Using SAS Proc Power to Perform Model-based Power Analysis for Clinical Pharmacology Studies

Peng Sun, Merck & Co., Inc., North Wales, PA

ABSTRACT

In this paper, we demonstrate that many hypothesis testing problems in clinical pharmacology studies can be reformulated as a one- or two-sample t-test or equivalence test. Thereby one can use SAS® Proc Power to perform the power analysis. Sample programs for different hypothesis testing problems are presented in a variety of study design settings. Compared to other commercial software such as NQuery and PASS, using Proc Power offers several advantages. First, one often obtains variance estimates by analyzing relevant clinical data in SAS, thus it is more streamlined and convenient to conduct power analysis in SAS as well. Second, the documentation in SAS is more comprehensive than those in NQuery and PASS; this helps the users better understand the procedure and hence the users are less likely to make mistakes. Third, as demonstrated in the examples, one can take advantage of the programming capabilities in SAS and make inputting parameters for power analysis more straightforward. Last, but not least, using SAS can better help the users keep a record of the power analysis, which is helpful for adapting the code for similar problems in the future.

KEYWORDS: SAS Proc Power, Model-based Power Analysis, Clinical Pharmacology Studies, t test, equivalence test

1. INTRODUCTION

The main objective of pharmacology studies is often to characterize the pharmacokinetics (PK) profile of the candidate drug, which is represented by the concentration-time curve and parameters derived from it, such as AUC, C_{max}, C_{trough}, T_{max}, and apparent $t_{1/2}$. There are a variety of clinical pharmacology studies, including first-in-man, multiple dose, single dose elderly, drug-drug interaction, bioequivalence, QTc, hepatic insufficiency, etc. Study designs such as parallel group, fixed sequence, and crossover have all been applied in those studies. For example, multiple dose studies and hepatic insufficiency studies often use a parallel group design; drug-drug interaction studies and bioequivalence studies, on the other hand, often utilize a fixed sequence design or a crossover design. In terms of hypothesis testing problems, the two most common ones are one-sided significance test and equivalence test versus two-sided significance tests. Although Pharmacodynamics (PD) endpoints are routinely collected in clinical pharmacology studies, most hypothesis testing problems are for PK endpoints and this is the main focus of this paper. The reader can easily adapt the methods presented here for power analysis of PD hypotheses.

For hypotheses formulated through PK parameters, the parameters of interest are typically AUC, C_{max}, or C_{trough}. Since experience has shown that those PK parameters are lognormally distributed, the hypotheses are typically stated using the population geometric mean (or equivalently, the population median) of the pharmacokinetic parameters. For example, in a multiple dose study, the hypothesis may be stated as "At one or more well tolerated doses, the steady state population geometric mean C_{trough} exceeds 35 nM"; in a bioequivalence study, the hypothesis may be stated as "The population geometric mean $AUC_{0-\infty}$ ratio (Formulation 2 / Formulation 1) is between 0.80 and 1.25". Note that, because of the lognormality assumption, PK hypotheses in clinical pharmacology studies can also be stated in terms of the population mean or mean difference of the log-transformed PK parameters. For the previous examples, "population mean C_{trough} exceeds 35 nM" is equivalent to "Population mean $\ln(C_{trough})$ exceeds $\ln(35$ nM)"; "population geometric mean $AUC_{0-\infty}$ ratio is between 0.80 and 1.25" is equivalent to "difference in population mean $\ln(AUC_{0-\infty})$ is between $\ln(0.80)$ and $\ln(1.25)$". Throughout this paper, the power analyses will be discussed in the log scale. As a result, the parameter of interest is either a population mean or a population mean difference. This makes inputting parameters for power calculations more straightforward, hence less error-prone.

The Proc Power procedure was first introduced in SAS 9.1 to perform prospective power and sample size analysis for a variety of statistical methods, including t-test, equivalence test, one way analysis of variance, logistic regression with binary response, etc. In this paper, we will demonstrate that, for most clinical pharmacology studies, one can use the one- and two-sample t-tests and equivalence tests in Proc Power to perform model-based power calculations. Because using Proc Power involves reformulating a more complex problem into a simple statistical test, one needs a deeper understanding of the underlying statistical model. To that end, the relevant theoretical results and methodologies will be presented in Section 2. This is followed by several clinical pharmacology study examples to illustrate the implementation of different model-based power analysis in Section 3, illustration of power assessment for higher order crossover designs in Section 4, and further discussion in Section 5.

2. THEORETICAL BACKGROUND

We start by reviewing the noncentral t distribution.

Definition: Let $Z \sim N(0,1)$, independently, $U \sim \chi^2(v)$, and δ be a real number, then $t = \dfrac{Z+\delta}{\sqrt{U/v}} \sim t(v, \delta)$.

Under the statistical models used in most clinical pharmacology studies, the test statistic follows a t distribution (central or non-central, depending on the true parameter value). In order to simplify the discussion, we adopt the approach of Berger and Hsu (1996) and present the construction of the t statistic in the context that is independent of the underlying study design. In general, the construction of a t statistic involves two sample statistics. The first statistic T has a normal distribution with mean μ and variances σ_T^2. The second statistic, denoted as $\mathrm{SE}(T)$, estimates the standard error of T. $\mathrm{SE}(T)$ is independent of T and $r[\mathrm{SE}(T)]^2 / \sigma_T^2 \sim \chi^2(r)$, where the degree of freedom r is a positive number that depends on the study design. It follows that, for any real number a,

$$t = \frac{T-a}{\mathrm{SE}(T)} = \frac{\dfrac{T-\mu}{\sigma_T} + \dfrac{\mu-a}{\sigma_T}}{\sqrt{\dfrac{r[\mathrm{SE}(T)]^2 / \sigma_T^2}{r}}} \sim t(r, \frac{\mu-a}{\sigma_T}) \tag{1}$$

We next present several examples to illustrate how T and $\mathrm{SE}(T)$ can be constructed in some common study designs, including fixed sequence, parallel group, one-way analysis of variance (ANOVA), and 2 x 2 crossover designs. Those examples will serve as the theoretical foundation for power calculations in those settings. Power analysis for higher order crossover designs will be discussed separately in Section 4.

2.1 EXAMPLES FOR CONSTRUCTION OF T STATISTICS

Design 1: Two-period Fixed Sequence Design

The data consists of bivariate measurements (X_{i1}, X_{i2}), $i = 1, \ldots, n$, such that $X_{i2} - X_{i1} \overset{IID}{\sim} N(\mu_D, \sigma_D^2)$. We therefore have $\overline{X}_2 - \overline{X}_1 \sim N(\mu_D, \sigma_D^2/n)$, where $\overline{X}_1$ and $\overline{X}_2$ are sample means for period 1 and period 2 data, respectively. Independently, $(n-1)S_D^2 / \sigma_D^2 \sim \chi^2(n-1)$, where S_D^2 is the sample variance of $\{X_{i2} - X_{i1}, i = 1, \ldots, n\}$. Therefore, (1) holds with:

$$T \equiv \overline{X}_2 - \overline{X}_1,$$
$$SE(T) \equiv \sqrt{S_D^2/n},$$
$$\sigma_T^2 \equiv \sigma_D^2/n,$$
$$\mu \equiv \mu_D,$$
$$r \equiv n-1.$$

Design 2: Two-sample Parallel Group Design with Common Variance

Group 1 consists of n_1 data points $X_{i1} \overset{IID}{\sim} N(\mu_1, \sigma^2)$, $i = 1, \ldots, n_1$; independently, group 2 consists of n_2 data points $X_{i2} \overset{IID}{\sim} N(\mu_2, \sigma^2)$, $i = 1, \ldots, n_2$. It follows that

$$\overline{X}_2 - \overline{X}_1 \sim N(\mu_2 - \mu_1, (\frac{1}{n_1} + \frac{1}{n_2})\sigma^2), \text{ and indepdently}$$
$$(n_1 + n_2 - 2)S_p^2 / \sigma^2 \sim \chi^2(n_1 + n_2 - 2),$$

where S_p^2 is the pooled sample variance

$$S_p^2 = \frac{1}{n_1+n_2-2}[(n_1-1)S_1^2 + (n_2-1)S_2^2],$$

with S_1^2 and S_2^2 being the sample variance of group 1 and group 2, respectively. Therefore, (1) holds with

$$T \equiv \bar{X}_2 - \bar{X}_1,$$

$$SE(T) \equiv \sqrt{(\frac{1}{n_1} + \frac{1}{n_2})S_p^2},$$

$$\mu \equiv \mu_2 - \mu_1,$$

$$\sigma_T^2 \equiv (\frac{1}{n_1} + \frac{1}{n_2})\sigma^2,$$

$$r \equiv n_1 + n_2 - 2.$$

Design 3: 2 X 2 Crossover Design

In a 2 x 2 crossover study, two sequences of subjects are enrolled. Subjects in sequence 1 receive treatment A first, and after a suitable washout period, crossover to treatment B. Subjects in sequence 2, on the other hand, receive treatment B first, then crossover to A. Let y_{ijk} denote the measurement for subject j in sequence i collected in period k ($j=1,...,n_i$ for $i=1,2$, $k=1,2$). It is assumed that

$$y_{ijk} = \mu + g_i + \pi_k + \tau_{d(i,k)} + s_{ij} + \varepsilon_{ijk},$$

where

- g_i: Effect associated with group (sequence) i.

- π_k: Effect associated with period k.

- $\tau_{d(i,k)}$: Effect of treatment $d(i,k)$ - the treatment in sequence i of period k (e.g, $d(1,1)=A$).

- s_{ij} : Subject effect with $s_{ij} \overset{IID}{\sim} N(0,\sigma_S^2)$.

- ε_{ijk} : Residual error with $\varepsilon_{ijk} \overset{IID}{\sim} N(0,\sigma^2)$, that is independent to s_{ij}.

Based on the above model, the populations mean of y_{ijk} by sequence and period is summarized in the table below:

Sequence	Period 1	Period 2
1 (AB)	$\mu + g_1 + \pi_1 + \tau_A$	$\mu + g_1 + \pi_2 + \tau_B$
2 (BA)	$\mu + g_2 + \pi_1 + \tau_B$	$\mu + g_2 + \pi_2 + \tau_A$

Following Hills and Armitage (1979), the treatment comparison can be analyzed as a two-sample t-test. Specifically, if we let $d_{1j} = \frac{1}{2}(y_{1j2} - y_{1j1})$ for $j = 1,...,n_1$ and $d_{2j} = \frac{1}{2}(y_{2j2} - y_{2j1})$ for $j = 1,...,n_2$, then $d_{1j} \overset{IID}{\sim} N(\frac{1}{2}(\pi_2 - \pi_1 + \tau_B - \tau_A, \frac{1}{2}\sigma^2))$ and, independently, $d_{2j} \overset{IID}{\sim} N(\frac{1}{2}(\pi_2 - \pi_1 + \tau_A - \tau_B, \frac{1}{2}\sigma^2))$. Hence $E(d_{1j}) - E(d_{2j}) = \tau_A - \tau_B$ and the comparison of τ_A and τ_B reduces to a two-sample t-test with sample sizes n_1 and n_2 and common variance $\frac{1}{2}\sigma^2$.

2.2 FORMULA FOR POWER CALCULATIONS

We next consider the derivation of the power functions in the general framework. We focus on three tests: the lower one-sided test, the upper one-sided test, and the equivalence test.

For a lower one-sided test, under the alternative hypothesis, the true parameter value μ lies to the left of the null value μ_0. Specifically, the hypothesis testing problem is stated as:

$$H_0 : \mu \geq \mu_0 \text{ versus } \mu < \mu_0.$$

For a size-α test, the power function can be easily derived:

$$1 - \beta(\mu, r, \sigma_T) = \Pr\left(\frac{T - \mu_0}{SE(T)} < -t_{r,1-\alpha}\right)$$

$$= \Pr\left(t(r, \frac{\mu - \mu_0}{\sigma_T}) < -t_{r,1-\alpha}\right),$$

where $t_{r,1-\alpha}$ is defined by $\Pr(t(r) \leq t_{r,1-\alpha}) = 1 - \alpha$.

In an upper one-sided test, under the alternative hypothesis, the true parameter value μ lies to the right of the null value μ_0 and the hypothesis takes the form:

$$H_0 : \mu \leq \mu_0 \text{ versus } \mu > \mu_0.$$

For a size-α test, the power function is:

$$1 - \beta(\mu, r, \sigma_T) = \Pr\left(\frac{T - \mu_0}{SE(T)} > t_{r,1-\alpha}\right)$$

$$= \Pr\left(t(r, \frac{\mu - \mu_0}{\sigma_T}) > t_{r,1-\alpha}\right).$$

In an equivalence test, under the alternative hypothesis, the true parameter value μ falls between two pre-specified values μ_L and μ_U. Hence the hypothesis takes the form:

$$H_0 : \mu \leq \mu_L \text{ or } \mu \geq \mu_U \text{ versus } H_1 : \mu_L < \mu < \mu_U. \qquad (2)$$

Westlake (1972) and Schuirman (1987) proposed what has become the standard test of (2). It is called the "two one-sided tests" (TOST). Following the TOST procedure, for a size-α test, one constructs the $(1-2\alpha)100\%$ confidence interval for μ and rejects the null hypothesis if the confidence interval falls entirely between μ_L and μ_U. Therefore the power function can be expressed as

$$1 - \beta(\mu, r, \sigma_T) = \Pr\left(\frac{T - \mu_L}{SE(T)} > t_{r,1-\alpha} \text{ and } \frac{T - \mu_U}{SE(T)} < -t_{r,1-\alpha}\right).$$

Applying law of iterative expectations, Phillips (1990) derived the power function:

$$1 - \beta(\mu, r, \sigma_T) = Q_r\left(-t_{r,1-\alpha}, \frac{\mu - \mu_U}{\sigma_T}; 0 \frac{(\mu_U - \mu_L)\sqrt{r}}{2\sigma_T t_{r,1-\alpha}}\right)$$

$$- Q_r\left(t_{r,1-\alpha}, \frac{\mu - \mu_L}{\sigma_T}; 0 \frac{(\mu_U - \mu_L)\sqrt{r}}{2\sigma_T t_{r,1-\alpha}}\right),$$

where the function

$$Q_\nu(t, \delta; a, b) = \frac{\sqrt{2\pi}}{\Gamma(\frac{\nu}{2}) \cdot 2^{\frac{\nu-2}{2}}} \int_a^b \Phi\left(\frac{tx}{\sqrt{\nu}} - \delta\right) x^{\nu-1} \phi(x) dx$$

is referred to as Owen's Q function. In the definition of Owen's Q function, $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of standard normal distribution, respectively.

3. SOME EXAMPLES

EXAMPLE 1: EQUIVALENCE TEST IN A TWO-PERIOD FIXED SEQUENCE STUDY

Consider a midazolam interaction study to be conducted using a two-period fixed-sequence design. In period 1, midazolam will be administered as a single dose. After suitable washout, in period 2, the candidate drug C will be administered daily for seven days, and on Day 7, the same single dose midazolam will be co-administered with the

candidate drug. The parameter of interest is midazolam $AUC_{0-\infty}$ and the hypothesis is stated as "The true midazolam geometric mean $AUC_{0-\infty}$ ratio (Period 2 / Period 1) lies within (0.5, 2.0)".

We adopt the notation and model of Design 1 and let X_{i1} and X_{i2} represent midazolam $\ln(AUC_{0-\infty})$ in period 1 and 2, respectively. Then the hypothesis of interest can be formally stated as:

$$H_0 : \mu_D \leq \ln(2) \text{ or } \mu \geq \ln(0.5) \text{ versus } H_1 : \ln(0.5) < \mu_D < \ln(2).$$

Although the analysis is in nature a paired t test (or equivalently, a one-sample t test), $\ln(AUC_{0-\infty})$ will be formally analyzed by a linear mixed effect model with period as fixed effect and subject as random effect. Specifically,

$$X_{ij} = \mu_j + s_i + \varepsilon_{ij},$$

where $s_i \overset{IID}{\sim} N(0,\sigma_S^2)$ and independently $\varepsilon_{ij} \overset{IID}{\sim} N(0,\sigma^2)$. This implies that

$$\sigma_D^2 = \text{var}(X_{i2} - X_{i1}) = \text{var}(\varepsilon_{i1} - \varepsilon_{i2}) = 2\sigma^2.$$

Based on a previous study, the within-subject variance σ^2 is estimated to be 0.0735. This implies that σ_D can be estimated as $\sqrt{2 \cdot 0.0735} = 0.3834$. The objective here is to provide power for n=6, 7, 8 under the assumption $\mu_D = 0$ (or equivalently, the true geometric mean ratio is equal to 1.0). This can be accomplished by regarding the problem as a one-sample equivalence test on $X_{i2} - X_{i1}$ and using the "onesamplemeans" option in proc power:

```
%let sigma2=0.0735; * macro variable for the within-subject variance;
%let sigma_D=%sysevalf((2*&sigma2)**0.5); * macro variable for the standard
deviation of the difference;
%let log_pt_5=%sysfunc(log(0.5)); * macro variable for log(0.5);
%let log_2=%sysfunc(log(2)); * macro variable for log(2);
proc power;
    onesamplemeans test=equiv alpha=0.05
    lower=&log_pt_5 upper=&log_2 std=&sigma_D
    mean=0
    ntotal=6 7 8
    power =.;
run;
```

The SAS code is rather self-explanatory. The keyword "test=equiv" specifies that the power analysis is for an equivalence test. The equivalence bounds (ln(0.5) and ln(2)) are specified by the keywords "lower" and "upper", respectively. The standard deviation of $X_{i2} - X_{i1}$ is communicated through the "std" keyword. Note that the SAS macro facility is extensively used to simplify inputting parameters for the power analysis. The result of the power calculation is displayed below.

```
      Computed Power

          N
 Index   Total  Power

  1        6    0.965
  2        7    0.989
  3        8    0.996
```
Hence the computed powers for N=6, 7, 8 are 0.965, 0.989, and 0.996, respectively.

EXAMPLE 2: UPPER ONE-SIDED TEST IN FIXED-SEQUENCE STUDY

Consider a rifampin interaction study to be conducted using a fixed-sequence design. In period 1, the candidate drug C will be administered as a single dose. After suitable washout, in period 2, single doses of 600 mg rifampin will be administered daily for 14 days, and on Day 14, 600 mg rifampin will be coadiminstered with the candidate drug. The parameter of interest is the $AUC_{0-\infty}$ for the candidate drug and the hypothesis is stated as "The true geometric mean $AUC_{0-\infty}$ ratio (Period 2 / Period 1) for the candidate drug C is greater than 0.5."

We again adopt the notation and model of Design 1 and let X_{i1} and X_{i2} represent midazolam $\ln(AUC_{0-\infty})$ in period 1 and 2, respectively. Then the hypothesis of interest can be formally stated as:

$$H_0 : \mu_D \leq \ln(0.5) \text{ versus } H_1 : \mu_D > \ln(0.5).$$

The objective here is to estimate the sample size that gives 90% power for a true GMR of 0.65. From a previous study, the within-subject variance σ^2 is estimated to be 0.0408. Since in the hypothesis testing framework, this is an upper one-sided test, one can use the "onesamplemeans" option with the keyword "sides=U", which specifies an upper one-sided test, to perform the power calculation.

```
%let sigma2=0.0408; * macro variable for the within-subject variance;
%let sigma_D=%sysevalf((2*&sigma2)**0.5); * macro variable for the standard
deviation of the difference;
%let log_pt_5=%sysfunc(log(0.5)); * macro variable for log(0.5);
%let log_true_gmr=%sysfunc(log(0.65)); * macro variable for log(0.65);
proc power;
   onesamplemeans test=t sides=U alpha=0.05
   nullmean=&log_pt_5 std=&sigma_D
   mean=&log_true_gmr
   ntotal=.
   power =0.9;
run;
```

The SAS output is displayed below:

```
Computed N Total

Actual    N
 Power    Total
0.909       12
```

Hence a total of 12 subjects are needed to achieve an actual power of 90.9%.

EXAMPLE 3: EQUIVALENCE TEST IN 2 X 2 CROSSOVER STUDY

Consider a definitive bioequivalence study to be conducted using a 2 x 2 crossover design, with the two treatments being two formulations (denoted as A and B) of the candidate drug C. The parameter of interest is $AUC_{0-\infty}$ and the hypothesis is stated as "The true geometric mean $AUC_{0-\infty}$ ratio for the candidate drug C (Formulation B / Formulation A) is between 0.80 and 1.25." We adopt the notation and model of Design 3 and let y_{ijk} denote the $\ln(AUC_{0-\infty})$ value collected at period k for the jth subject in sequence i. The hypothesis can therefore be formally stated as:

$$H_0 : \tau_B - \tau_A \leq \ln(0.80) \text{ or } \tau_B - \tau_A \geq \ln(1.25) \text{ versus } H_1 : \ln(0.80) < \tau_B - \tau_A < \ln(1.25).$$

In this study, the sample sizes in the two sequences will be equal ($n_1 = n_2 = n$). It is expected that the true geometric mean ratio (GMR) is 1.1. The objective is therefore to choose a sample size so that there is at least 95% probability to conclude equivalence given the true GMR is 1.1. Based on Design 3, the power analysis can be conducted using a two-sample t test with n subjects per group and a common variance $\frac{1}{2}\sigma^2$. From a prior study, the within-subject variance σ^2 is estimated to be 0.1003 and the "twosamplemeans" option can be used to perform the power analysis:

```
%let sigma2=0.1003; * macro variable for the within-subject variance;
%let std_derived=%sysevalf((&sigma2/2)**0.5); * macro variable for the derived
common standard deviation;
%let log_pt_8=%sysfunc(log(0.8)); * macro variable for log(0.8);
%let log_1_pt_25=%sysfunc(log(1.25)); * macro variable for log(1.25);
%let log_true_gmr=%sysfunc(log(1.1)); * macro variable for log(1.1);
proc power;
   twosamplemeans test=equiv_diff alpha=0.05
   lower=&log_pt_8 upper=&log_1_pt_25 std=&std_derived
   meandiff=&log_true_gmr
   npergroup=.
   power =0.95;
run;
```

Note that the keyword "test=equiv_diff" specifies an equivalence test for the population mean difference of two independent samples. The SAS output is displayed below.

```
Computed N Per Group

Actual    N Per
 Power    Group
0.952       68
```

Based on the calculation, a sample size of 68 per treatment sequence is needed to achieve an actual power of 95.2%.

EXAMPLE 4: LOWER ONE-SIDED TEST IN 2 X 2 CROSSOVER STUDY

Consider a ketoconazole interaction study to be conducted using a 2 x 2 crossover design. Denote the two treatments as A and B, respectively. For treatment A, 300 mg of the candidate drug C will be administered as a single dose. In treatment B, single doses of 400 mg kenotconazole will be administered daily for 5 days and a single 300 mg dose of the drug C will be co-administered on Day 2. The parameter of interest is $AUC_{0-\infty}$ for the candidate drug and the hypothesis is stated as "The true geometric mean $AUC_{0-\infty}$ ratio for the candidate drug C (B / A) is less than 2.0. " We adopt the notation and model of Design 3 and let y_{ijk} denote the $\ln(AUC_{0-\infty})$ value collected at period k for the jth subject in sequence i. The hypothesis can therefore be formally stated as:

$$H_0 : \tau_B - \tau_A \geq \ln(2.0) \text{ versus } H_1 : \tau_B - \tau_A < \ln(2.0).$$

The objective is to identify the true GMR that gives 80% power for n=4 per sequence. The within-subject variance σ^2 is estimated to be 0.03821. We therefore use the "twosamplemeans" option with the keyword "sides=L", which specifies a lower one-sided test, to perform the power calculation.

```
%let sigma2=0.03821; * macro variable for the within-subject variance;
%let std_derived=%sysevalf((&sigma2/2)**0.5); * macro variable for the derived
common standard deviation;
%let log_2=%sysfunc(log(2.0)); * marco variable for log(2.0);
%let log_true_gmr=%sysfunc(log(1.518)); * macro variable for log(1.518);
proc power;
    twosamplemeans test=diff side=L alpha=0.05
    nulldiff=&log_2 std=&std_derived
    meandiff=&log_true_gmr
    npergroup=4
    power =.;
run;
```

Note that Proc Power does not allow direct calculation of true mean difference given power and sample size. Hence in this example, the true GMR that gives 80% power can only be obtained by trial and error. The SAS output is displayed below.

```
Computed Power

    Power
    0.800
```

Therefore for n=4 per sequence and a true GMR of 1.518, the power is 80%.

4. POWER CALCULATION FOR HIGHER ORDER CROSSOVERS

First consider a 3 period crossover design with 3 treatments (denoted as A, B, and C) and 6 treatment sequences (see Diagram (a)). We adopt the same model (i.e, treatment and period as fixed effects, and subject as a random effect) and notation as those in Design 3. Without loss of generality, assume the parameter of interest is $\tau_B - \tau_A$. Let T denote the estimator for $\tau_B - \tau_A$. Using linear model theory, it can be shown that under the assumption of equal number of subjects per sequence, T can be expressed:

$$T = \frac{1}{6n} \sum_{d(i,k)=B} y_{ijk} - \frac{1}{6n} \sum_{d(i,k)=A} y_{ijk},$$

where n denote the number of subjects per treatment sequence. It follows that $\sigma_T^2 = \frac{\sigma^2}{3n}$, where σ^2 is the within subject variance.

Diagram (a). Three Period Crossover Design

Sequence	Period		
	1	2	3
1	A	B	C
2	A	C	B
3	B	A	C
4	B	C	A
5	C	A	B
6	C	B	A

The error degree of freedom r for the 3 period crossover design is:

$$r = df_{total} - df_{period} - df_{trt} - df_{sujbect}$$
$$=(18n-1)-2-2-(6n-1)$$
$$= 12n - 4.$$

All power calculations can be conducted through the "onesamplemeans" option in Proc Power by formulating the problem as a one-sample t test or equivalence test. The idea is to modify the input for "ntotal" and "std" so that the correct degrees of freedom r and variance σ_T^2 are entered. In a one-sample t test or equivalence test, the degrees of freedom is calculated as ntotal − 1. This suggests that we should set "ntotal" to $12n - 3$, which ensures that the degrees of freedom is equal to $12n - 4$. Since for a one-sample t test or equivalence test with "ntotal" set to $12n - 3$, the variance of the estimator is calculated as $\dfrac{\text{std}^2}{12n\text{-}3}$, the correct input for "std" can be obtained by solving the following equation:

$$\frac{\text{std}^2}{12n\text{-}3} = \sigma_T^2 = \frac{\sigma^2}{3n}.$$

Hence "std" should be set to $\left(\dfrac{4n-1}{n}\right)^{1/2}\sigma$.

For crossover designs with more than 3 periods, to be specific, we focus on the class of Williams designs (Williams (1949)), which are widely used in practice. Examples of layouts for 4, 5, and 6 period Williams designs are displayed in Diagrams (b), (c), and (d) respectively.

Diagram (b). Four Period Williams Design

Sequence	Period			
	1	2	3	4
1	A	D	B	C
2	B	A	C	D
3	C	B	D	A
4	D	C	A	B

Diagram (c). Five Period Williams Design

Sequence	Period				
	1	2	3	4	5
1	A	E	B	D	C
2	B	A	C	E	D
3	C	B	D	A	E
4	D	C	E	B	A
5	E	D	A	C	B
6	A	B	E	C	D
7	B	C	A	D	E
8	C	D	B	E	A
9	D	E	C	A	B
10	E	A	D	B	C

Diagram (d). Six Period Williams Design

Sequence	Period					
	1	2	3	4	5	6
1	A	F	B	E	C	D
2	B	A	C	F	D	E
3	C	B	D	A	E	F
4	D	C	E	B	F	A
5	E	D	F	C	A	B
6	F	E	A	D	B	C

For the designs in (b), (c), and (d), provided that there are equal number of subjects in each sequence, the estimator of $\tau_B - \tau_A$ can be expressed in the form of:

$$T = \frac{1}{sn} \sum_{d(i,k)=B} y_{ijk} - \frac{1}{sn} \sum_{d(i,k)=A} y_{ijk},$$

where n denote the number of subjects per treatment sequence and s the number of sequences. Hence the "onesamplemeans" option in Proc Power can again be used to perform power calculations and the appropriate input for "ntotal" and "std" can be similarly derived. The results are summarized in the table below.

Design	ntotal	std
Three period (six sequences)	$12n-3$	$\left(\frac{4n-1}{n}\right)^{1/2}\sigma$
Four period Williams design	$12n-5$	$\left(\frac{12n-5}{2n}\right)^{1/2}\sigma$
Five period Williams design	$40n-7$	$\left(\frac{40n-7}{5n}\right)^{1/2}\sigma.$
Six period Williams design	$30n-9$	$\left(\frac{30n-9}{3n}\right)^{1/2}\sigma$

We next present two examples.

EXAMPEL 5: EQUIVALENCE TEST IN 3 PERIOD CROSSOVERS

Consider a formulation study for the candidate drug C to be conducted using the 3 period crossover design in Diagram (a). The three treatments are Phase I formulation, Phase II formulation, and Phase III formulation. The parameter of interest is $AUC_{0-\infty}$ for the candidate drug and the primary hypothesis is stated as "The true geometric mean $AUC_{0-\infty}$ ratio for the candidate drug C (Phase III formulation / Phase II formulation) is between 0.70 and 1.43. "

The study plans to enroll equal number of subjects per treatment sequence. Let y_{ijk} denote the $\ln(AUC_{0-\infty})$ value collected at period k for the jth subject in sequence i. The objective is to calculate the power given n=3 per treatment sequence and a true GMR of 1.20. The within-subject variance σ^2 is estimated to be 0.0389. By the discussion in Design 4, the power analysis can be conducted using the "onesamplemeans" option by setting "ntotal" to $12n-3$ and "std" to $\left(\frac{4n-1}{n}\right)^{1/2}\sigma$, where n is the number of subject per treatment sequence.

```
%let n_per_seq=3; * macro variable the number of subjects per treatment sequence;
%let sigma2=0.0389; * macro variable for the within subject variance;
%let ntotal_derived=%sysevalf(12*&n_per_seq-3); * macro variable for the derived
ntotal in a 3 period crossover;
%let std_derived=%sysevalf(((4-1/&n_per_seq)*&sigma2)**0.5); * macro variable for
the derived std in a 3 period crossover;
%let log_pt_7=%sysfunc(log(0.70)); * macro variable for log(0.70);
%let log_1_pt_43=%sysfunc(log(1.43)); * macro variable for log(1.43);
%let log_true_gmr=%sysfunc(log(1.20)); * marco variable for log(1.20);
proc power;
    onesamplemeans test=equiv alpha=0.05
    lower=&log_pt_7 upper=&log_1_pt_43 std=&std_derived
    mean=&log_true_gmr
```

```
        ntotal=&ntotal_derived
        power =.;
    run;
```

The power is 83.3% for n=3 per sequence and a true GMR of 1.20.

```
    Computed Power

        Power

        0.833
```

EXAMPEL 6: UPPER ONE-SIDED TEST IN 5 PERIOD CROSSOVERS

Consider a formulation study for the candidate drug C that compares one old formulation with four new formulations. The study will be conducted using a 5 period crossover design (Diagram (c)) and the parameter of interest is $AUC_{0-\infty}$ for the candidate drug. The primary hypothesis can be stated as "The true geometric mean $AUC_{0-\infty}$ ratio (New formulation 1 / Old formulation) is above 0.70". The study plans to enroll equal number of subjects per treatment sequence. Let y_{ijk} denote the $\ln(AUC_{0-\infty})$ value collected at period k for the jth subject in sequence i. The objective is to calculate the power given n=2 per treatment sequence and a true GMR of 0.80. The within-subject variance σ^2 is estimated to be 0.0285. By the discussion in Design 4, the power analysis can be conducted using the "onesamplemeans" option by setting "ntotal" to $40n-7$ and "std" to $\left(\dfrac{40n-7}{5n}\right)^{1/2}\sigma$, where n is the number of subject per treatment sequence.

```
    %let n_per_seq=2; * macro variable the number of subjects per treatment sequence;
    %let sigma2=0.0285; * macro variable for the within subject variance;
    %let ntotal_derived=%sysevalf(40*&n_per_seq-7); * macro variable for the derived
    ntotal in a 5 period crossover;
    %let std_derived=%sysevalf(((8-7/5/&n_per_seq)*&sigma2)**0.5); * macro variable for
    the derived std in a 5 period crossover;
    %let log_pt_7=%sysfunc(log(0.70)); * marco variable for log(0.70);
    %let log_true_gmr=%sysfunc(log(0.80)); * macro variable for log(0.80);
    proc power;
        onesamplemeans test=t side=U alpha=0.05
        nullmean=&log_pt_7 std=&std_derived
        mean=&log_true_gmr
        ntotal=&ntotal_derived
        power =.;
    run;
```
The power is 79.8% for n=2 per sequence and a true GMR of 0.80.

```
    Computed Power

        Power
        0.798
```

5. DISCUSSION

In this paper, we demonstrate that many hypothesis testing problems in clinical pharmacology studies can be reformulated as a one- or two-sample t-test or equivalence test. Thereby one can use SAS Proc Power to perform the power analysis. Sample programs for different hypothesis testing problems are presented in a variety of study design settings.

Besides SAS, NQuery and PASS are also widely used in practice for power calculations. The power analyses for one and two-sample t-tests and equivalence tests in two-sample parallel group design and 2 x 2 crossover design are straightforward in both packages. However, neither packages offer a direct solution for power calculations in one-sample equivalence tests[1]. To the extent possible, the calculations in all examples have been validated using NQuery and PASS. We conclude this paper by outlining several advantages of using SAS Proc Power versus NQuery and PASS to perform power analysis. First, since one often obtains variance estimates by analyzing relevant clinical data in SAS, it is more streamlined and convenient to conduct power analysis in SAS as well. Second, the documentation in SAS is more comprehensive than those in NQuery and PASS: this helps the users better understand the procedure and hence the users are less likely to make mistakes. Third, as demonstrated in the examples, one can take advantage of the programming capabilities in SAS and make inputting parameters for power analysis more

---

[1] NQuery has no such capability directly. In order to cover more distributions, PASS uses a simulation approach even though a closed-form solution is available for one-sample equivalence test based on t statistic.

straightforward. Last, but not least, using SAS can better help the users keep a record of the power analysis, which is helpful for adapting the code for similar problems in the future.

REFERENCES

- Berger, R. L. and J. C. Hsu (1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets. Statistical Science 11 (4), 283-319.

- Hills, M. and P. Armitatge (1979). The two-period cross-over clinical trials. British Journal of Clinical pharmacology 8, 7-20.

- Owen, D. B. (1965). A special case of bivariate non-central t-distribution. Biometrica 52, 437-446.

- Phillips, K. F. (1990). Power of the two one-sided tests procedure in bioequivalence. Journal of Pharmacokinetics and Biopharmaceutics 18, 137-144.

- Schuirman, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. Journal of Pharmacokinetics and Biopharmaceutics 15, 657-680.

- Westlake, W. J. (1972). Use of confidence intervals in the analysis of comparative bioavailability trials. Journal of Pharmaceutical Sciences 61, 1340-1341.

- Williams, E. J. (1949). Experimental designs balanced for estimation of residual effects of treatments. Australian Journal of Scientific Research 2, 149-168.

ACKNOWLEDGMENTS

CONTACT INFORMATION

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

Name: Peng Sun
Enterprise: Merck & Co., Inc.
Address: PO Box 1000
City, State ZIP: North Wales, PA 19454
Work Phone: 267-305-7419
E-mail: peng_sun@merck.com

---

[2] Biostatistics and Research Decision Sciences, Merck & Co., Inc.