

Paper 322-2012

## A Simple Way to Program Power Calculations

Rober Lew<sup>1</sup>, Hongsheng Wu<sup>1,2</sup>

<sup>1</sup>MAVERIC, VA Cooperative Studies Program, Boston V.A. Healthcare System,  
Boston, MA, United States

<sup>2</sup>Department of Computer Science and Systems, Wentworth Institute of  
Technology, 550 Huntington Ave, Boston MA, United States

### Abstract

We propose simulating large sets of statistical power calculations and selecting from a long list of study designs those with appropriate power. Typical commercial programs fill in all but one parameter and solve for the remaining parameter. We avoid solving equations and thereby obtain power for atypical designs. First, for simplicity, we compute the power for the test comparing two proportions without assuming either approximate normality or a common variance. Then, we extend the log-rank test power calculation to account for various patterns of on-study censoring, study duration, and patient recruitment period. We generate a new approximation for the censoring adjustment. In so doing, we set out a easy way for others to modify our program for other related power calculations.

### 1. Background

Many standard power calculation formulas simplify calculations, for example, the two-sample test comparing proportions that assumes a common variance and approximate normality. This paper formalizes a simulation approach for obtaining more exact results. For the log-rank test it facilitates calculations that account for various censoring patterns, study durations, and enrollment periods. Many statisticians simulate power as we do, so that our method is not new. However, elaborating a simple but general programming structure for such calculations may facilitate the calculation of power in atypical settings.

Older textbooks on the analysis of variance [1] contain graphs of power functions. One interpolates within a region of the graph to obtain an approximate power. Conceptually our approach is the same. We simulate a large dataset of power calculations or power profiles and select a subset to close in on the desired parameter values. Rather than interpolate, we rerun the program with a finer grid of values within the region of interest as if magnifying this region of the graph.

Unlike a graph, the number of parameters can be large; as in the setting of survival analysis; two hazard rates, two sample sizes, types I and II error, and length of the patient recruitment period. We require knowing the explicit power formula. The SAS® examples are nested do-loops. This facilitates adaptation of our program logic into

another programming language. The datasets of interest can be sorted and displayed in a user-chosen format such as a graph.

Our approach complements commercial packages by narrowing vague requests. For example, consider calculating power for a case-control study where the requested odds ratio is 2.0. By generating a large set of possible designs with pairs of proportions and sample sizes associated with odds ratios near 2.0, investigators can narrow the choices. The large set indicates the robustness of the narrower design.

The paper has five sections. Section 2 treats the problem of the difference between proportions to show the approach in a simple familiar setting. Section 3 discusses life table power calculations and addresses adjustment for on-study censoring. For this problem we consider approximations that either minimize or add flexibility to parametric assumptions. Section 4 provides some rationale for the approximations. Section 5 is discussion.

## 2. Comparing Means and Proportions

### 2.1 Basic Framework and Notation

For normally distributed data the null hypothesis of no difference (H0) opposes the one-sided alternative hypothesis (HA) of a pre-specified difference  $d > 0$ . Let  $\delta$  denote the random variable 'difference'. Under H0 let  $P(\delta \geq c) = \alpha$ . Under HA let  $P(\delta \leq c) = \beta$ . The critical value,  $c$ , lies between 0 and  $d$ . For the difference,  $\delta$ , assume the variance is  $\sigma^2$  under both H0 and HA. This implies that  $d - c = \sigma z_\beta$  and  $c = \sigma z_\alpha$  and hence the power equation is  $d = \sigma (z_\beta + z_\alpha)$ .

For binomial data let the two groups have probabilities of success,  $p$  under H0 and  $p+d$  under HA. For one observation per group the difference between proportions,  $\delta$ , has variance  $\sigma_0^2 = 2p(1-p)$  under H0 and variance  $\sigma_A^2 = p(1-p) + (p+d)(1-p-d)$  under HA. (The chi-square-test formulation differs in that  $p + d/2$  replaces  $p$  under H0).

The power equation is then:

$$d = \sigma_A z_\beta + \sigma_0 z_\alpha.$$

With unequal sample sizes,  $n$  and  $m$ ;

$$\sigma_0^2 = p(1-p)/n + p(1-p)/m \quad \text{and} \quad \sigma_A^2 = p(1-p)/n + (p+d)(1-p-d)/m,$$

A program that solves for  $d$ , given  $\sigma_A^2$ , must solve a quadratic equation. To avoid any type of non-linear equation we always solve for  $z_\beta$ . To illustrate, we fix all parameters except  $p$ ,  $d$ , and  $z_\beta$ :

*Data powerprofiles;*

```

Z_alpha = 1.96;    n = 200;  m = 400;  *two-sided test at 5%;
Do ip = 1 to 10;
  p = 0.10 + 0.05*ip;
  Do id = 1 to 5;
    d = 0.04 + 0.04*id;
    var0 = p*(1-p)/n + p*(1-p)/m;
    varA = p*(1-p)/n + (p+d)*(1-p-d)/m;

    Z_beta = ( d - z_alpha*sqrt(var0) )/ sqrt(varA);
    Output;
  End;
End

```

Generalizing, with loops for all parameters except Z-beta, the program would then produce a large dataset. In SAS, to select a subset with  $Z\_beta \sim 1.645$  (for example), we code as follows:

```
Data powerprofile_1.645; set powerprofiles; if z_beta > 1.6 and z_beta < 1.7;
```

Poor choice for the ranges for the loops might result in an empty dataset, easily resolved by widening these ranges. Otherwise, the empty set connotes an infeasible design.

## 2.2 Binomial Data: More Precise Calculations

Fisher's exact test is often used when the normal approximation to the binomial distribution is a poor fit. Thus, the exact binomial solution that follows merely illustrates some general issues. When  $\delta$  is a discrete random variable then the equation,  $P(\delta \geq c) = \alpha$ , may not hold without a randomized critical region [2]. Our approach only yields exact binomial probabilities close to  $\alpha$ . When generating the power equation from the equations,  $d - c = \sigma_A z_\beta$  and  $c = \sigma_0 z_\alpha$ , proves awkward, then we also simulate a range of values of  $c$ .

The calculation of  $P(\delta \geq c)$  dictates the loop structure. Fixing the parameter values for the outer loops, then, given  $d$ , we explore values of the critical value,  $c$ , between 0 and  $d$ .

```

m = 6;  n = 7;  p = 0.2;  d = 0.6;
Do ic = 1 to 9;
  c = ic*d/10;
  calculate one-sided type I error
  calculate power
  Output;
End

```

Discrete variables introduce issues. With two independent binomial variables,  $x$  and  $y$ , the difference could be defined as an integer  $\delta = x - y$ . The formal task is to find  $c$  for which the type I error is  $P(\delta \geq c) = \alpha$  under  $H_0$  and the power  $P(\delta \geq c) = 1 - \beta$  under  $H_A$ . By choosing a non-integer critical value,  $c$ , we need not distinguish between the potentially inconsistent equations,  $P(\delta > c) = \alpha$  and  $P(\delta \geq c) = \alpha$ . We choose to study the difference-between-proportions case by defining  $\delta = y/n - x/m$ . In this case, if the product,  $mnc$ , is non-integer, we can safely set  $P(\delta \geq c) = P(my - nx \geq mnc)$ .

To find the type I error given  $c$ , let the difference between proportions,  $\delta$ , satisfy  $\delta = y/n - x/m$ , where  $y$  and  $x$  are binomial distributions; specifically, under  $H_0$ ,  $y$  is  $\text{Bin}(p, n)$  and  $x$  is  $\text{Bin}(p, m)$ . Thus,

$$1 - P(\delta \geq c) = P(y/n - x/m < c) = E_x ( P(y < nc + nx/m) )$$

where  $E_x$  is the expected value over the range of  $x$  from 0 to  $m$ .

Given  $x = k$ ,  $P(y < nc + nk/m) = P(y \leq G(k))$ ,

where  $G(k) = \text{Greatest integer in } (nc + nk/m)$ .

Thus,

$$1 - \alpha = P(\delta < c) = \sum P(y \leq G(k) | n, p) P(x = k | m, p) \quad \text{where } k = 0, 1, \dots, m.$$

Finally, consider the type II error =  $1 - \text{power}$ . Under  $H_A$ ,  $y$  is  $\text{Binomial}(p+d, n)$ . Thus, the complement of the power equals

$$P(\delta < c) = \sum P(y \leq G(k) | n, p + d) P(x = k | m, p) \quad \text{where } k = 0, 1, \dots, m.$$

Appendix A provides a SAS program and results.

Asymmetric distributions and two-sided hypotheses complicate matters. Thus, we consider only one-sided tests and simple alternatives, consistent with a 'simple vs. simple' hypothesis test that compares results assuming that the distribution of  $\delta$  is symmetric and either centered at 0 under  $H_0$  or at  $d > 0$  under  $H_A$ . To calculate the power for a two-sided test we typically need to specify the value for  $\delta$  under  $H_A$  as either  $d$  or  $-d$ . A two-sided test posits choices of  $\delta = d$  and of  $\delta = -d$ . Without symmetry an extra logical principle may be needed to decide which choice determines the power.

### 3. Life Table Power Calculation With Censoring During the Study Period

#### 3.1 Background

Losses to follow-up during a study can reach levels of 20% or more, as in a 5-year study of elderly subjects with non-melanoma skin cancer. We could either follow fewer patients for a longer time or more patients for a shorter time. Besides the administrative censoring that occurs when a study ends, we needed to compare the efficiency and costs of these designs in terms of the on-study censoring. This clinical trial had ‘rolling’ enrollment; that is, enrollment of subjects for several years. Hence, in a 5-year study with 2 years of enrollment the first subject to enroll would have up to 5 years of follow-up and the last subject to enroll would have up to 3 year of follow-up.

For time-to-event models let  $Y$  denote an outcome event and  $C$  denote a censoring event. Assume that  $C$  is independent of  $Y$ . Dependence between  $Y$  and  $C$  or non-random censoring calls for sensitivity analyses to model the distribution of the unobserved missing data[3, 4]. Our approach makes it easy to add do-loops and explore missing value patterns using a few simple explicit discrete, subjectively chosen models for missing data.

This section presents a simplified formula for the probability that an outcome is observed.

$$P(\text{Outcome } Y \text{ is observed}) = \sum_{\text{time } t} P(Y \text{ occurs at } t \text{ and } C \text{ occurs after } t)$$

The sum is over all times  $t$  where  $0 < t \leq T(i)$  where  $T(i)$  is the maximal time-on-study for subject  $i$ , as opposed to the study duration,  $S$ . Assume  $1 \leq T(i) \leq S$ . Then if the times,  $T(i)$ , are symmetrically distributed about a midpoint,  $m$ , then the mean and median time-on-study is  $m = (S + 1)/2$ . For asymmetric distributions, choose  $m$  as the mean of median time, as appropriate.

For each subject a heuristic approximation is:

$$P(\text{Outcome } Y \text{ is observed} \mid T(i)) \approx$$

$$P(\text{Outcome } Y \text{ occurs before } T(i)) * (P(C > \text{mean time to a } Y\text{-event} \mid T(i))).$$

The mean time to a  $Y$ -event refers only to subjects who have  $Y$ -events while on study, a conditional mean excluding all those who finish the study event-free. Hence, this mean time is less than  $T(i)$ .

Section 4 develops a simple approximation for the entire sample, not just one subject, detailing the simplifications in the exact formula. This facilitates creations of new approximations appropriate to more complex situations, such as  $C$  dependent on  $Y$ .

The approximation is thus a ‘**rule of thumb**’:

$$P(\text{Observe } Y) \approx (1 - \exp(-\lambda m)) * (1 - \frac{am}{2}) \quad (1)$$

where  $m$  is the mean or median time on study, the rate,  $a$ , is the annual proportion of censored events, and the hazard rate,  $\lambda$ , is the hazard rate in an exponential distribution. One could replace the exponential distribution with a more realistic distribution.

### 3.2 Power for the Log-rank Test

We first assume underlying exponential distributions for the power formula[5-7] for the two-sample log-rank test and ignore on-study censoring. Appendix B contains a fuller program.

```
Data powerprofiles;
Z_alpha = 0.05; ny = 50; nx = 50;

  do ix = 0 to 50;      lamda_x = 0.005 + 0.001*ix;
  do iy = 1 to 50;      lamda_y = lamda_x + 0.001*iy;

    haz_ratio =      lamda_y/lamda_x; *Force lamda_y > lamda_x;

    py = 1 - exp(-lamda_y);          *One year death rates;
    px = 1 - exp(-lamda_x);
    differ = py -px;

      *Compute Z_beta;
      output;
      end;
  end;

data slice;
set powerprofiles; if haz_ratio >1.5 and haz_ratio <2.5;
proc means n min max; var lamda_x lamda_y px py differ power;
```

Note that the suffixes 'y' and 'x' mark the two samples, 'lamda' denotes the hazard rate, and 'px' and 'py' denote the one-year death rates assuming 'year' is the unit of time. In the last line, the 'proc means' statement provides a summary of the profiles in the subset. The proportions help those unfamiliar with hazard rates to evaluate the designs.

To compute  $Z_{\beta}$  we insert in the innermost loop the additional lines:

```
dx = nx*px;      * deaths among nx subjects in arm x;
dy = ny*py;      * deaths among ny subjects in arm y;
dxy = sqrt( dx*dy/(dx + dy));
z_beta = log(haz_ratio)*dxy - z_alpha;
power = probnorm(z_beta);      * P( Z < Z_beta);
power = round( 100*power );    *power as percentage;
```

We add another loop to account for on-study censoring. Suppose the maximum follow-up time is one year ( $T = 1$ ) and we specify  $P(\text{Censored before time } T) = 20\%$  in the first line as `'p_censor = 0.2;'`.

Now, with changes within in boldface, the innermost loop might become:

```
Do icensor = 1 to 9;
  pc_meantime = icensor*P_censor/10; q = 1 - pc_meantime;

  dx = nx*px*q;      * deaths observed among nx subjects in arm x;
  dy = ny*py*q;      * deaths observed among ny subjects in arm y;
  dxy = sqrt( dx*dy/(dx + dy));
  z_beta = log(haz_ratio)*dxy - z_alpha;
  power = probnorm(z_beta);      * P( Z < Z_beta);
  power = round( 100*power );    *power as percentage;
  output;
end;
```

This last loop merely takes a robust look at the rule of thumb in approximation (1) by varying `pc_meantime` around 0.5 from 0.1 to 0.9. Note that the adjustment for censoring does not depend on distributional assumptions or on secondary analyses of existing data other than setting a value for `p_censor`.

In this case, variations in on-study censoring matter. Under the rule of thumb the proportion censored before the mean time to an event should be  $10\% = p\_censor * 0.5 = 20\% * 0.5$ . The program in Appendix B slices the large dataset to find  $N = 24$  power profiles with 90% power and differences between proportions of 8%. The values of the proportions `px` and `py` are close to zero. The program in Appendix B that produces the table of means below shows that we can only allow about a 5% rate of events censored on-study, not the rule-of-thumb 10% rate.

#### **Output From the SAS MEANS Procedure (Proc Means)**

Variable	N	Minimum	Maximum	Mean
haz_ratio	24	2.6400000	2.8400000	2.7462500
pc_meantime	24	0.0200000	0.1000000	0.0491667
px	24	0.0478189	0.0535149	0.0503134
py	24	0.1297720	0.1349777	0.1320518
differ	24	0.0800000	0.0800000	0.0800000
power	24	90.0000000	90.0000000	90.0000000

The rule of thumb does better with larger sample sizes and proportions `px` and `py` further away from zero; in other words, the mean of the variable `pc_meantime` is much closer to 10%.

### 3.3 Varying the Length of the Study and of the Recruitment Period

In a proposed 5-year study of the incidence of non-melanoma skin cancer in a high-risk population, we considered varying both the length of the study and the length of the recruitment period. For example, consider uniform rates of recruitment for two years (Rtime) and follow up of all patients until the study ends at  $S = 5$  years. The first patient entered would then have up to  $S$  years of follow-up and the last patient entered would have  $S - Rtime$  years of follow-up. On average patients would then have up to  $S - (Rtime/2)$  years of follow-up. For the proposed study with  $S = 5$  and  $Rtime = 2$ , this average was 4.

Varying  $S$  and  $Rtime$  within practical limits indicates which time-to-event designs with 90% power might minimize cost. We chose the limits  $4 \leq S \leq 6$  and  $1 \leq Rtime \leq 3$ . We assumed an underlying exponential distribution for the incidence of non-melanoma skin and a cumulative incidence of 30% at 5 years.

This adds two loops to the example of a program for survival analysis power in section 3.5, an  $S$ -loop and an  $Rtime$ -loop. With average follow-up time,  $avefup = S - Rtime/2$ , the major change in the SAS code is:

```
py = 1 - exp(-lamda_y*avefup); *rates at mean follow-up time;
px = 1 - exp(-lamda_x*avefup);
```

For the planned study expected on-study 5-year losses to follow-up were about 10%. We needed to enroll a total of about 900 patients to obtain 800 evaluable patients. For a 6-year study we needed to enroll about 700 patients to obtain about 600 evaluable patients. We chose  $Rtime = 2$  because the power was insensitive to this parameter. Data from a previous study indicated that we could enroll about 50 patients per year at a typical site and 100 patients in two years. Thus, we compared the costs of a 6-year 7-site study to a 5-year 9-site study.

## **4. Rationale for approximations**

### 4.1 Probability of observing an event

The number of observed events (as opposed to the number of subjects enrolled) drives the power calculation for the log-rank test.

Let  $Y$  denote the outcome event and  $C$  denote the censoring event. Assume that  $C$  is independent of  $Y$ . Let 'E' denote 'expected value'. Then,

$$P(\text{observe event } Y) = E_i(P(\text{observe } Y \text{ for subject } i)).$$

With rolling enrollment, each subject has a maximum follow-up period or a time-on-study,  $T(i)$ . Hence, we fix the subject and suppress the index 'i' for now. Then,



$$P(\text{observe } Y(i) \mid T(i)) = P(\text{observe } Y \mid T).$$

By independence, for each subject

$$P(\text{Observe } Y \mid T) = \int_0^T P(C > t) f_y(t) dt = \int_0^T (1 - F_C(t)) f_y(t) dt$$

where  $f_y(t)$  is the probability density function of a Y-event at time  $t$

and  $F_C(t)$  is the cumulative probability of a censoring event on or before time  $t$ .

Let  $F_y(T) = P(y \leq T) = \int_0^T f_y(t) dt$  and define  $f_y(t \mid T) = \frac{f_y(t)}{F_y(T)}$ .

Thus,

$$P(\text{Observe } Y \mid T) = F_y(T) (1 - E_{y \mid T}(F_C(t)))$$

where  $E_{y \mid T}(F_C(t))$  is the conditional expectation of  $F_C(t)$  over the distribution of Y-event times during the interval,  $[0, T]$ , with density function,  $f_y(t \mid T)$ . Let  $\theta(T)$  be the mean time to a Y-event for this density function on the interval,  $[0, T]$ . Note that  $\theta(T) \leq T$ . If Y-events are uniform on  $[0, T]$  then  $\theta(T) = T/2$ . If exponential with hazard rate,  $\lambda$ , then  $\theta(T) < T/2$  because the density function is a strictly decreasing function.

$$\text{Assume } E_{y \mid T}(F_C(t)) \approx F_C(E_{y \mid T}(t)) = F_C(\theta(T)).$$

In other words,  $P(\text{Observe } Y \mid T)$

$$\approx P(\text{Event } Y \text{ occurs before } T \mid T) * P(\text{Censoring occurs after time } \theta(T) \mid T).$$

The approximation is exact if  $F_C(t)$  is linear in  $t$  on the interval,  $[0, T]$ .

#### 4.2 Convenient Assumptions

Assume, for convenience, that Y-events follow an exponential distribution with hazard rate,  $\lambda$ . Also, assume that,  $\theta(T) = T/2$ , a conservative assumption with respect to the exponential distribution. Then,

$$P(\text{Observe } Y \mid T)$$

$$\begin{aligned} &\approx P(\text{Event } Y \text{ occurs before } T | T) * P(\text{Censoring occurs after time } T/2 | T) \\ &\approx (1 - \exp(-\lambda T)) * (1 - \frac{aT}{2}) \end{aligned}$$

where  $a$  is the annual rate of censored events during the study and the term,  $aT/2$ , approximates the probability of censoring before time  $T/2$ . The choice of the exponential assumption merely simplifies the next step, the aggregation over the subjects.

Consider a uniform distribution of rolling enrollment times for which the on-study follow-up times,  $T$ , range from a minimum of 1 to a maximum of  $S$ . To obtain the expected value over the uniform distribution, first integrate the expression above from 1 to  $S$  and then divide by  $(S-1)$ . The integral is:

$$(S - 1) \left( 1 - \frac{a(S+1)}{4} \right) - \frac{e^{-\lambda} - e^{-S\lambda}}{\lambda} + \frac{a}{2} \left( \frac{e^{-\lambda} - Se^{-S\lambda}}{\lambda} \right) + \frac{a}{2} \left( \frac{e^{-\lambda} - e^{-S\lambda}}{\lambda^2} \right).$$

The median and mean value of  $T$  is  $m = (S + 1)/2$ . By the mean value theorem, the integral above is approximately the range of the integral multiplied by the integrand evaluated at the midpoint.

$$(S - 1) (1 - \exp(-\lambda(S + 1)/2)) * (1 - \frac{a(S+1)}{4}).$$

We computed the exact integral and the approximate integral for values  $S = 2, 3$ , and  $4$ , for  $a = 0.12$  to  $0.20$  by  $0.02$ , and for  $\lambda = 0.2$  to  $2.0$  by  $0.2$ . The approximation was an upper bound for the exact integral and the maximum relative error, defined as  $(\text{exact} - \text{approximate})/\text{exact}$ , was  $0.063$ .

With  $m = (S + 1)/2$ , we obtain the approximation (1) in section 3.1:

$$P(\text{Observe } Y) \approx E_T(1 - \exp(-\lambda T)) * (1 - \frac{aT}{2}) \approx (1 - \exp(-\lambda m)) * (1 - \frac{am}{2}).$$

Here,  $m$  is the mean or median time-on-study, the rate,  $a$ , is the expected annual proportion of censoring events, and the hazard rate,  $\lambda$ , is the hazard rate in an exponential distribution. If the minimal and maximal times-on-study were  $s$  and  $S$  then  $m = (s + S)/2$ . Thus, the choice of  $m$  automatically scales the times.

To illustrate let the annual rate of censored events is 20%, the median time-on-study is 2.0 years, and 40% of subjects have an event within one year. Then  $a = 0.20$ ,  $m = 2.0$ , and  $\lambda = 0.51$  (obtained by solving:  $1 - \exp(-\lambda) = 0.4$ ). Then,  $P(\text{Observe } Y) \approx (1 - \exp(-0.51 * 2.0)) * (1 - 0.2 * \frac{2.0}{2}) = 0.6394 * 0.8 = 0.51152$ .

## 5. Discussion

Our approach provides an organized review of many possible choices for designs. The precise power formulas matter less than the review. However, the user must know the formulas rather than rely on a package to use a correct formula.

Fluency with many statistical packages and computer languages might easily provide an adequate review. Resources include NQUERY, PASS, PROC POWER in SAS version 9, the library of power calculations associated with package/languages such as STATA, STATISTICA, S-PLUS, and R, and much online software distributed by industry and academia. Less fluent statisticians might prefer our approach. Complex settings engender more complex formulas not easily found in these packages. Menu-driven software may disappoint the user who fills in all but one parameter in the formula and then expects the package to solve for the remaining one.

Tabulation with loop-friendly discrete distributions avoids solving non-linear equations. One can also make notes in the program giving the formula in Arabic and Greek letters, citing the source, and showing the correspondence between the Greek variable names in a formula and the Arabic variable names in the program code.

The detailed steps in approximating the probability of observing an outcome facilitates improved approximations. For example, in discussing on-study censoring, we noted the ease of extending from the exponential to another distribution such as the Weibull distribution[8]. Also, we can use do-loops to explore simple discrete distributions of events when no obvious parametric family of distributions comes to mind. In contrast to an exponential distribution for which 30% of patients have an event in year one, one might consider a discrete distribution that had no events at baseline (time 0) and 10% of events at 4, 8, and 12 months after baseline. The loop could cycle through a variety of distributions on the time-grid of 0, 4, 8, and 12 months.

Our approach does not help to choose the right formula. Jung and Hu assert that a better formula is needed for survival analysis power calculations for three or more groups. Power calculations can incorporate covariate effects[9]. Group sequential designs are far more complex[10]. Right-censored events require special methods[11].

In summary we have presented a simple programming approach to making atypical power calculations, applied the approach to adjustment for on-study censoring in survival analysis, and put forth an approximation and rule of thumb to simplify the adjustment.

## APPENDIX A: Exact Binomial Tails

The power calculation for binomial data,  $x$  and  $y$ , with respective sample sizes,  $m$  and  $n$ , evaluates the random variable,  $\delta = y/n - x/m$ . The expression,  $P(\delta \geq c) = P(my - nx \geq mnc)$  will have the non-integer advantage provided the product,  $mnc$ , is not an integer. Thus, the user selects a range of values for  $c$ . The program adds to each value of  $c$  a tiny increment such as  $\text{eps} = \text{sqrt}(2)/10000mn$ .

Let  $G(U)$  denote the greatest integer,  $I$ , such that  $I \leq U$ . In SAS the function,  $\text{FLOOR}(U)$  yields  $G(U)$ .

Let  $P(k | n p)$  denote  $P(k \text{ successes in } n \text{ trials} | n \text{ and } P(\text{success}) = p)$ , the binomial density function.

The SAS system function **PROBBNML** calculates cumulative binomial probabilities. Hence,  $P(k | n p)$  is the difference between successive values of **PROBBNML**. **PROBBNML** returns a missing value if  $k < 0$  or  $k > n$ . To remedy this one adds code such as:

```
bp = PROBBNML(p,n,k) ; if k < 0 then bp = 0; if k > n then bp = 1;
```

### One-Sided Type I Error, $\alpha$

Let the difference between proportions,  $d$ , satisfy  $d = y/n - x/m$ , where  $y$  and  $x$  are binomial distributions; specifically, under  $H_0$ ,  $y$  is  $\text{Bin}(p, n)$  and  $x$  is  $\text{Bin}(p, m)$ . Since,  $1 - P(\delta \geq c) = P(\delta < c)$ , we evaluate

$$P(\delta < c) = P(y/n - x/m < c) = E_x ( P(y < nc + nx/m) )$$

where  $E_x$  is the expected value over the range of  $x$  from 0 to  $m$ .

Let  $x = k$  and consider the expression,  $P(y < nc + nk/m)$ . By choice of a non-integer value for  $cmn$

$$P(y < nc + nk/m) = P(y \leq G(k)) \text{ where } G(k) \text{ is greatest integer in } nc + nk/m.$$

Thus, the complement of the type I error,  $1 - \alpha$ , equals

$$P(\delta < c) = \sum P(y \leq G(k) | n, p) P(x = k | m p) \quad \text{where } k = 0, 1, \dots, m.$$

### Power, $1 - \beta$

Under  $H_A$ , the only difference (in boldface) between this and the preceding calculation is that now the binomial variable  $y$  is  $\text{Bin}(\mathbf{p+d}, n)$ . Thus, the complement of the power,  $\beta$ , equals

$$P(\delta < c) = \sum P(y \leq G(k) | n, p + d) P(x = k | m, p) \text{ where } k = 0, 1, \dots, m.$$

```

*Variable x is Bin(m,p) and y is Bin(n,p+d) with sample
sizes m < n;

Data powerprofiles;
  do m = 6 to 8;
    do in = 1 to 5; n = m + in;                * n > m;
      eps = sqrt(2)/10000*n*m;
    do ip = 0 to 6; p = 0.1 + 0.05*ip;
    do id = 0 to 4; d = 0.5 + 0.05*id; if p+d >= 1 then d = 0.5;

      do ic = 1 to 9;    c = ic*d/10 + eps;  *non-integer;

*Calculate one-sided type I error and power;
  sum_alpha = 0; sum_power = 0;

  do x = 0 to m;
    pcumx = probbnml(p,m,x);
    if x = 0 then px = pcumx;
    if x > 0 then px = pcumx - probbnml(p,m,x-1);

    k = floor( c*n + (x*n/m));

*NULL;
    if k >= 0 and k <= n then pcumy = probbnml(p,n,k);
    if k > n then pcumy = 1; if k < 0 then pcumy = 0;

    sum_alpha = sum_alpha + pcumy*px;
*ALTERNATIVE Hypothesis;
    if k >= 0 and k <= n then pcumy = probbnml(p+d,n,k);
    if k > n then pcumy = 1; if k < 0 then pcumy = 0;

    sum_power = sum_power + pcumy*px;
  end;

  alpha = 1 - sum_alpha;
  power = 1 - sum_power;
  Output;
end;    * ic-loop;
end;    * id-loop;
end;    * ip-loop;
end;    * in-loop;
end;    * m -loop;

data slice; set powerprofiles;
if alpha <= 0.054 and alpha >= 0.046 and power > 0.895 and power
< 0.905;

```

```
proc print noobs; var m n p d c alpha power;
```

Output from SAS Print Procedure (Proc Print)

m	n	p	d	c	alpha	power
6	7	0.20	0.60	0.300	0.053096	0.90087
6	8	0.15	0.55	0.275	0.048711	0.90196
6	8	0.15	0.60	0.300	0.048251	0.89930
6	11	0.10	0.50	0.250	0.048574	0.90156
8	10	0.15	0.55	0.275	0.053133	0.90207
8	10	0.30	0.60	0.360	0.053464	0.90072

### APPENDIX B: Program for Power Analysis Accounting for On-Study Censoring

The program below requires the hazard rate lamda\_y to exceed the hazard rate lamda\_x. The 'proc means' statements produce the correlation and the table of output in section 3.2.

```
*Assume exponential distribution. Time in years;
*Then P(die before time t) = 1 - exp(-lamda*t);
*Assume 20% of patients censored on-study - p_censor = 0.2;

Data powerprofiles;
Z_alpha = 0.05; ny = 50; nx = 50; p_censor = 0.2;

do ix = 0 to 50; lamda_x = 0.005 + 0.001*ix;
do iy = 1 to 50; lamda_y = lamda_x + 0.01*iy;

haz_ratio = lamda_y/lamda_x; *Force lamda_y > lamda_x;
py = 1 - exp(-lamda_y); *One year death rates;
px = 1 - exp(-lamda_x);
differ = py -px;

*Compute Z_beta;
Do icensor = 1 to 9;
pc_meantime = icensor*P_censor/10; q = 1 - pc_meantime;

dx = nx*px*q; * deaths among nx subjects in arm x;
dy = ny*py*q; * deaths among ny subjects in arm y;
dxy = sqrt( dx*dy/(dx + dy));
z_beta = log(haz_ratio)*dxy - z_alpha;
power = probnorm(z_beta); * P( Z < Z_beta);
power = round( 100*power ); *power as percentage;
output;
end;
end;
end;
```

```

data slice; set powerprofiles;
if haz_ratio > 2 and haz_ratio < 3.2 and power > 89 and power < 91 and
differ > 0.078 and differ < 0.082;

haz_ratio = 0.01*round(100*haz_ratio);
differ = 0.01*round(100*differ);

proc sort; by haz_ratio;
proc means n min max mean;
    var haz_ratio pc_meantime px py differ power;

```

## References

1. Scheffé, H., *The Analysis of Variance*. A Wiley publication in mathematical statistics, ed. Wiley. 1959, New York.
2. D.R. Cox, D.V.H., *Theoretical Statistics* 1979, New York: Halsted Press.
3. Daniels, M.J., Hogan, Joseph W. , *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis (Chapman & Hall/CRC Monographs on Statistics & Applied Probability)*. 2008, Boca Raton: Chapman & Hall/CRC.
4. Geert Molenberghs, M.K., *Missing Data in Clinical Studies (Statistics in Practice)* 2007.
5. Buyse, M.E., M.J. Staquet, and R.J. Sylvester, *Cancer clinical trials : methods and practice*. Oxford medical publications. 1984, Oxford ; New York: Oxford University Press. xiv, 481 p.
6. Collett, D., *Modelling survival data in medical research*. 2003, Boca Raton, Fla: Chapman & Hall/CRC.
7. Tsiatis, A.A., *A Large Sample Study of Cox's Regression Model*. Ann. Statist., 1981. **9**(1): p. 93-108.
8. Heo, M., M.S. Faith, and D.B. Allison, *Power and sample size for survival analysis under the Weibull distribution when the whole lifespan is of interest*. Mechanisms of ageing and development, 1998. **102**(1): p. 45-53.
9. Hernandez, A.V., E.W. Steyerberg, and J.D. Habbema, *Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical*

- power and reduces sample size requirements. Journal of clinical epidemiology, 2004. 57(5): p. 454-60.*
10. Jennison, C.a.T., Bruce W., *Group sequential methods with applications to clinical trials.* 2000, Boca Raton Chapman & Hall/CRC.
  11. Finkelstein, D.M., *A Proportional Hazards Model for Interval-Censored Failure Time Data.* Biometrics, 1986. 42(4): p. 845-854.

## ACKNOWLEDGMENTS

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

## CONTACT INFORMATION

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

Robert Lew  
MAVERIC, VA Cooperative Studies Program,  
150 South Huntington Ave.  
Boston, MA 02130  
Robert.Lew2@va.gov