Sample Size and Assurance Probability Calculation in Multi-regional Clinical Trials

Zuoshun Zhang[1]
Biostatistics, Celgene Corporation, Summit, New Jersey, USA

Abstract

With the increasing globalization of drug development, it has become important that data from multi-regional clinical trials (MRCTs) can be accepted by regulatory agencies across regions as the primary source of evidence to support marketing approval of drugs. The Japanese MRCT guidance presented two criteria for evaluating consistency, which were generalized and evaluated using unconditional, conditional or joint approaches with the overall significant test. For using the three approaches to evaluate regional consistency requirements, we propose a unify method in sample size and assurance probability calculation. The method involves simulating sufficient test statistic vectors across regions and its outputs are accurate in comparison with results from several published papers using complex analytical derivation and numerical integration. The method can be conveniently implemented by a statistical practitioner and applicable to more general consistency criteria across regions in statistical planning of MRCT.

Key Words: Multi-regional clinical trial, MRCT, Consistency criterion, Sample size, Assurance probability

1. Introduction

As specified in ICH E17 guideline, it has become important that data from multi-regional clinical trials (MRCTs) can be accepted by regulatory agencies across regions as the primary source of evidence to support marketing approval of drugs (International Conference on Harmonization 2017). The guideline presented general recommendations in the planning and design of MRCTs, including sample size allocation and statistical analysis principles. The primary objective of an MRCT generally corresponds to an evaluation of the treatment effect averaged across all regions. The overall sample-size is determined to ensure that this objective can be met. The MRCT should be planned to include an evaluation of the consistency of treatment effects among regions.

In order to solve Japanese "drug lag" problem, MHLW (Ministry of Health, Labour and Welfare) Published a guidance in 2007: Basic Principles on Global Clinical Trials. The guidance is in question-and-answer form with 12 questions and answers, and a flow chart to determine when to use global confirmatory trials (MRCTs) and when to have an additional domestic (ie, Japan) study. For using MRCTs, the guideline proposed two methods (Method 1 and Method 2) for Japanese sample size requirement in order to obtain consistent result.

After the MHLW (2007) publication, many authors further evaluated and generalized the consistency criteria. Kawai et al (2008) provided formula and numerical example for

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

Method 2. Ikeda & Bretz (2010) and Quan et al. (2009) provided an approximate formula for Method 1. Uesaka (2009) proposed 4 criteria as generalization of Method 2, and calculated the assurance probability jointly with the overall significance effect. Ko et al (2010) used the same 4 criteria, and calculated the assurance probability conditionally on the overall significance. Quan et al (2010) gave 6 definitions for consistency across all regions, with Li et al. (2012) for calculation using R. Tsong et al (2012) proposed testing across all regions at different level α_k, either unconditional or conditional on overall significance.

In designing MRCT, the specific region(s) need to have sufficient number of subjects in order to achieve the desired assurance probability. The assurance probabilities can be defined in different ways in relationship with overall test statistic: (i) unconditional approach; (ii) conditional approach; (iii) joint approach. The calculation is usually via rather complicated analytical derivation and numerical integration, which may put extra burden for practical statisticians.

In this article, we propose to use simulation of sufficient test statistic vectors for calculations in MRCT designs. For continuous endpoints, the overall standard test statistic can be conveniently expressed as a weight summation of test statistic vectors on different regions. Under the alternative hypothesis, the statistic vectors over the regions can be expressed using standard normal random vectors and constants involving sample size or power. By simulating a large number of independent standard normal random vectors, we can easily estimate regional sample sizes and assurance probabilities. This method can be easily adapted to satisfying different requirements. Similar approaches were used in Bang et al (2005) for multiple-testing designs and Zhang (2012, 2017) for co-primary designs.

In Section 2, we present notation for six consistency criteria and three approaches in evaluating assurance probabilities. Section 3 presents statistical settings for MRCT with continuous endpoints and steps in calculating the assurance probabilities. In Section 4, 5 and 6, we present assurance probability calculation using the unconditional, conditional and joint approaches and compare with outputs from other papers. Section 6 also presents outputs of conditional assurance probability for cases with different regional effects, relationship with sample size changes, and consistency for multiple regions. Section 7 shows the method can be used for calculating power and sample size for testing regional effects. Section 8 has a discussion and three SAS programs are presented in the appendices.

2. Consistency Criteria for Multi-regional Clinical Trials

For a multi-regional clinical trial, let $k = 1, \cdots, K$ denote different regions. Furthermore, let s denote a specific region and sc denote the combined regions other than region s. Let the observed treatment effect be D for the whole study; D_k for region $k = 1, \cdots, K$; D_s and D_{sc} for region s and sc. For constant $0 < \rho < 1$, the following six consistency criteria are specified as:

A1: $D_s/D_{sc} \geq \rho$,
A2: $D_s/D \geq \rho$,
A3: $\rho \leq D_s/D_{sc} \leq 1/\rho$,
A4: $\rho \leq D_s/D \leq 1/\rho$,
A5: $D_k/D \geq \rho, where\ 1 \leq k \leq K' \leq K$,
A6: $\rho \leq D_k/D \leq 1/\rho, where\ 1 \leq k \leq K' \leq K$.

A7: $Z_k \geq z_{\alpha_k}, where\ 1 \leq k \leq K^{'} \leq K.$

Criteria A2 corresponds to Method 1 of Ministry of Health, Labor, and Welfare of Japan (MHLW 2007). Criteria A1 to A4 correspond to Criteria E1, E2, C1, and C2 in Uesaka (2009) and Criteria (i) to (iv) in Ko et al (2010). Criteria A5 and A6 generalize criteria A2 and A4 to consistency requirements in more than one region. Criteria A7 generalizes Tsong et al (2012), where they proposed testing across all *K* regions at level α_k for $k = 1, \cdots, K$.

We use a one-sided level α test for claiming the overall statistical significance with the critical region as $(Z > z_\alpha)$. The assurance probabilities are evaluated under the alternative hypothesis H_1 with pre-specified treatment effects of the *K* regions for the overall study power $(1 - \beta)$. The constant $0 < \rho < 1$ is the ratio of treatment effect for the specific region(s) to retain. There are three approaches in evaluating assurance probability for Criteria $A_m, m = 1, \cdots, 6$:

(i) The unconditional approach in MHLW (2007) defined as

$$P_u(A_m) = P_{H1}(A_m),$$

(ii) The joint approach in Uesaka (2009) defined as

$$P_j(A_m) = P_{H1}(A_m \cap (Z > z_\alpha)),$$

(iii) The conditional approach in Ko et al (2010) defined as

$$P_c(A_m) = P_{H1}(A_m | Z > z_\alpha) = P_{H1}(A_m \cap (Z > z_\alpha))/P_{H1}(Z > z_\alpha).$$

For example, the assurance probabilities for Criteria A_2 can be explicitly written as:

$$P_u(A_2) = P_{H1}(D_s/D \geq \rho),$$
$$P_j(A_2) = P_{H1}(D_s/D \geq \rho, Z > z_\alpha),$$
$$P_c(A_2) = P_{H1}(D_s/D \geq \rho | Z > z_\alpha).$$

In order to evaluate the assurance probabilities using numerical integration, one has to derive tediously analytical expressions for each of the criteria. In the following, we present an uniform approach using simulation of sufficient test statistic vectors across regions. The simulation method is straightforward to carry out using SAS programming and its outputs are in general accurate when comparing with numerical integration methods.

3. Multi-regional Clinical Trials Using Continuous Endpoint

In this section, we present statistical settings for MRCT with continuous endpoints. We express the overall standard test statistic as a weighted summation of independent regional test statistics, and propose a general approach for estimating sample sizes and assurance probabilities by simulating regional test statistic vectors.

3.1 Statistical Settings

Let $i = A$ or C denote the test and control arms. The trial randomizes n subjects, with $n_A = Q_A n$ subjects in arm A and $n_C = Q_C n = (1 - Q_A)n$ subjects in arm C, respectively. Let f_k

denote the fraction of subjects from region k, where $k = 1, \cdots, K$ and $f_1 + \cdots + f_K = 1$. then region k enrolls $n_k = f_k n$ subjects for both arms, and $n_{Ak} = Q_A n_k = Q_A f_k n$ subjects for arm A and $n_{Ck} = Q_C n_k = (1 - Q_A) f_k n$ subjects for arm C, respectively.

For subject j in arm i, region k, the response follows normal distribution with a known common variance, ie,

$$Y_{ik,j} \sim N(\mu_{ik}, \sigma^2).$$

For region $k = 1, \cdots, K$, the mean difference is $\Delta_k = \mu_{Ak} - \mu_{Ck}$. The overall mean difference for the study can be expressed as a weight average of regional mean differences as $\Delta = \Sigma \mathrm{f_k} \Delta_k = \Sigma \mathrm{f_k} (\mu_{Ak} - \mu_{Ck})$. We assume $\Delta > 0$ indicating benefit of the test arm A over the control arm C. The study is designed to reject the overall null hypothesis versus the alternative hypothesis

$$H_0 : \Delta \leq 0 \; vs. \; H_1 : \Delta > 0.$$

For the one-sided level α test with power $(1 - \beta)$ at the overall difference $\delta = \sum f_k \, \delta_k$, where $\delta_k > 0$ denote the expected difference for region k under the alternative hypothesis H_1. The overall study sample size n satisfying the following equation:

$$n Q_A (1 - Q_A) (\delta / \sigma)^2 = (z_\alpha + z_\beta)^2.$$

For regions $k = 1, \ldots, K$ or $k = s, sc$, the observed mean difference can be expressed as

$$D_k = \overline{\mathrm{Y}}_{\mathrm{Ak\cdot}} - \overline{\mathrm{Y}}_{\mathrm{Bk\cdot}} \sim \mathrm{N}\left[\Delta_k, \frac{\sigma^2}{Q_A (1 - Q_A) n_k}\right],$$

and standard test statistic can be expressed as

$$Z_k = \frac{D_k}{\sigma / \sqrt{Q_A (1 - Q_A) n_k}} \sim \mathrm{N}\left[\frac{\Delta_k}{\sigma} \sqrt{Q_A (1 - Q_A) n_k}, 1\right],$$

The overall observed mean difference can be expressed as

$$D = \overline{\mathrm{Y}}_{\mathrm{A\cdot}} - \overline{\mathrm{Y}}_{\mathrm{B\cdot}} = \sum\nolimits_{\mathrm{k}} \mathrm{f_k D_k} \sim \mathrm{N}\left[\Delta, \frac{\sigma^2}{Q_A (1 - Q_A) n}\right]$$

and the overall standard test statistic can be expressed as

$$Z = \frac{D}{\sigma / \sqrt{Q_A (1 - Q_A) n}} = \sum\nolimits_{k} \sqrt{\mathrm{f_k}} \mathrm{Z_k} \sim \mathrm{N}\left[\frac{\Delta}{\sigma} \sqrt{Q_A (1 - Q_A) n}, 1\right].$$

Under the alternative hypothesis H_1 with $\Delta_k = \delta_k$, for $k = 1, \ldots, K$, or $k = s, sc$, the test statistics can be expressed as

$$Z_k = W_k + B_k \sqrt{n} = W_k + C_{k,\alpha,\beta},$$

where W_k distributes as a standard normal variable with mean 0 and variance 1, and the constant $B_k = \sqrt{f_k Q_A (1 - Q_A)} \; (\delta_k / \sigma)$ and $C_{k,\alpha,\beta} = \sqrt{f_k} \; (z_\alpha + z_\beta)(\delta_k / \delta)$.

For regions $k = 1, \ldots, K$, or $k = s, sc$, let $e_k = \delta_k / \sigma$ denote the standardized regional effect size. The ratio of the regional effect and overall effect is expressed as $\delta_k / \delta = e_k / (\sum f_k \, e_k)$, which equals constant 1 if all regions have the same effect sizes.

The treatment effect ratios used in Condition (A1) to (A6) can be expressed using Z scores as:

$$D_s/D_{sc} = \sqrt{f_{sc}/f_s}\,(Z_s/Z_{sc}),$$

and

$$D_k/D = \sqrt{1/f_k}\,(Z_k/Z)\,, k = 1, \dots, K \text{ or } k = s, sc.$$

3.2 Steps for Calculating Assurance Probabilities

Using the above formulas, we can calculate the assurance probabilities for the six consistency criteria using either conditional, unconditional, or joint approaches as specified in Section 2. First, we simulate a large number of independent standard normal random vectors $(W_1, \dots, W_K)$ or (W_s, W_{sc}). Each of the random vectors corresponds to one clinical trial. Then, we translate the vectors to $(Z_1, \dots, Z_K)$ or $(Z_s,\ Z_{sc})$ using constants B_k or $C_{k,\alpha,\beta}$, where $k = 1, \dots, K$, or $k = s, sc$. The overall test statistic Z is derived as a weighted average of the components in vector $(Z_1, \dots, Z_K)$ or $(Z_s,\ Z_{sc})$ under the alternative hypothesis H_1. At last, the assurance probabilities are estimated as proportions of random vectors satisfying corresponding conditions.

In the following, we list steps for estimating the conditional assurance probability for Criteria A2, ie, $P_c(A_2) = P_{H1}(D_s/D \geq \rho | Z > z_\alpha)$:

(1) Let the number of the simulated trials be M (we use $M = 50000$ in all examples). Generate M independent K-dimensional random vectors $(W_1, \dots, W_K)$. For each vector, the components $W_1, \dots, W_K$ are independent random numbers from normal distribution with mean 0 and variance 1.

(2) Transform the vectors $(W_1, \dots, W_K)$ to the test statistical vector $(Z_1, \dots, Z_K)$ under the alternative hypotheses H_1 using either formula

$$Z_k = W_k + B_k\sqrt{n}, k = 1, \dots, K,$$

or formula

$$Z_k = W_k + C_{k,\alpha,\beta}, k = 1, \dots, K.$$

(3) Calculate the overall test statistic as the weighted average

$$Z = \sum_k \sqrt{\mathrm{f_k}}\mathrm{Z_k}.$$

(4) Among the M vectors, count the number of vectors satisfying both conditions $D_s/D = \sqrt{1/f_s}\,(Z_s/Z) \geq \rho$ and $Z > z_\alpha$, and count the number of vectors satisfying condition $Z > z_\alpha$. The assurance probability is estimated as the quotient.

In calculating the assurance probabilities for the specific region s as in Criteria A1 to A4, we can also combine all other regions to form a region sc and simulate 2-dimensional random vectors (W_s, W_{sc}).The same approach can be applied to estimate unconditional, conditional or joint assurance probabilities (power) for all seven criteria A1 to A7.

4. Unconditional Assurance Probabilities

The easiest approach is to understand MHLW (2007) literally and calculate the assurance probabilities without consideration of the overall efficacy test. We simulated $M = 50000$ independent normal random vectors for calculation using the SAS program in Appendix 1. The outputs for Criteria A2 (Method 2 in MHLW) are similar to those using approach of Ikeda and Bretz (2010).

Table 4.1: The unconditional assurance probabilities for observing Criteria (A1) to (A4) given $\alpha = 0.025, \rho = 0.50$ and power 80% or 90% ($\beta = 0.10, or\ 0.20$)

Proportion in Region s	$\beta = 0.10$				$\beta = 0.20$			
	Assurance Probability				Assurance Probability			
f_s	A1	A2	A3	A4	A1	A2	A3	A4
0.10	0.66	0.67	0.44	0.49	0.69	0.70	0.49	0.55
0.20	0.72	0.75	0.54	0.65	0.75	0.78	0.60	0.71
0.30	0.75	0.80	0.59	0.74	0.79	0.84	0.66	0.80
0.40	0.78	0.85	0.61	0.82	0.82	0.89	0.69	0.87
0.50	0.79	0.89	0.62	0.87	0.83	0.93	0.69	0.92

5. Joint Assurance Probabilities

Uesaka (2009) evaluated assurance probabilities for Criteria A1 to A4 jointly with the overall significant output. He derived integral formulas and presented numerical examples for: (1) Equal allocation to each region (AR1), (2) Total sample size minimization rule (AR2), and (3) Minimum sample size rule for the region of interest (AR3). In this section, we use the simulation method as specified in Section 3 to calculate sample sizes for rule AR1 and present outputs in Table 5.1 for effect size $\delta/\sigma = 0.40$ across all regions with joint assurance probability $\gamma = 0.8$ or 0.9. We also compare with sample sizes in Table 1 in Uesaka (2009). The differences between the two methods in overall sample size are within $\pm 1.4\%$. A SAS program is presented in Appendix 2.

Table 5.1: Total sample size and the sample size of the specific region in the case of equal allocation rule: $\alpha = 0.025, \beta \leq 0.2, \rho = 0.50$, 1:1 randomization, uniform effect size $\delta/\sigma = 0.40$, equal allocation among regions

	Sample size n_s/n					
Criterion	Joint assurance probability $\gamma = 0.8$			Joint assurance probability$\gamma = 0.9$		
	50%	33.3%	25%	50%	33.3%	25%
A1	144/287	109/326	93/371	223/445	189/568	180/721
A2	115/229	91/274	80/321	158/316	138/415	137/548
A3	211/422	154/463	133/534	339/677	251/754	223/891
A4	115/230	95/286	88/351	160/319	144/431	146/582

Note: Criteria A1, A2, A3 and A4 correspond to Criteria E1, E2, C1 and C2 in Uesaka (2009). The allocation percentage 50%, 33.3% and 25% correspond to trials with 2, 3 and 4 regions with equal allocation among regions. The trials are designed to have at least 80% power.

The joint assurance probability will be below the study power. Table 5.2 presented assurance probabilities for a study with 80% and 90% power with allocation ratio to the specified region ranging from 0.10 to 0.50 calculated using a SAS program in Appendix 1.

Table 5.2: The joint assurance probabilities for observing Criteria (A1) to (A4) given $\alpha = 0.025, \beta = 0.10, 0.20$ and $\rho = 0.50$ with uniform effects among regions

Proportion in Region s	$\beta = 0.10$				$\beta = 0.20$			
	Assurance Probability				Assurance Probability			
f_s	A1	A2	A3	A4	A1	A2	A3	A4
0.10	0.55	0.56	0.39	0.43	0.63	0.64	0.47	0.52
0.20	0.60	0.62	0.48	0.57	0.69	0.71	0.57	0.67
0.30	0.63	0.67	0.52	0.65	0.73	0.77	0.62	0.75
0.40	0.66	0.71	0.54	0.70	0.76	0.82	0.65	0.81
0.50	0.67	0.74	0.55	0.74	0.77	0.85	0.66	0.85

6. Conditional Assurance Probabilities

For cases with uniform true effect size across regions, Ko et al (2010) calculated the assurance probabilities for Criteria A1 to A4 conditionally on overall significance using rather complicated analytical derivation and numerical integration. In this section, we show the assurance probability can also be estimated accurately using the simulation approach from Section 3. We also apply the simulation approach to more general cases.

6.1 Cases with Uniform Regional Effects

Ko et al (2010) presented assurance probabilities conditionally on overall significance for a specified region s. We use $M = 50000$ simulated trials to estimate those assurance probabilities. For allocation ratio between 0.10 to 0.50, Table 4.1 presents the conditional assurance probabilities from Table 2 in Ko et al (2010) and from the simulation approach using the SAS program in Appendix 1. The differences between the two approaches are within ± 0.01.

Table 6.1: The conditional assurance probabilities for observing Criteria (A1) to (A4) given $\alpha = 0.025, \beta = 0.20$ and $\rho = 0.50$

Proportion in Region s	Assurance Probability from Table 2 in Ko at al. (2010)				Assurance Probability Using Simulation			
f_s	A1	A2	A3	A4	A1	A2	A3	A4
0.10	0.69	0.70	0.49	0.55	0.68	0.69	0.48	0.54
0.20	0.75	0.78	0.60	0.71	0.75	0.78	0.60	0.71
0.30	0.80	0.84	0.66	0.81	0.80	0.84	0.65	0.81
0.40	0.83	0.89	0.68	0.88	0.82	0.89	0.68	0.88
0.50	0.85	0.93	0.69	0.93	0.84	0.93	0.69	0.93

In the following figure, the assurance probabilities from the simulation approach were plotted.

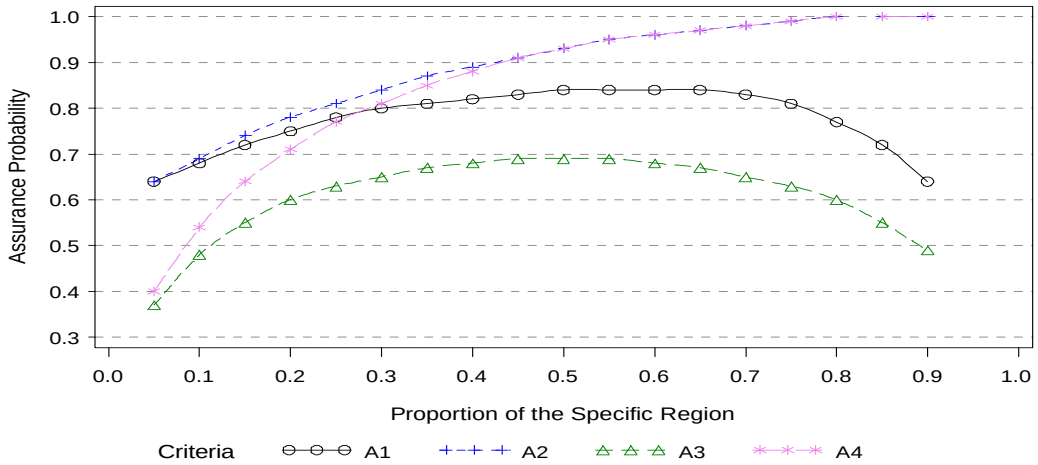

Figure 6.1: Conditional assurance probabilities with allocation ratio of the specific region f_s, for Criteria (A1) to (A4) given $\alpha = 0.025, \beta = 0.20$ and $\rho = 0.80$

6.2 Cases with Different Regional Effects

In design of MRCT, there may have cases where the assurance probabilities need to be evaluated with different effect sizes among the regions. In such cases, it will be tedious in calculating the assurance probabilities using numerical integration. Ko et al (2010) mentioned the issue and would hope to address it in the future. Uesaka (2009) presented some numerical example. Using the simulation approach as specified in Section 3, the assurance probabilities can be easily estimated. In the next table, we present the conditional assurance probabilities for cases having different regional effects using the SAS program in Appendix 1, where the sample sizes are determined using the overall averaged effect size to achieve the specified power.

Table 6.2: The assurance probability and sample size for observing Criteria (A1) to (A4) given $\alpha = 0.025, \beta = 0.20$ and $\rho = 0.50$, 1:1 randomization

Effect Size			Proportion in Region s	Sample Size	Conditional Assurance Probability			
Region s	Region sc	Overall						
δ_s/σ	δ_{sc}/σ	δ/σ	f_s	n	A1	A2	A3	A4
0.20	0.30	0.28	0.2	404	0.61	0.65	0.54	0.62
		0.27	0.3	434	0.64	0.70	0.58	0.69
		0.26	0.4	468	0.66	0.76	0.61	0.76
		0.25	0.5	506	0.68	0.83	0.62	0.83
0.30	0.20	0.22	0.2	652	0.88	0.90	0.58	0.74
		0.23	0.3	596	0.92	0.94	0.62	0.85
		0.24	0.4	548	0.93	0.96	0.63	0.92
		0.25	0.5	506	0.93	0.98	0.62	0.96

6.3 Relationship with Sample Size Changes

In clinical practice, the actual sample sizes can deviate from the planned sample size. For Criteria A2, the following figure shows that the conditional assurance probabilities increase slowly with sample size when the allocation proportion to the specified region f_s is fixed. The SAS program in Appendix 2 is used for calculation.

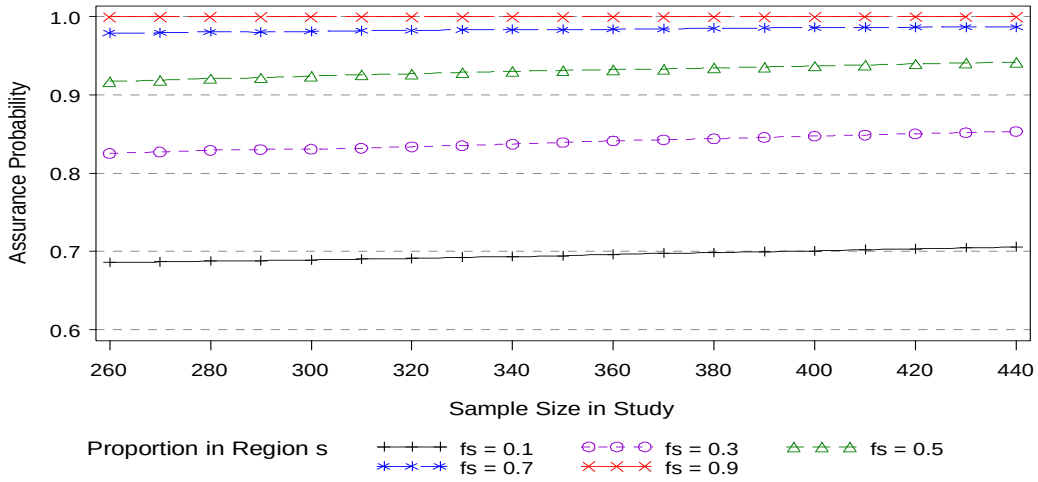

Figure 6.2: Conditional assurance probabilities against the sample size for Criteria A2 by proportion of the specific region, f_s, given $\alpha = 0.025$ and $\rho = 0.50$, and the effect size $\delta/\sigma = \delta_s/\sigma = \delta_{sc}/\sigma = 0.30$

The following table presents conditional assurance probabilities when the overall sample size changes and the sample size in the specified region fixed. The conditional assurance probabilities decrease slowly when the overall sample sizes increase.

Table 6.3: Assurance probabilities for observing Criteria (A1) to (A4) given $\alpha = 0.025, \rho = 0.50$, 1:1 randomization, uniform effect size $\delta/\sigma = 0.30$, and fixed sample size $n_s = 84$ for the specified region

Sample Size		Overall Power	Conditional Assurance Probability			
Region s	Overall					
n_s	n	%	A1	A2	A3	A4
84	250	66	0.79	0.84	0.63	0.82
	290	72	0.78	0.82	0.63	0.79
	330	78	0.78	0.81	0.63	0.77
	370	82	0.77	0.80	0.62	0.75
	410	86	0.76	0.79	0.62	0.73

6.4 Consistency for Multiple Regions

For MCRT requires consistency in multiple regions, we can still use the simulation approach in Section 3 to calculate the assurance probabilities. In the following, we present results for Criteria A5 and A6 with three region and all of them required to

be consistent, ie, $K' = K = 3$. Since the expected regional effects are the same, it is no surprise that the equal allocation among regions has the highest assurance probabilities. The study should be powered over 95% in order for the assurance probability exceed 0.80 for all three regions retaining 40% of the overall effect. A SAS program is present in Appendix 3.

Table 6.4: The conditional assurance probabilities for observing Criteria (A5) and (A6) given $\alpha = 0.025, \rho = 0.40$, $K' = K = 3$, and uniform regional effects

Type II Error	Regional Proportions	Assurance Probability For Consistency in all Three Regions		Assurance Probability For Consistency in Region 1	
β	(f_1, f_2, f_3)	A5	A6	A2	A4
0.20	(1/3, 1/3, 1/3)	0.70	0.70	0.90	0.89
0.10	(0.1, 0.1, 0.8)	0.55	0.48	0.75	0.69
	(0.1, 0.3, 0.6)	0.65	0.62	0.75	0.69
	(0.2, 0.3, 0.5)	0.72	0.71	0.84	0.83
	(1/3, 1/3, 1/3)	0.75	0.75	0.91	0.91
0.05	(1/3, 1/3, 1/3)	0.79	0.79	0.93	0.93

7. Power and Sample Size for Testing Regional Effects

For MRCT with $k = 1, \cdots, K$ regions, Tsong et al (2012) proposed testing across all K regions at level α_k, either unconditional or conditional on overall significance. For unconditional approach, they derived formulas for power and sample size. For conditional approach, they derived a rather complicated analytical formula for power and sample size using multiple integration and provided numerical examples. In this section, we show the same simulation approach in Section 3 can also apply to the testing approach.

Using the same notation as Section 3, let the regional tests be one-sided level α_k and the overall test be one-sided level α. For a subset of regions $1 \leq K' \leq K$, the conditional power can be expressed as

$$P_{H1}\left(Z_k \geq z_{\alpha_k},\ k = 1, \cdots, K' | Z > z_\alpha\right).$$

We use this formula as a basis for evaluating power with different sample size. A SAS program is presented in Appendix 4 which applies to any combinations of regional tests with the same or differeent regional levels and treatment effects. The following graph provided overall power and conditional power for regional tests for an MRCT with four regions.

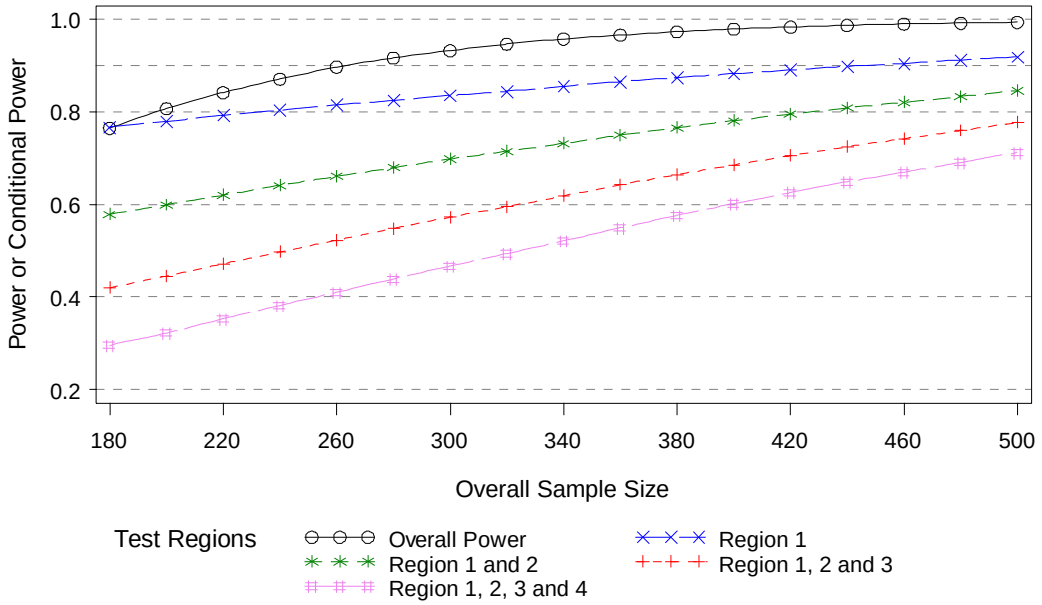

Figure 7.1: Power for testing regional effects conditional on overall significance with equal regional allocation for $K = 4$ regions, using 1:1 randomization, equal regional effect size $\delta/\sigma = 0.40$, overall 1-sided test with $\alpha = 0.025$, and 1-sided tests for 1, 2, 3 or all 4 regions at level $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0.20$.

8. Discussion and Conclusion

For design of MRCT, we show that the assurance probabilities can be accurately and efficiently calculated through simulation of sufficient test statistic vectors. The method involves simulation of independent test scores from individual regions and represents the overall test score as a weighted summation of the regional scores and represents ratio of treatment effects using the standard test scores. It avoids complex derivation of analytic formula and correlation coefficients. The same SAS program can provide calculation for the unconditional, conditional and joint assurance probability for different consistency criteria. The method can also be adapted to other approaches for MRCT designs.

Acknowledgments

The author is grateful to Angela Hu, Lillian Mellars, Mingyu Li and Xiaolong Luo for discussion and working on assurance probability and sample size calculation in a MRCT and to Huiling Li and Angela Hu for discussion and comments.

References

International Conference on Harmonization (2017). E17: General principle for planning and design of Multi-Regional Clinical Trials.

Ministry of Health, Labor, and Welfare of Japan. (2007). *Basic Principles on Global Clinical Trials*. Tokyo: MHLW

China Food and Drug Administration. (2015). Guidance for International Multicenter Clinical Trials (IMCT). Beijing, China.

Ikeda, K., Bretz, F. (2010). Sample size and proportion of Japanese patients in multi regional trials. Pharmaceutical Statistics 9:207–216.

Kawai, N., Chuang-Stein, C., Komiyama, O., Ii, Y. (2008) An approach to rationalize partitioning sample size into individual regions in a multiregional trial. Drug Information Journal 42:139–147.

Uesaka H (2009) Sample Size Allocation to Regions in a Multiregional Trial, Journal of Biopharmaceutical Statistics, 19:4, 580-594

Ko F, Tsou H H, Liu J, Hsiao C (2010) Sample Size Determination for a Specific Region in a Multiregional Trial, Journal of Biopharmaceutical Statistics, 20:4, 870-885.

Quan, H., Zhao, P. L., Zhang, J., Roessner, M., Aizawa, K. (2010). Sample size considerations for Japanese patients in a multi-regional trial based on MHLW guidance. *Pharmaceutical Statistics* 9(1):100–112

Quan H, Li M, Chen J, Gallo P, Binkowitx B, Ibia E, Tanaka, Y, Ouyang P, Luo X, Li G, Menjoge S, Talerice S, Ikeda K (2010) Assessment of consistency of treatment effects in multiregional clinical trials, Drug Information Journal, Vol 44: 619-634

Li M, Quan H, Chen J, Tanaka Y, Ouyang P, Luo X, Li G (2012). R Functions for Sample Size and Probability Calculations for Assessing Consistency of Treatment Effects in Multi-Regional Clinical Trials, Journal of Statistical Software, April 2012, Volume 47, Code Snippet 1.

Tsong, Y., Chang, W. J., Dong, X., Tsou, H. H. (2012). Assessment of regional treatment effect in a multiregional clinical trial. Journal of Biopharmaceutical Statistics 22(5):1019–1036.

Quan, H., Mao, X., Chen, J., Shih, W. J., Ouyang, S. P, Zhang, J., Zhao, P-L, Binkowitz, B. (2014). Multi-regional clinical trial design and consistency assessment of treatment effects, Statistics in Medicine 33(13):2191–2205.

Bang, H., Jung, S.-H., George, S. L. (2005). Sample size calculation for simulation-based multiple-testing procedures. J Biopharm Stat. 15(6):957–967.

Zhang, Z (2012). Power and sample size estimation in clinical trials with multiple co-primary endpoints, In *JSM Proceedings*, Biopharmaceutical Section. Alexandria, VA: American Statistical Association. 542-553.

Zhang, Z (2017). Sample size calculation in clinical trials with binary co-primary endpoints or multiple testing procedures, In *JSM Proceedings*, Biopharmaceutical Section. Alexandria, VA: American Statistical Association. 68-81.

Teng Z, Chen Y, Chang M (2017). Unified additional requirement in consideration of regional approval for multiregional clinical trials, Journal of Biopharmaceutical Statistics, 27:6, 903-917.

Appendix 1

A SAS program for calculation of unconditional, conditional or joint assurance probabilities with overall significant test in Table 4.1, 5.2, 6.1 and Figure 6.1

```
*input design parameters for sample size estimation;
%let alpha=0.025;       * One-sided significant level;
%let beta=0.10;         * Type 2 error (power=1-beta);
%let rho =0.50;         * Regional consistent constant;
%let eff1=0.35;         * Treatment effect size (delta/sigma): region 1 (s);
%let eff2=0.35;         * Treatment effect size (delta/sigma): region 2 (sc);
%let f1min=0.1;         * Low bound of allocation ratio in Region 1 (s);
```

```
%let f1max=0.5;          * Upper bound of allocation ratio in Region 1 (s);
%let f1by=0.10;          * Increase step for allocation ratio in Region 1 (s);
%let ntrial=50000;       * Total number of simualted trials;
%let seed=778899;        * Seed for generating random numbers;

data d0;
  alpha=&alpha; z1_alpha=-probit(alpha);
  beta=&beta; z1_beta=-probit(beta);
  rho=&rho; eff1=&eff1; eff2=&eff2; ntrial=&ntrial;
run;

proc iml;
  Mean = {0, 0}; Cov = {1 0, 0 1};
  call randseed(&seed);
  W = RandNormal(&ntrial, Mean, Cov);
  varNames = "W1":"W2";
  create sd0 from W[colname=varNames];
  append from W;
  close sd0;
quit;

data sd1; if _n_=1 then set d0; set sd0; run;

data sd2;
  set sd1;
  do f1=&f1min to &f1max by &f1by;
    f2 = 1-f1;
          eff = f1*eff1 + f2*eff2;
          c1 = sqrt(f1)*(z1_alpha + z1_beta)*(eff1/eff);
          c2 = sqrt(f2)*(z1_alpha + z1_beta)*(eff2/eff);
    Z1 = W1 + c1;
    Z2 = W2 + c2;
          Z = Z1*sqrt(f1) + Z2*sqrt(f2);
          ratio1_2 = (Z1/Z2)*sqrt(f2/f1);
          ratio1  12 = (Z1/Z)*sqrt(1/f1);
          output;
  end;
run;

proc sort data=sd2 out=sd3; by f1; run;

data sd4;
  set sd3;
  by f1;
  retain Zcnt uncA1 uncA2 uncA3 uncA4 jntA1 jntA2 jntA3 jntA4;
  if first.f1 then do;
          Zcnt=0;
          uncA1=0; uncA2=0; uncA3=0; uncA4=0;
          jntA1=0; jntA2=0; jntA3=0; jntA4=0;
  end;
  Zcnt = Zcnt + (Z>z1_alpha);
```

```
  uncA1 = uncA1 + (ratio1_2>=rho);
  uncA2 = uncA2 + (ratio1_12>=rho);
  uncA3 = uncA3 + (ratio1_2>=rho)*(ratio1_2=<1/rho);
  uncA4 = uncA4 + (ratio1_12>=rho)*(ratio1_12=<1/rho);
  jntA1 = jntA1 + (ratio1_2>=rho)*(Z>z1_alpha);
  jntA2 = jntA2 + (ratio1_12 >=rho)*(Z>z1_alpha);
  jntA3 = jntA3 + (ratio1_2>=rho)*(ratio1_2=<1/rho)*(Z>z1_alpha);
  jntA4 = jntA4 + (ratio1_12>=rho)*(ratio1_12=<1/rho)
          *(Z>z1_alpha);
  format uncPA1 uncPA2 uncPA3 uncPA4 jntPA1 jntPA2
          jntPA3 jntPA4 conPA1 conPA2 conPA3 conPA4 4.2;
  if last.f1 then do;
    P_Z = Zcnt/&ntrial;
    uncPA1 = uncA1/&ntrial;
    uncPA2 = uncA2/&ntrial;
    uncPA3 = uncA3/&ntrial;
    uncPA4 = uncA4/&ntrial;
    jntPA1 = jntA1/&ntrial;
    jntPA2 = jntA2/&ntrial;
    jntPA3 = jntA3/&ntrial;
    jntPA4 = jntA4/&ntrial;
    conPA1 = jntA1/Zcnt;
    conPA2 = jntA2/Zcnt;
    conPA3 = jntA3/Zcnt;
    conPA4 = jntA4/Zcnt;
    output;
  end;
run;

proc print;
          var alpha z1_alpha beta z1_beta rho eff1 eff2 f1 f2 uncPA1 uncPA2 uncPA3
          uncPA4 jntPA1 jntPA2 jntPA3 jntPA4 conPA1 conPA2 conPA3 conPA4;
run;
```

Appendix 2

A SAS program for evaluating sample size and joint assurance probabilities in Table 5.1

```
*input design parameters;
%let nmin = 300;      * Low bound of overall sample size n;
%let nmax = 550;   * Upper bound of overall sample size n;
%let nby=1;             * Increase step in sample size;
%let f1min=0.25;     * Low bound of allocation ratio in Region 1 (s);
%let f1max=0.25;  * Upper bound of allocation ratio in Region 1 (s);
%let f1by=0.1;         * Increase step for allocation ratio in Region 1 (s);
%let eff1=0.40;      * Treatment effect size (delta/sigma): region 1 (s);
%let eff2=0.40;      * Treatment effect size (delta/sigma): region 2 (sc);
%let alpha=0.025;  * One-sided significant level;
%let rho = 0.50;       * Consistent constant;
%let qa=0.50;          * Ranndomization proportion in Arm A (Qa = na/n);
%let ntrial=50000;   * Total number of simualted trials;
```

```
%let seed=778899;   * Seed for generating random numbers;

data d0;
  alpha=&alpha; z1_alpha=-probit(alpha);
  beta=&beta; z1_beta=-probit(beta);
  rho=&rho; eff1=&eff1; eff2=&eff2;
  qa=&qa; ntrial=&ntrial;
run;

proc iml;
  Mean = {0, 0};
  Cov = {1 0, 0 1};
  call randseed(&seed);
  W = RandNormal(&ntrial, Mean, Cov);
  varNames = "W1":"W2";
  create sd0 from W[colname=varNames];
  append from W;
  close sd0;
quit;

data sd1; if _n_=1 then set d0; set sd0; run;

data sd2;
  set sd1;
  do f1=&f1min to &f1max by &f1by;
  do n=&nmin to &nmax by &nby;
    f2 = 1-f1;
    b1 = sqrt(f1*&qa*(1-&qa))*&eff1;
    b2 = sqrt(f2*&qa*(1-&qa))*&eff2;
    Z1 = W1 + b1*sqrt(n);
    Z2 = W2 + b2*sqrt(n);
    Z = Z1*sqrt(f1) + Z2*sqrt(f2);
    ratio1_2 = (Z1/Z2)*sqrt(f2/f1);
    ratio1_12 = (Z1/Z)*sqrt(1/f1);
    output;
  end;
  end;
run;

proc sort data=sd2 out=sd3; by f1 n; run;

data sd4;
  set sd3;
  by f1 n;
  n1=round(n*f1);
  retain Zcnt uncA1 uncA2 uncA3 uncA4 jntA1 jntA2 jntA3 jntA4;
  if first.n then do;
     Zcnt=0;
          uncA1=0; uncA2=0; uncA3=0; uncA4=0;
          jntA1=0; jntA2=0; jntA3=0; jntA4=0;
  end;
```

```
  Zcnt = Zcnt + (Z>z1  alpha);
  uncA1 = uncA1 + (ratio1_2>=rho);
  uncA2 = uncA2 + (ratio1_12>=rho);
  uncA3 = uncA3 + (ratio1_2>=rho)*(ratio1_2=<1/rho);
  uncA4 = uncA4 + (ratio1_12>=rho)*(ratio1_12=<1/rho);
  jntA1 = jntA1 + (ratio1_2>=rho)*(Z>z1_alpha);
  jntA2 = jntA2 + (ratio1_12 >=rho)*(Z>z1_alpha);
  jntA3 = jntA3 + (ratio1_2>=rho)*(ratio1_2=<1/rho)*(Z>z1_alpha);
  jntA4 = jntA4 + (ratio1_12>=rho)*(ratio1_12=<1/rho)
          *(Z>z1_alpha);
  format uncPA1 uncPA2 uncPA3 uncPA4 jntPA1 jntPA2
          jntPA3 jntPA4 conPA1 conPA2 conPA3 conPA4 4.2;

  if last.n then do;
    P_Z = Zcnt/&ntrial;
    uncPA1 = uncA1/&ntrial;
    uncPA2 = uncA2/&ntrial;
    uncPA3 = uncA3/&ntrial;
    uncPA4 = uncA4/&ntrial;
    jntPA1 = jntA1/&ntrial;
    jntPA2 = jntA2/&ntrial;
    jntPA3 = jntA3/&ntrial;
    jntPA4 = jntA4/&ntrial;
    conPA1 = jntA1/Zcnt;
    conPA2 = jntA2/Zcnt;
    conPA3 = jntA3/Zcnt;
    conPA4 = jntA4/Zcnt;
    output;
  end;
run;

proc print;
          var  alpha z1_alpha beta z1_beta rho eff1 eff2 f1 f2 n1 n uncPA1 uncPA2
          uncPA3 uncPA4 jntPA1 jntPA2 jntPA3 jntPA4 conPA1 conPA2 conPA3
          conPA4;
run;
```

Appendix 3

A SAS program for evaluating conditional assurance probabilities for MCRT with three regions and requiring consistency in all three regions and in region 1 as in Table 6.4

```
*Input design parameters;
%let alpha=0.025;*One-sided significant level;
%let beta=0.10;  *Type 2 error (probability of power=1-beta);
%let rho =0.40;  *Consistent constant;
%let f1min=0.10; *Low bound of allocation ratio in Region 1;
%let f1max=0.30; *Upper bound of allocation ratio in Region 1;
%let f1by=0.10;  *Increase step for allocation ratio in Region 1;
%let f2min=0.10; *Low bound of allocation ratio in Region 2;
```

```
%let f2max=0.30; *Upper bound of allocation ratio in Region 2;
%let f2by=0.10;  *Increase step for allocation ratio in Region 2;
%let eff1=0.30;  *Treatment effect size (delta/sigma): region 1;
%let eff2=0.30;  *Treatment effect size (delta/sigma): region 2;
%let eff3=0.30;  *Treatment effect size (delta/sigma): region 3;
%let ntrial=50000; *Total number of simualted trials;
%let seed=778899;  *Seed for generating random numbers;

data d0;
  alpha=&alpha; z1_alpha=-probit(alpha);
  beta=&beta; z1_beta=-probit(beta);
  rho=&rho; eff1=&eff1; eff2=&eff2; eff3=&eff3;
  ntrial = &ntrial;
run;

proc iml;
  Mean = {0, 0, 0};
  Cov = {1 0 0, 0 1 0, 0 0 1};
  call randseed(&seed);
  W = RandNormal(&ntrial, Mean, Cov);
  VarNames = "W1":"W3";
  create sd0 from W[colname=varNames];
  append from W;
  close sd0;
quit;

data sd1;
  if _n_=1 then set d0; set sd0;
run;

data sd2;
  set sd1;
  do f1=&f1min to &f1max by &f1by;
  do f2=&f2min to &f2max by &f2by;
    f3 = 1 - f1 - f2;
          eff = f1*eff1 + f2*eff2 + f3*eff3;
          c1 = sqrt(f1)*(z1_alpha + z1_beta)*eff1/eff;
          c2 = sqrt(f2)*(z1_alpha + z1_beta)*eff2/eff;
          c3 = sqrt(f3)*(z1_alpha + z1_beta)*eff3/eff ;
          Z1 = W1 + c1;
          Z2 = W2 + c2;
          Z3 = W3 + c3;
          Z  = Z1*sqrt(f1) + Z2*sqrt(f2) + Z3*sqrt(f3);
          ratio1_123 = (Z1/Z)*sqrt(1/f1);
          ratio2_123 = (Z2/Z)*sqrt(1/f2);
          ratio3_123 = (Z3/Z)*sqrt(1/f3);
     output;
  end;
  end;
run;
```

```
data sd3; set sd2; case = 100*f1 + 10*f2 + f3; run;

proc sort data=sd3; by case; run;

data sd4;
  set sd3;
  by case;
  retain Zcnt CA2r1 CA4r1 CA5 CA6;
  if first.case then do;
          Zcnt=0; CA2r1=0; CA4r1=0; CA5=0; CA6=0;
  end;
  Zcnt = Zcnt + (Z>z1_alpha);
  CA2r1 = CA2r1 + (ratio1_123 >=rho)*(Z>z1_alpha);
  CA4r1 = CA4r1 + (ratio1_123 >=rho)*(ratio1_123 =<1/rho)
          *(Z>z1_alpha);
  CA5 = CA5 + (ratio1_123 >=rho)*(ratio2_123 >=rho)*
          (ratio3  123 >=rho)*(Z>z1  alpha);
  CA6 = CA6 + (ratio1_123 >=rho)*(ratio2_123 >=rho)*
          (ratio3_123 >=rho)*(ratio1_123 =<1/rho)*
          (ratio2_123 =<1/rho)*(ratio3_123 =<1/rho)*(Z>z1_alpha);
  if last.case then do;
    PZ = Zcnt/&ntrial;
    conPA2r1 = CA2r1/Zcnt;
    conPA4r1 = CA4r1/Zcnt;
    conPA5 = CA5/Zcnt;
    conPA6 = CA6/Zcnt;
    output;
  end;
run;

proc print;
    var  alpha z1  alpha beta z1_beta rho eff1 eff2 f1 f2 f3 PZ   conPA2r1 conPA4r1
    conPA5 conPA6;
run;
```

Appendix 4

A SAS program for evaluating conditional power for regional tests for an MRCT with four regions in Figure 7.1

```
*Input design parameters;
%let alpha=0.025; * One-sided significant level;
%let qa=0.5;      * Randomization ratio to Arm A (Qa);
%let alpha1=0.20; * One-sided significant level for region 1;
%let alpha2=0.20; * One-sided significant level for region 2;
%let alpha3=0.20; * One-sided significant level for region 3;
%let alpha4=0.20; * One-sided significant level for region 4;
%let eff1=0.40; * Treatment effect (delta/sigma): region 1;
%let eff2=0.40; * Treatment effect (delta/sigma): region 2;
%let eff3=0.40; * Treatment effect (delta/sigma): region 3;
%let eff4=0.40; * Treatment effect (delta/sigma): region 4;
```

```
%let nmin=180;    * Low bound of sample size n;
%let nmax=500;    * Upper bound of sample size n;
%let nby=20;     * Increase step in sample size;
%let f1=0.25;     * Allocation ratio in Region 1;
%let f2=0.25;     * Allocation ratio in Region 2;
%let f3=0.25;     * Allocation ratio in Region 3;
%let ntrial=50000;* Total number of simualted trials;
%let seed=778899; * Seed for generating random numbers;

data d0;
  alpha=&alpha;
  z1_alpha=-probit(alpha);
  z1_alpha1=-probit(&alpha1);
  z1_alpha2=-probit(&alpha2);
  z1_alpha3=-probit(&alpha3);
  z1_alpha4=-probit(&alpha4);
  eff1=&eff1; eff2=&eff2; eff3=&eff3; eff4=&eff4;
  ntrial = &ntrial;
run;

proc iml;
  Mean = {0, 0, 0, 0};
  Cov = {1 0 0 0, 0 1 0 0, 0 0 1 0, 0 0 0 1};
  call randseed(&seed);
  W = RandNormal(&ntrial, Mean, Cov);
  VarNames = "W1":"W4";
  create sd0 from W[colname=varNames];
  append from W;
  close sd0;
quit;

data sd1; if _n_=1 then set d0; set sd0; run;

data sd2;
  set sd1;
  f1=&f1; f2=&f2; f3=&f3; f4 = 1-f1-f2-f3;
  do n=&nmin to &nmax by &nby;
    b1 = sqrt(f1*&qa*(1-&qa))*&eff1;
    b2 = sqrt(f2*&qa*(1-&qa))*&eff2;
    b3 = sqrt(f3*&qa*(1-&qa))*&eff3;
    b4 = sqrt(f4*&qa*(1-&qa))*&eff4;
    Z1 = W1 + b1*sqrt(n);
    Z2 = W2 + b2*sqrt(n);
    Z3 = W3 + b3*sqrt(n);
    Z4 = W4 + b4*sqrt(n);
    Z  = Z1*sqrt(f1) + Z2*sqrt(f2) + Z3*sqrt(f3) + Z4*sqrt(f4);
    output;
  end;
run;

proc sort data=sd2 out=sd3; by f1 f2 f3 n; run;
```

```
data sd4;
  set sd3;
  by f1 f2 f3 n;
  retain Zcnt cdr1 cdr12 cdr123 cdr1234;
  if first.n then do;
          Zcnt=0; cdr1=0; cdr12=0; cdr123=0; cdr1234=0;
  end;
  Zcnt = Zcnt + (Z>z1_alpha);
  cdr1 = cdr1 + (Z>z1_alpha)*(Z1>z1_alpha1);
  cdr12 = cdr12 + (Z>z1_alpha)*(Z1>z1_alpha1)*(Z2>z1_alpha2);
  cdr123 = cdr123 + (Z>z1_alpha)*(Z1>z1_alpha1)*(Z2>z1_alpha2)*
          (Z3>z1_alpha3);
  cdr1234 = cdr1234 + (Z>z1_alpha)*(Z1>z1_alpha1)*(Z2>z1_alpha2)*
          (Z3>z1_alpha3)*(Z4>z1_alpha4);
  if last.n then do;
    pbZcnt = Zcnt/&ntrial;
    pbcdr1 = cdr1/Zcnt;
    pbcdr12 = cdr12/Zcnt;
    pbcdr123 = cdr123/Zcnt;
    pbcdr1234 = cdr1234/Zcnt;
    output;
  end;
run;

proc print;
    var  z1_alpha z1_alpha1 z1_alpha2 z1_alpha3 z1_alpha4 eff1 eff2 eff3 eff4 f1 f2 f3 f4
    n pbZcnt pbcdr1 pbcdr12 pbcdr123 pbcdr1234;
run;
```