Power and Sample Size Estimation in Clinical Trials with Multiple Co-primary Endpoints

Zuoshun Zhang[1]

Biostatistics and Programming, Celgene Corporation, Summit, New Jersey, USA

Abstract

For some diseases, there needs to show the significance of a treatment in multiple co-primary endpoints. In the design stage, it will be helpful to have easily accessible tools for power and sample size estimation. Sozu et al. (2011) provided formulas and gave examples for 2 and 3 co-primary endpoints using integration. For any number of co-primary endpoints, we give an alternative way for accurate and quick estimation for power and sample size. We factor the positive definite correlation matrix into a product using Cholesky decomposition and generate independent standard normal random numbers. The random numbers are transformed into correlated vectors corresponding to the sufficient test statistics of simulated trials, from which the power and sample size are estimated. For 2 or 3 co-primary continuous endpoints, we estimated the sample sizes and compared with those in Sozu et al (2011); the differences are within 1% for each of the 100 cases. We further estimate sample sizes for cases with 4 and 5 co-primary endpoints. We provided a program using SAS IML and base language, which can be easily adapted to other designs including those with binary endpoints.

Key Words: co-primary endpoints, power, sample size, correlation matrix, positive definite matrix, Cholesky decomposition

1. Introduction

Many clinical trials characterize treatment efficacy using a single primary endpoints. However for some diseases, clinically meaningful benefits are characterized through multiple primary efficacy endpoints, where the trials are designed to show statistically significant outcomes for all those endpoints simultaneously. A non-significant result in any one of the specified efficacy endpoints would lead to a non-win scenario for the trial. Such primary endpoints are termed co-primary endpoints in clinical trial terminology. Examples can be found in migraine, acute pain, Alzheimer's disease, menopausal symptoms, and vaccine trials. Offen et al. (2007) and Dmitrienko et al. (2009) reviewed co-primary designs with medical and statistical perspectives.

Several methods for power and sample size calculations have been proposed for clinical trials with multiple co-primary endpoints (Xiong et al., 2005; Eaton and Muirhead, 2007; Hung and Wang, 2009; Song, 2009; Kordzakhia et al., 2010; Sozu et al., 2010, 2011). Xiong et al. (2005) proposed a power formula for two co-primary endpoints in superiority clinical trials when they are bivariate normally distributed and their variance–covariance matrix is known. Eaton and Muirhead (2007) provided computable bounds of the power function under the assumption of multivariate normality. Hung and Wang (2009) gave bounds for sample size. Song (2009) discussed sample size calculations with multiple co-primary binary endpoints for non-inferiority clinical trials. Kordzakhia et al. (2010)

[1] Address correspondence to Zuoshun Zhang, Celgene Corporation, 33 Technology Drive, Warren, NJ 07059, USA; E-mail: zzhang@celgene.com

presented a generalization in testing co-primary endpoints using a method of balanced adjustment. Sozu et al. (2010, 2011) provided formulas for power and sample size calculations with multiple co-primary continuous and binary endpoints in superiority clinical trials. In addition, Kong et al. (2004) used simulation and presented type I error and power estimation in non-inferiority and equivalence trials with correlated multiple endpoints for an example from vaccine development.

Sozu et al. (2010, 2011) provided general formulas and illustrated their method for 2 or 3 co-primary endpoints using numerical integration. For any number of co-primary endpoints, we give an alternative way for accurate and quick estimation of power and sample size. In section 2, we provide statistical settings for designs with multiple co-primary continuous endpoints. In section 3, we give a general approach using simulation to estimate power and sample size. We factor the positive definite correlation matrix into a product using Cholesky decomposition and generate independent standard normal random numbers. The random numbers are transformed into correlated vectors corresponding to the sufficient test statistics of simulated trials, from which the power and sample size are estimated. For 2 or 3 co-primary continuous endpoints, we estimate the sample sizes and compare with those in Sozu et al (2011); the differences are within 1% for each of the 100 cases. We further estimate sample sizes for cases with 4 and 5 co-primary continuous endpoints. Section 4 presents sample sizes for co-primary binary endpoints. Section 5 gives further discussion and conclusion. In the appendix, we give a program using SAS IML and BASE language for power and sample size estimation with $K = 3$ co-primary continuous endpoints, which can be easily modified to other cases with multiple co-primary endpoints.

2. Clinical Trials with Multiple Co-Primary Continuous Endpoints

In this section, we follow Sozu et al. (2011) to set up notations for statistical design with co-primary continuous endpoints. We will use slightly different notations in order to have uniform presentation for overall sample sizes for trials with different ratios of randomization between two arms (Rosenberger and Lachin 2002).

2.1 Statistical Settings

In a randomized clinical trial with two arms, let $i = A$ or B denote the test and control arms. Overall, the trial will randomize n subjects, where $n_A = Qn$ subjects will be randomized to arm A and $n_B = (1 - Q)n$ subjects will be randomized to arm B. There are $k = 1, \cdots, K$ co-primary continuous endpoints with joint normal distributions.

For a subject in arm A, the response vector $Y_{A,j} = \left(Y_{A1,j}, \cdots, Y_{AK,j}\right)^T$, where $j = 1, \dots, n_A$; and for a subject in arm B, the response vector $Y_{B,j} = \left(Y_{B1,j}, \cdots, Y_{BK,j}\right)^T$, where $j = 1, \dots, n_B$. The endpoints are jointly distributed as
$$Y_{A,j} \sim N_K(\mu_A, \Sigma), \ Y_{B,j} \sim N_K(\mu_B, \Sigma),$$
where
$$\Sigma = \begin{pmatrix} \sigma_1^2 & \cdots & r_{1K}\sigma_1\sigma_K \\ \vdots & \ddots & \vdots \\ r_{1K}\sigma_1\sigma_K & \cdots & \sigma_K^2 \end{pmatrix}$$
with $var[Y_{Ak,j}] = var[Y_{Bk,j}] = \sigma_k^2$, $cov[Y_{Ak,j}, Y_{Ak',j}] = cov[Y_{Bk,j}, Y_{Bk',j}] = \sigma_k\sigma_{k'}r_{kk'}$, and $\mu_A = (\mu_{A1}, \cdots, \mu_{AK})^T, \mu_B = (\mu_{B1}, \cdots, \mu_{BK})^T$. The correlation matrix is given by

$$R = \begin{pmatrix} 1 & \cdots & r_{1K} \\ \vdots & \ddots & \vdots \\ r_{1K} & \cdots & 1 \end{pmatrix}.$$

We are interested in testing the hypotheses on the mean difference $\delta_k = \mu_{Ak} - \mu_{Bk}$, where a positive value of δ_k indicates the benefit of the test arm A over the control arm B, $k = 1, \dots, K$.

For $k = 1, \dots, K$, the null and alternative hypotheses are
$$H_{k,0}: \delta_k \leq 0 \ vs. \ H_{k,1}: \delta_k > 0.$$
The co-primary design seeks to assert the superiority of the test arm A over the control arm B in all K primary endpoints simultaneously. The success of the trial will be to reject all K null hypotheses at the one-sided level α, and the overall power will be $1 - \beta$ at the pre-specified standardized effect size $d_k = \delta_k/\sigma_k$, for $k = 1, \dots, K$.

2.2 Power and Sample Size Formula

Assume the variance and co-variance matrix $\sum$ are known. For $k = 1, \dots, K$, the following standard Z-statistic can be used to test the mean difference for each endpoint:
$$Z_k = \frac{\overline{Y}_{Ak} - \overline{Y}_{Bk}}{\sqrt{\text{var}(\overline{Y}_{Ak} - \overline{Y}_{Bk})}} = \frac{\overline{Y}_{Ak} - \overline{Y}_{Bk}}{\sigma_k/\sqrt{Q(1 - Q)n}}$$
where $\overline{Y}_{Ak}$ and $\overline{Y}_{Bk}$ are sample means and given by
$$\overline{Y}_{Ak} = \sum_{j=1}^{n_A} Y_{Ak,j}/n_A, \ \overline{Y}_{Bk} = \sum_{j=1}^{n_B} Y_{Bk,j}/n_B.$$

Then $Z = (Z_1, \cdots, Z_K)^T$ has multi-normal distribution with covariance matrix
$$R = \begin{pmatrix} 1 & \cdots & r_{1K} \\ \vdots & \ddots & \vdots \\ r_{1K} & \cdots & 1 \end{pmatrix}.$$

The co-primary null hypotheses will be rejected when all $Z_k \geq z_\alpha, k = 1, \dots, K$, for the one-sided test at the significant level α, where z_α is the upper 100αth percentile of the standard normal distribution; and the power function will be evaluated at the pre-specified standardized effect size $d_k = \delta_k/\sigma_k$, i.e., when Z_k has mean d_k, for $k = 1, \dots, K$. Therefore, the power function is
$$1 - \beta = Prob\left[\bigcap_{k=1}^{K}\{Z_k > z_\alpha\}\right] = Prob\left[\bigcap_{k=1}^{K}\{W_k > c_k\}\right],$$
where $c_k = z_\alpha - d_k\sqrt{Q(1 - Q)n}$ and $W = (W_1, \cdots, W_K)^T$ has multivariate normal distribution with mean vector 0 and covariance matrix R.

Let $C_k = -c_k = -z_\alpha + d_k\sqrt{Q(1 - Q)n}$, then the sample size can be expressed as
$$n = \frac{1}{Q(1 - Q)}\left(\frac{C_K + z_\alpha}{d_K}\right)^2$$
where C_k is the solution of the following integral equation
$$1 - \beta = \Phi_K\left(\frac{d_1}{d_K}(z_\alpha + C_K) - z_\alpha, \cdots, \frac{d_{K-1}}{d_K}(z_\alpha + C_K) - z_\alpha, C_K\right),$$
and Φ_K is the cumulative distribution function of the multivariate normal distribution with mean vector 0 and covariance matrix R.

Sozu et al. (2011) calculated sample sizes for cases with 2 and 3 ($K = 2, 3$) co-primary endpoints by evaluating the above integral using numerical integration via SAS programming language. In this note, we will use simulation to estimate power and sample size for $K = 2, 3, 4, 5$ co-primary endpoints. The procedures and SAS programs are essentially the same without additional complexity with any number of co-primary endpoints.

3. Power and Sample Size Estimation for Continuous Endpoints

3.1 General Approach for Simulation

We first consider two extreme theoretical cases as guidance, even though they may not be practical. Let the pre-specified standardized effect size be the same for all K co-primary endpoints, i.e., $d_1 = \cdots = d_K$. For the following two cases: (i) perfect correlation between any two endpoints, the sample size will be the same as one primary endpoint design with the one-sided test at level α and power $(1 - \beta)$; (ii) no correlation between any two different endpoints, the sample size can be determined as one primary endpoint design with the one-sided test at level α and power $1 - \beta_1 = \sqrt[K]{(1 - \beta)}$.

For $k = 1, \cdots, K$, let us define the minimum of the standardized effect sizes as the following

$$d = \min\left\{d_k = \delta_k/\sigma_k : k = 1, \cdots, K\right\}.$$

Then the sample size for the K co-primary endpoints design will be bounded by

$$n_{min} = \frac{1}{Q(1 - Q)}\left(\frac{z_\alpha + z_\beta}{d}\right)^2, n_{max} = \frac{1}{Q(1 - Q)}\left(\frac{z_\alpha + z_{\beta_1}}{d}\right)^2.$$

For using simulation to estimate power and sample size, we could expect the estimated sample size n would be approximately between n_{min} and n_{max}. In the following, we will list all the steps for power and sample size estimation using SAS programming:

(1) Using SAS PROC IML, first check the correlation matrix R is positive-definite by showing all K eigenvalues being positive; otherwise adjust the correlation matrix to make it positive-definite. Then, de-composite (Cholesky decomposition) the symmetric positive-definite correlation matrix R into the product of a lower triangular matrix and its conjugate transpose $R = AA^T$, where A is a $K \times K$ matrix with positive diagonal entries and 0 above diagonal entries.

(2) Let the number of the simulated trials be M (we will use $M = 20000$ in all examples). Generate MK $i.i.d.$ standard normal distributed random numbers $X_m = (X_{m1}, \cdots, X_{mK})^T, m = 1, \cdots, M$. Make the linear transformation
$$W_m = (W_{m1}, \cdots, W_{mK})^T = AX_m.$$
It is easy to show $Cov(W_m, W_m^T) = R$. Therefore, each W_m represents an outcome statistic for a simulated trial, $m = 1, \cdots, M$.

(3) Let n be the sample size for the trial. For $m = 1, \cdots, M$, find all cases which lead to reject of all co-primary null hypotheses at the one-sided level α, which satisfies the following condition

$$\left\{W_k > z_\alpha - d_k\sqrt{Q(1-Q)n} : \ k = 1, \cdots, K\right\}.$$

The simulated power will be the proportion of all such cases among the M simulated trials.

(4) For given power $(1 - \beta)$, the estimated sample size will be the smallest integer n such that the associated estimated power is greater than or equal to $(1 - \beta)$. The sample size n could be set between $n_{min} - 10$ and $n_{max} + 10$ in order to find the right sample size for the design and to test the sensitivity of the simulation algorithm.

3.2 Two Co-Primary Continuous Endpoints (K = 2)

For the standardized effect sizes of the two co-primary endpoints range from 0.20 to 0.40, Sozu et al. (2011) calculated sample size for the one-sided level $\alpha = 0.025$ and power $1 - \beta = 0.80$ (80%), and reported their results in Table 1 of their paper. For all the cases, we use a SAS program with $M = 20000$ simulated trials to estimate the sample size and list them in the following Table 3.1. The sample sizes range from 230 to 1035. For all the cases, the differences between their results using numerical integration and our results using simulation are within ± 3 and $\pm 1\%$ of the total sample sizes.

For the standardized effect sizes of the two co-primary endpoints range from 0.20 to 0.40, we further estimated sample sizes for the one-sided level $\alpha = 0.025$ and power $1 - \beta = 0.90$ (90%), and reported the results in Table 3.2. We plot power and sample sizes for the case of the standardized effect sizes as 0.30 for both co-primary endpoints for the one-sided level $\alpha = 0.025$ and various correlation coefficients.

Table 3.1: Sample Size with $K = 2$ Co-primary Continuous Endpoints, 1:1 Randomization, One-sided Level $\alpha = 0.025$ and Power $1 - \beta = 0.80$ (80%)

Standardized Effect Size		Correlation Coefficient r_{12}				Sample Size for the Smallest Effect with Power	
d_1	d_2	0.0	0.3	0.5	0.8	$1 - \beta$	$\sqrt[K]{(1-\beta)}$
0.20	0.20	1035	1009	982	918	785	1031
0.20	0.25	866	851	835	801		
0.20	0.30	805	797	793	785		
0.20	0.35	787	786	785	784		
0.20	0.40	784	784	784	784		
0.25	0.25	662	645	628	588	503	660
0.25	0.30	571	557	546	519		
0.25	0.35	527	520	513	504		
0.25	0.40	508	506	504	502		
0.30	0.30	460	448	437	408	351	459
0.30	0.35	404	395	386	365		
0.30	0.40	374	368	361	352		
0.35	0.35	338	329	321	300	258	337
0.35	0.40	301	295	287	272		
0.40	0.40	259	252	246	230	198	258

Table 3.2: Sample Size with $K = 2$ Co-primary Continuous Endpoints, 1:1 Randomization, One-sided Level $\alpha = 0.025$ and Power $1 - \beta = 0.90$ (90%)

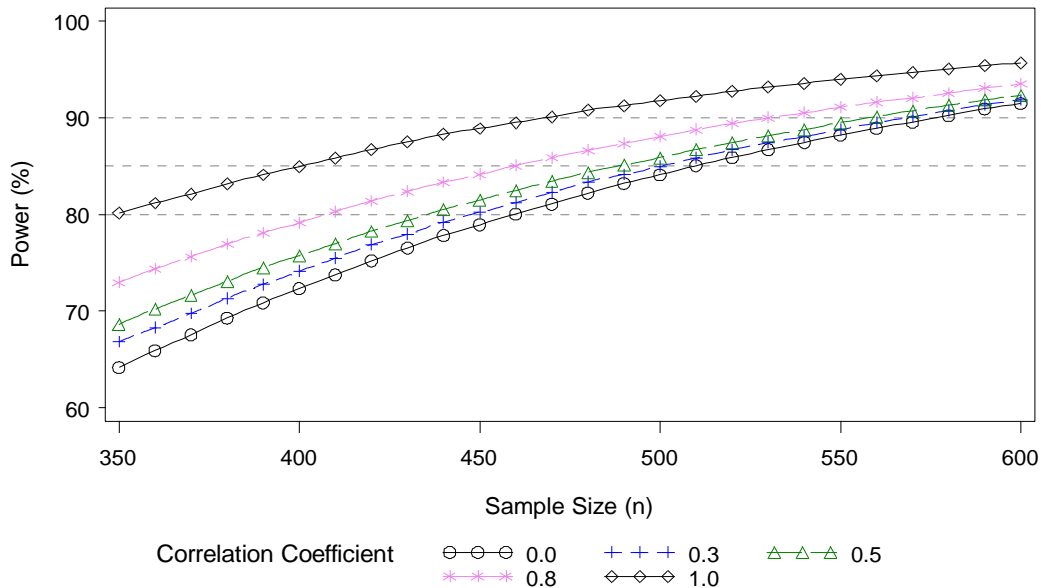| Standardized Effect Size		Correlation r_{12}				Sample Size for the Smallest Effect with Power	
d_1	d_2	0.0	0.3	0.5	0.8	$1 - \beta$	$\sqrt[K]{(1 - \beta)}$
0.20	0.20	1295	1279	1256	1193	1052	1291
0.20	0.25	1108	1091	1079	1063		
0.20	0.30	1059	1058	1054	1053		
0.30	0.30	576	569	558	530	470	574
0.30	0.35	511	502	497	479		
0.30	0.40	479	476	475	469		
0.40	0.40	324	320	314	299	266	323
0.40	0.45	294	290	285	274		
0.40	0.50	277	273	270	266		

Figure 3.1: Power and Sample Size with $K = 2$ Co-primary Continuous Endpoints, 1:1 Randomization, One-sided Level $\alpha = 0.025$, Standard Effect Size $d_1 = d_2 = 0.3$, and Different Correlation Coefficients

3.3 Three Co-Primary Continuous Endpoints (K = 3)

For $K = 3$, Sozu et al. (2011) calculated sample size for the one-sided level $\alpha = 0.025$ and power $1 - \beta = 0.80$ (80%), and reported their results in Table 2 of their paper. For all the cases, we use a SAS program with $M = 20000$ simulated trials to estimate the sample size and list them in the following Table 3.3. The differences between their results using numerical integration and our results are within ± 3 and $\pm 1\%$.

Table 3.3: Sample Size with $K = 3$ Co-primary Continuous Endpoints, 1:1
Randomization, One-sided Level $\alpha = 0.025$ and Power $1 - \beta = 0.80$ (80%)

Standardized Effect Size			Correlation $r_{12} = r_{13} = r_{23}$				Sample Size for the Smallest Effect with Power	
d_1	d_2	d_3	0.0	0.3	0.5	0.8	$1 - \beta$	$\sqrt[K]{(1 - \beta)}$
0.20	0.20	0.20	1176	1136	1089	984	785	1172
0.20	0.20	0.30	1029	1003	976	914		
0.20	0.20	0.40	1027	1001	975	914		
0.20	0.30	0.30	819	808	798	786		
0.20	0.30	0.40	805	797	792	785		
0.20	0.40	0.40	785	785	784	784		
0.30	0.30	0.30	523	505	484	438	351	521
0.30	0.30	0.40	463	450	437	407		
0.30	0.40	0.40	389	379	370	354		
0.40	0.40	0.40	294	284	273	246	198	293

Table 3.4: Sample Size with $K = 3$ Co-primary Continuous Endpoints, 1:1
Randomization, One-sided Level $\alpha = 0.025$ and Power $1 - \beta = 0.90$ (90%)

Standardized Effect Size			Correlation $r_{12} = r_{13} = r_{23}$				Sample Size for the Smallest Effect with Power	
d_1	d_2	d_3	0.0	0.3	0.5	0.8	$1 - \beta$	$\sqrt[K]{(1 - \beta)}$
0.20	0.20	0.20	1433	1403	1360	1268	1052	1428
0.20	0.25	0.25	1146	1124	1104	1065		
0.20	0.25	0.30	1108	1091	1081	1058		
0.20	0.30	0.30	1061	1056	1053	1050		
0.30	0.30	0.30	637	624	605	564	470	635
0.30	0.35	0.35	539	527	515	486		
0.30	0.35	0.40	516	508	498	478		
0.30	0.40	0.40	487	483	478	468		
0.40	0.40	0.40	359	351	340	317	266	357

Table 3.5: Sample Size with $K = 3$ Co-primary Continuous Endpoints, 2:1
Randomization, One-sided Level $\alpha = 0.025$ and Power $1 - \beta = 0.90$ (90%)

Standardized Effect Size			Correlation Coefficients r_{12}, r_{13}, r_{23}				Sample Size for the Smallest Effect with Power	
d_1	d_2	d_3	0 0 0	0.2 0.3 0.4	0.4 0.5 0.6	0.6 0.7 0.8	$1 - \beta$	$\sqrt[K]{(1 - \beta)}$
0.20	0.20	0.20	1612	1573	1531	1467	1183	1606
0.30	0.30	0.30	717	699	681	652	526	714
0.30	0.35	0.35	607	594	581	561		
0.30	0.35	0.40	580	575	563	551		
0.30	0.40	0.40	548	544	539	533		
0.40	0.40	0.40	403	394	383	367	296	402

It is worth to note that when $K = 3$, it may not be feasible for all values of r_{12}, r_{13}, r_{23} ranging from 0 to 1 to make a correlation matrix. For example, matrix $R = \begin{pmatrix} 1 & 0.8 & 0.2 \\ 0.8 & 1 & 0.8 \\ 0.2 & 0.8 & 1 \end{pmatrix}$ can't be a correlation matrix. Using SAS IML, we find its eigenvalues are 2.236, 0.8, -0.036, which indicates it is not positive definite.

3.4 Four or More Co-Primary Continuous Endpoints ($K \geq 4$)

The method for estimating power and sample size using simulation with SAS IML and BASE can be easily generated to cases for more than 3 co-primary endpoints. The time for calculation will increase linearly with the number of co-primary endpoints K, which is usually within 1-2 minutes in a PC with Windows 7. In the following, we listed some cases for $K = 4, 5$.

Table 3.6: Sample Size with $K = 4$ Co-primary Continuous Endpoints, 1:1 Randomization, One-sided Level $\alpha = 0.025$ and Power $1 - \beta = 0.90$ (90%)

Standardized Effect Size				Correlation $r_{12} = r_{13} = r_{14} = r_{23} = r_{24} = r_{34}$				Sample Size for the Smallest Effect with Power	
d_1	d_2	d_3	d_4	0.0	0.3	0.5	0.8	$1 - \beta$	$\sqrt[K]{(1 - \beta)}$
0.20	.20	.20	.20	1538	1491	1454	1325	1052	1524
0.30	.30	.30	.30	684	663	646	589	470	678
0.30	.30	.30	.35	649	633	614	567		
0.30	.30	.35	.35	611	596	578	540		
0.30	.35	.35	.35	561	548	534	500		
0.30	.40	.40	.40	502	497	487	473		
0.40	.40	.40	.40	385	373	364	332	266	381

Table 3.7: Sample Size with $K = 5$ Co-primary Continuous Endpoints, 1:1 Randomization, One-sided Level $\alpha = 0.025$ and Power $1 - \beta = 0.90$ (90%)

Standardized Effect Size					Correlation $r_{12} = r_{13} = r_{14} = r_{15} = r_{23} = r_{24} = r_{25} = r_{34} = r_{35} = r_{45}$				Sample Size for the Smallest Effect with Power	
d_1	d_2	d_3	d_4	d_5	0.0	0.3	0.5	0.8	$1 - \beta$	$\sqrt[K]{(1 - \beta)}$
0.20	.20	.20	.20	.20	1609	1560	1506	1362	1052	1598
0.30	.30	.30	.30	.30	716	693	670	606	470	710
0.30	.35	.35	.35	.35	575	561	544	502		
0.40	.40	.40	.40	.40	403	390	377	341	266	400

3.5 Relationship of Sample Size and Number of Co-Primary Endpoints

In general with increase of the number K of co-primary endpoints, we have to increase sample size in order to achieve some pre-determined power. Such increase could be rather substantial in general, e.g., with $K = 5$ co-primary endpoints and low correlation

among them, the sample size increase could be more than 50% comparing to one primary endpoint. As pointed by Offen et al. (2007), there are cases where co-primary endpoints won't increase sample size dramatically: (a) all co-primary endpoints are highly correlated, (b) there is a big gap in the standardized effect size between the smallest and other co-primary endpoints and the study is powered by the smallest effect.

The following figure presents the relationship between power and sample size for $K = 1, 2, 3, 4, 5$ co- primary endpoints with correlation coefficients 0.3 between any two of the endpoints and the smallest standardized effect size 0.3 and all other standardized effect sizes 0.35.
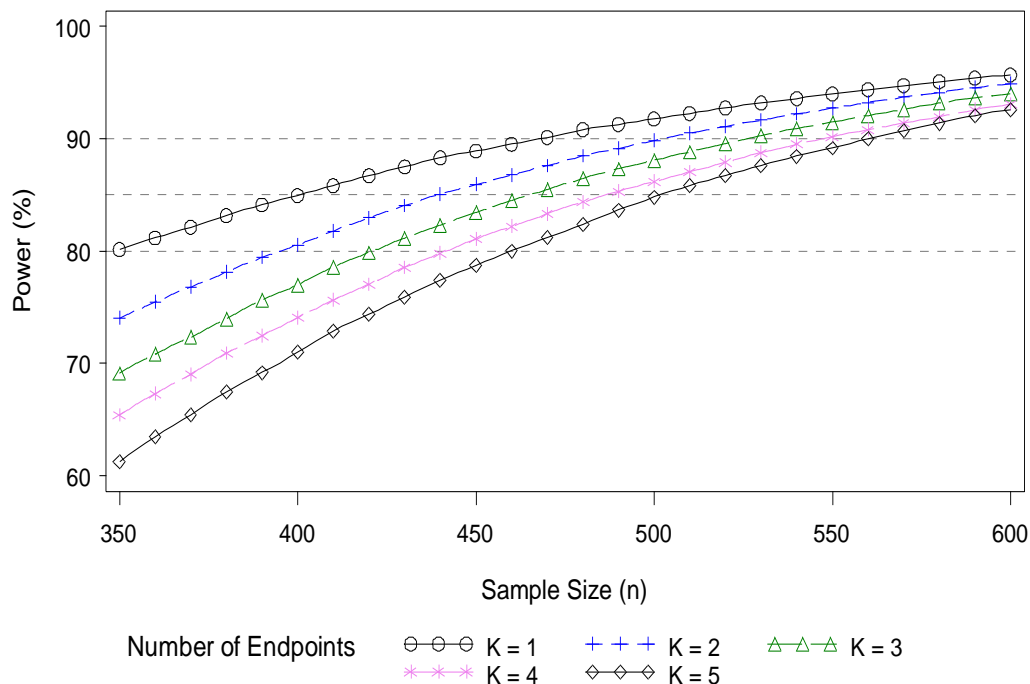

Figure 3.2: Power and Sample Size with K Co-primary Endpoints, 1:1 Randomization, One-sided Level $\alpha = 0.025$, Standard Effect Size $d_1 = 0.3$, $d_2 = \cdots = d_K = 0.35$, and Correlation Coefficients 0.30

4. Power and Sample Size Estimation for Co-Primary Binary Endpoints

For clinical trials with co-primary binary endpoints, similar approaches using simulation as in the above section can be used for power and sample size estimation. With normal approximation, test statistics of the correlated co-primary binary endpoints can be put into variables with multivariate normal distribution. Sozu et al. (2011) provides formulas and calculate sample sizes for $K = 2$ or 3 co-primary binary endpoints using numerical integration.

For $K = 3$ co-primary binary endpoints, we use a SAS program with $M = 20000$ simulated trials to estimate sample sizes for chi-square test using normal approximation without continuity correction. For all 24 cases with the one-sided level $\alpha = 0.025$ and power $1 - \beta = 0.80\ (80\%)$, the sample size differences between the

results using simulation in Table 4.1 and the corresponding results in Sozu et al. (2011) Table 6 are within ± 4 and $\pm 1\%$.

The simulation approach can be easily adapted to cases with 4 or more co-primary binary endpoints and to cases with co-primary continuous and binary endpoints. We are going to give details in a separate report.

Table 4.1: Sample Size with $K = 3$ Co-primary Binary Endpoints, 1:1 Randomization, One-sided Level $\alpha = 0.025$ and Power $1 - \beta = 0.80$ (80%), Chi-square Test with Normal Approximation Without Continuity Correction

Proportions						Correlation $r_{12} = r_{13} = r_{23}$				Sample Size for the Smallest Effect with Power	
Endpoint 1		Endpoint 2		Endpoint 3							
π_{11}	π_{21}	π_{12}	π_{22}	π_{13}	π_{23}	0.0	0.3	0.5	0.8	$1 - \beta$	$\sqrt[K]{(1-\beta)}$
.60	.50	.60	.50	.60	.50	1159	1120	1074	971	776	1164
.60	.50	.60	.50	.65	.50	1014	988	962	901		
.60	.50	.65	.50	.65	.50	804	795	786	776		
.65	.55	.65	.55	.65	.55	1124	1086	1041	941	752	1130
.65	.55	.65	.55	.70	.55	983	958	933	874		
.65	.55	.70	.55	.70	.55	777	769	762	752		

Note: Proportions $\pi_{11}, \pi_{12}, \pi_{13}$ and $\pi_{21}, \pi_{22}, \pi_{23}$ are Arm 1 and Arm 2 response rates of the 3 co-primary end-points under the alternative hypothesis.

5. Discussion and Conclusion

For design of clinical trials with co-primary continuous and binary endpoints, we show that power and sample size calculation can be accurately and efficiently carried out through simulation. For cases we calculated with 2 or 3 continuous and binary endpoints, the differences are within 1% of the sample size calculated using numerical integration. In the appendix, we provide a program using SAS IML and BASE language for power and sample size estimation for 3 co-primary continuous endpoints, which can be easily adapted for other cases with multiple correlated endpoints, including different number *K* of co-primary continuous endpoints and co-primary binary endpoints with test statistics using normal approximation. The complexity of the simulation approach is essentially the same for different number of co-primary endpoints.

References

Berger RL (1982). Multiparameter hypothesis testing and acceptance sampling. Technometrics, 24:295-300.

Chow SC, Shao J, Wang H. (2007). *Sample Size Calculations in Clinical Research* (2nd edn). Chapman and Hall/CRC: Boca Raton, FL.

Chuang-Stein, C, Stryszak, P, Dmitrienko, A and Offen, W (2007), Challenge of multiple co-primary endpoints: a new approach. Statist. Med. 26: 1181–1192.

Dmitrienko, A, Tamhane, AC, and Bretz, F (eds) (2009). Multiple Testing Problems in Pharmaceutical Statistics. CRC Press, New York.

Eaton ML, Muirhead RJ (2007). On a multiple endpoints testing problem. *Journal of Statistical Planning and Inference* 137:3416-3429.

Hung HM, Wang SJ (2009). Some controversial multiple testing problems in regulatory applications. *J Biopharm Stat.* 19:1-11.

Julious SA, McIntyre NE (2012). Sample sizes for trials involving multiple correlated must-win comparisons. Pharmaceut. Statist. Published online in Wiley Online Library.

Julious SA (2010). Sample Sizes for Clinical Trials. Chapman & Hall/CRC, Boca Raton, Florida.

Kong L, Kohberger RC, Koch GG (2004) Type I error and power in noninferiority/equivalence trials with correlated multiple endpoints: an example from vaccine development trials. *J Biopharm Stat.* 14: 893-907

Kordzakhia, G, Siddiqui, O and Huque, M.F (2010), Method of balanced adjustment in testing co-primary endpoints. Statist. Med. 29: 2055–2066.

Offen W, Chuang-Stein C, Dmitrienko A, Littman G, Maca J, Meyerson L, Muirhead R, Stryszak P, Boddy A, Chen K, Copley-Merriman K, Dere W, Givens S, Hall D, Henry D, Jackson JD, Krishen A, Liu T, Ryder S, Sankoh AJ, Wang J, Yeh CH. (2007). Multiple co-primary endpoints: medical and statistical solutions. *Drug Information Journal* 41:31-46.

Rosenberger, WF, Lachin, JM (2002). Randomization in Clinical Trials: Theory and Practice. Wiley, New York.

Sankoh, AJ, D'Agostino, RB and Huque, MF (2003). Efficacy endpoint selection and multiplicity adjustment methods in clinical trials with inherent multiple endpoint issues. Statist. Med. 22: 3133–3150.

Song, JX (2009). Sample size for simultaneous testing of rate differences in non-inferiority trials with multiple endpoints. Comp. Stat. & Data Anal. 53: 1201-1207.

Sozu T, Sugimoto T and Hamasaki T (2010). Sample size determination in clinical trials with multiple co-primary binary endpoints. Statist. Med. 29: 2169-2179.

Sozu T, Sugimoto T, Hamasaki T (2011). Sample size determination in superiority clinical trials with multiple co-primary correlated endpoints. J Biopharm Stat. 21:650-668.

Xiong C, Yu K, Gao F, Yan Y, Zhang Z (2005). Power and sample size for clinical trials when efficacy is required in multiple endpoints: application to an Alzheimer's treatment trial. *Clinical Trials* 2:387-393.

Appendix

SAS program using simulation for power and sample size estimation of $K = 3$ co-primary continuous endpoints

```
*input design parameters for sample size estimation;
%let k=3;          *number of co-primary endpoints;
%let q=0.5;        *randomization ratio between the two arms as q:(1-q);
%let d1=0.2;       *standard effect size for endpoint 1;
%let d2=0.2;       *standard effect size for endpoint 2;
%let d3=0.2;       *standard effect size for endpoint 3;
%let corrc=%str(1 0.3 0.3, 0.3 1 0.3, 0.3 0.3 1);  *correlation matrix;
%let nmin=900;             *sample size low bound;
%let nmax=1200;            *sample size upper bound;
%let ntrial=20000;         *total number of simulated trials in power estimation;
%let alpha=0.025;          *one-sided significant level;
%let seed=448877999;  *seed for generating standard normal random  numbers;
*genarating random numbers;
```

```
data sd1;
  retain seed &seed;
  do t=1 to &ntrial; do s=1 to &k;
    call rannor(seed,x); output;
  end; end;
run;

proc transpose data=sd1 out=sd2;
  by t;  id s;  var x;
run;

*transform random numbers into correlated test statistics;
Proc IML;
  corrc = {&corrc}; call eigen(egva, egvec, corrc);
  A=root(corrc); print CORRC egva egvec A;
  use sd2; read all;
  X = _1||_2||_3; W = X*A;
  create dw0 from W; append from W;
quit;

*estimating power as percentage of trials with significant co-primary outcomes;
data dw1;
  set dw0;
  k=&k; d1=&d1; d2=&d2; d3=&d3; q=&q; ntrial=&ntrial; alpha=&alpha;
  z_crt=-probit(alpha);
run;

data dw2;
  set dw1; do n=&nmin to &nmax; output; end;
run;

proc sort data=dw2 out=dw3; by n d1 d2 d3 ntrial; run;

data dw4;
  set dw3;
  by n;
  e1=sqrt(q*(1-q)*n)*d1;
  e2=sqrt(q*(1-q)*n)*d2;
  e3=sqrt(q*(1-q)*n)*d3;
  retain sigcnt;
  if first.n then sigcnt=0;
    sigcnt = sigcnt + (col1>z_crt-e1)*(col2>z_crt-e2)*(col3>z_crt-e3);
  if last.n then do; sigpcent=sigcnt*100/ntrial; output; end;
run;

data dw5; set dw4(keep=d1 d2 d3 n sigpcent); run;

proc print;run;
```