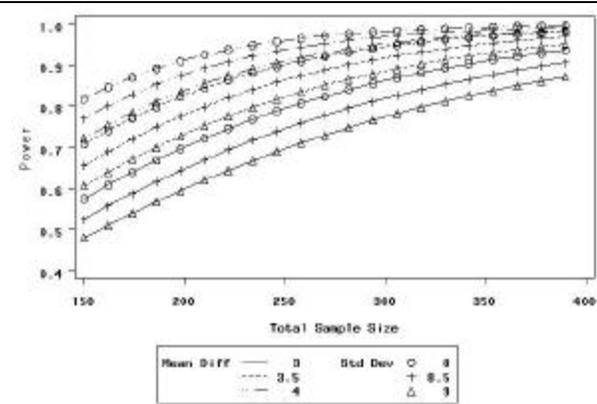


Figure 3

5) SECTION 3: FISHER'S EXACT TEST

The syntax below shows how easy it is to have SAS do a power calculation for a Fishers Exact test.

```
Proc power;
  Twosamplefreq test=fisher
  Proportiondiff=0.10 to 0.15 by 0.01
  Refproportion=0.20
  Npergroup=150
  Power=.;
Run;
```

The fact that power is set to . is the instruction to SAS to calculate power. Other parameters specify a Fisher's exact test with 150 patients per active treatment group. This assumes 20% of patients on the standard comparator (i.e. a competing drug) experience an adverse event, and we power the test to detect a .10 to .15 difference between it and our experimental drug.

Output

| Prop. Diff (Ref. Diff=0.20) | Power |
|--------------------------------------|--------------|
| 0.10 (10% of patients with an AE) | 0.466 |
| 0.11 | 0.541 |
| 0.12 | 0.614 |
| 0.13 | 0.683 |
| 0.14 | 0.745 |
| 0.15 (5% of patients with AE) | 0.800 |

The table shows we could have no more than 5% of patients on our experimental drug experience an adverse event compared to 20% on a standard comparator to have 80% power to detect the difference.

6) SECTION 4: SURVIVAL ANALYSIS

The syntax below shows how easy it is to have SAS do a power calculation for a survival analysis.

| <pre>proc power; twosamplesurvival test=logrank gexphs= 0.3567 0.5978 .6931 grouplossexphazards=(0.3567 0.3567) accrualtime = 1 followuptime = 1 groupweights = (1 2) power = . ntotal=225; run;</pre> <p>This assumes 30% of placebo patients are sustained responders (exponential hazard =0.3567) compared to 45 or 50% for the treatment group (exp. hazard = 0.5978 or 0.6931). Twice as many patients are on treatment as placebo, and all patients are enrolled at the beginning of the study with a 30% drop-out rate.</p> | <table border="1"> <thead> <tr> <th>Exp Hazard Group 2</th> <th>Power</th> </tr> </thead> <tbody> <tr> <td>0.598</td> <td>0.620</td> </tr> <tr> <td>0.693</td> <td>0.843</td> </tr> </tbody> </table> <p>The table above shows (assuming 30% of patients on placebo are sustained responders, compared to 50% on treatment, and 30% of patients drop-out before the study completes) a total of 225 patients (150 on the experimental treatment and 75 on placebo) would provide 84% power to detect a difference.</p> | Exp Hazard Group 2 | Power | 0.598 | 0.620 | 0.693 | 0.843 |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|-------|-------|-------|--------------|--------------|
| Exp Hazard Group 2 | Power | | | | | | |
| 0.598 | 0.620 | | | | | | |
| 0.693 | 0.843 | | | | | | |
| <p>Many additional features not displayed in the SAS code above are available with this procedure including specifying hazard ratios, median survival times, or actual survival curves (either exponential or piecewise linear).</p> | | | | | | | |

The final sample size for this study should be 375 patients (150 on the standard treatment, 150 on the experimental treatment and 75 on placebo). This provides sufficient power for the primary analysis in section 2, as well as sufficient power for the two secondary endpoints in Section 3 and 4.

7) THE VERSATILITY OF V9.1 POWER CALCULATIONS

Figure 4

Generally the researcher is in one of two states of nature: 1) limited money so N is fixed or 2) money available and power fixed (Figure 4). In state 1) the statistician is told, “We only have enough money for 120 subjects, what can we do with that sample size?” In state 2) the statistician is told to calculate the sample size required for a certain power (80% or 90%).

Additionally, the study might have equal or unbalanced costs/risks associated with different treatments. One treatment might be more costly and the researcher would wish to assign fewer subjects to the expensive treatment in an effort to minimize the costs. Ethically, if patients are at different risks in different treatment groups, the statistician is obliged to investigate having fewer patients in the higher risk group. Figure 4 shows how easy it is to request that SAS perform the complex calculations required to

minimize costs/risks for both balanced and unbalanced studies. In both of the examples, in Figure 4, very simple coding instructs SAS to calculate sample sizes for both balanced and unbalanced designs (blue arrows).

8) A FEATURE COMPARISON TABLE AMONG (SASV9.1, SAS V8.2, NQUERY 5.0, UNIFYPOW 2002.08.17A, STATXACT, SAMPSIZE 2.0)

| CONTINUOUS METHODS | | | | | | |
|-------------------------------------------------|-----------------|-----------------|-------------------|-----------------------------|-----------------|---------------------|
| Difference | SAS V9.1 | SAS V8.2 | Nquery 5.0 | UnifyPow 2002.08.17a | StatXact | Sampsize 2.0 |
| One-sample t-test | Yes | Yes | Yes | Yes | | Yes |
| One-sample t-test with lognormal data | Yes | Yes | | Yes | | |
| Paired t-test | Yes | Yes | Yes | Yes | | Yes |
| Paired t-test of mean ratio with lognormal data | Yes | Yes | | Yes | | |

| | | | | | | |
|------------------------------------------------------------------------------|-----------------|-----------------|-------------------|-----------------------------|-----------------|---------------------|
| Two-sample t-test | Yes | Yes | Yes | Yes | | Yes |
| Two-sample Satterthwaite t-test (unequal variances) | Yes | | Yes | | | |
| Two-sample pooled t-test of mean ratio with lognormal data | Yes | Yes | Yes | Yes | | |
| One-way ANOVA | Yes | Yes | Yes | Yes | | |
| Two-way ANOVA | Yes | | Yes | Yes | | |
| General ANOVA | Yes | | Yes | Yes | | |
| Repeated measures ANOVA | Yes | | Yes | Yes | | |
| Simple linear regression | Yes | | Yes | Yes | | Yes |
| Multiple Linear Regression | Yes | | Yes | Yes | | |
| Wilcoxon Signed rank test for one group | | | | Yes | | |
| Wilcoxon Signed rank Test for matched pairs | | | | Yes | | |
| Wilcoxon (Mann-Whitney) rank sum test (2 independent groups) | | | Yes | Yes | | |
| Equivalence Tests | SAS V9.1 | SAS V8.2 | Nquery 5.0 | UnifyPow 2002.08.17a | StatXact | Sampsize 2.0 |
| One-sample equivalence test | Yes | Yes | Yes | | | |
| One-sample equivalence test for lognormal data | Yes | | | | | |
| Paired additive equivalence of mean difference | Yes | Yes | Yes | | | |
| Paired multiplicative equivalence of mean ratio with lognormal data | Yes | | | | | |
| Two-sample additive equivalence of mean difference | Yes | Yes | Yes | Yes | | Yes |
| Two-sample multiplicative equivalence of mean ratio with lognormal data | Yes | | Yes | | | |
| Confidence Intervals | SAS V9.1 | SAS V8.2 | Nquery 5.0 | UnifyPow 2002.08.17a | StatXact | Sampsize 2.0 |
| One group mean | Yes | Yes | Yes | | | |
| Mean of paired differences | Yes | Yes | Yes | | | |
| Two-sample for mean differences | Yes | Yes | Yes | | | |
| Difference or ratio of two means in general ANOVA | Yes | | Yes | | | |
| CATEGORICAL METHODS | | | | | | |
| Difference tests | SAS V9.1 | SAS V8.2 | Nquery 5.0 | UnifyPow 2002.08.17a | StatXact | Sampsize 2.0 |
| Single Binomial test | Yes | | Yes | Yes | Yes | Yes |
| Chi-square or likelihood ratio test for two independent proportions | Yes | | Yes | Yes | Yes | |
| Fisher's exact test for two independent proportions | Yes | | Yes | Yes | Yes | |
| Mc Nemars test | Yes | | Yes | Yes | | Yes |
| Cochran Mantel-Haenszel test | | | Yes | | | |
| Cochran-Armitage Trend test | | | Yes | | Yes | |
| Wilcoxon-Mann Whitney U test for two independent Ordered Categorical Samples | | | Yes | Yes | Yes | Yes |

| Equivalence tests | SAS V9.1 | SAS V8.2 | Nquery 5.0 | UnifyPow 2002.08.17a | StatXact | Sampsize 2.0 |
|---------------------------------------------------------|-----------------|-----------------|-------------------|---------------------------------|-----------------|-------------------------|
| Non-inferiority of single proportion vs Fixed reference | Yes | | | Yes | | |
| Non-inferiority of two paired proportions | Yes | | Yes | | | |
| Non-inferiority of two independent proportions | Yes | | Yes | Yes | | |
| Equivalence of two independent proportions | | | Yes | | Yes | Yes |
| Confidence Interval | SAS V9.1 | SAS V8.2 | Nquery 5.0 | UnifyPow 2002.08.17a | StatXact | Sampsize 2.0 |
| Single Proportion | | | Yes | | | |
| Difference between two independent proportions | | | Yes | | | |
| Log Odds ratio For two independent proportions | | | Yes | | | |
| Log Odds ratio for two paired proportions | | | Yes | | | |
| SURVIVAL TESTS | | | | | | |
| Difference tests | SAS V9.1 | SAS V8.2 | Nquery 5.0 | UnifyPow 2002.08.17a | StatXact | Sampsize 2.0 |
| Log Rank | Yes | | Yes | | | |
| Gehan rank test | Yes | | | | | |
| Tarone-Ware rank test | Yes | | | | | |

CONCLUSION

SAS V9.1 Proc Power is full featured enough to satisfy the needs of major research institutions and is free with V9.1 STAT.

REFERENCES

Conference proceedings are a valuable source of information of SAS topics. Many years of NUGI and NESUG proceedings are on-line and accessible for free. An Excel Spreadsheet listing of the article titles, authors and web locations is available from www.About-Consulting.com.

CONTACT INFORMATION

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

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Deborah Bauer, MS 9 Great Valley Parkway Malvern, PA 19355 610-889-8839 EMAIL: DEBORAH.BAUER@SANOFI-SYNTHELABO.COM | Russell Lavery, Independent Contractor About-Consulting, LLC (About-Consulting.Com) Chadds Ford, PA 610-388-9455 EMAIL: RUSS.LAVERY@VERIZON.NET |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

SAS is a registered trademark or trademark of SAS Institute, Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are registered trademarks of their respective companies.