Revisit Sample Size Estimation in Phase II Selection Designs

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

Abstract

The statistical methods for ranking and selection were introduced and used in the design of phase 2 oncology clinical trials, where subjects were randomized to several promising treatment arms with the goal to select one arm for further development. The methods were generalized to selection designs with time-to-event endpoint and different designs with binary endpoint. In order to facilitate its wider application, there need readily applicable methods for sample size calculation. For classical selection designs with binomial endpoint, we show that the sample size can be calculated exactly using a SAS program. For selection designs with time-to-event endpoint, we adapt Bechhofer's method for normal endpoint and estimate sample size of total number of events; and we further show via simulation the designs have approximately the specified correct selection probability. For a class of flexible selection designs with binary endpoint, we point out a possible flaw of the minimum advantage requirement based on response rate, and propose a new class of designs based on the number of observed responses and calculate sample size with specified correct selection probability.

Key Words: clinical trials, phase 2 selection designs, sample size, binary endpoint, time-to-event endpoint

1. Introduction

There are numerous challenges in conducting oncology clinical trials with limited resources, time, and patient population. One common practice is to screen a drug candidate with several promising dosing regimens in a Phase II setting using randomized selection designs and advance the best regimen for further development.

Bechhofer (1954) pioneered research on the selection designs and Gibbons et al. (1977) published a book on the topic. Simon et al. (1985) introduced the methodology to oncology field. Liu et al. (2006) summarized different kinds of selection designs in cancer trials. In addition, Liu et al. (1993, 1995) studies selection designs for time-to-event endpoint. Sargent & Goldberg (2001) proposed a flexible selection design. Steinberg et al. (2002) considered the selection designs with an early selection at an interim analysis. Cheung (2009) studied selection designs with an active control. Thall et al. (1989) proposed a two-stage design with selection at the first stage and comparison to the control at the second stage. Wason et al. (2012) and Wu et al. (2013) further studied multi-stage designs with interim treatment selection at the first stage. Jung (2013) discussed randomized phase 2 oncology clinical trials in more details.

In Section 2 for selection designs with binary endpoint, we show that the sample size can be calculated using exact binomial distribution for classical binary endpoint and a

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

computational program using SAS is provided. We verify sample sizes in Simon et al. (1985) and clarify a comment in Liu (2006). In Section 3, we review selection designs with normal endpoints from Bechhofer (1954) and provide constants for a range of correct selection probabilities. Section 4 is devoted to sample size estimation for selection designs with time-to-event endpoint. Using Bechhofer's theory and normal approximation to survival distribution, we provide sample size estimation and verify the correct selection probability with simulation examples. The method is applicable where fixed number of events is required and is an extension to Liu et al. (1993, 1995, 2006) where fixed accrual time and follow-up time are required. In Section 5, we point out a dilemma in a flexible selection design proposed by Sargent & Goldberg (2001). As an alternative, we propose a new class of selection designs with minimum advantage requirements and calculate sample sizes.

2. Selection Designs with Binary Endpoint

This section follows the approach in Simon et al. (1985), which was reviewed in Liu et al. (2006). For a trial with K arms, let the ordered response rate (range from 0 for no response and 1 for 100% response) be:

$$p_{[1]} \le p_{[2]} \le \cdots \le p_{[K-1]} \le p_{[K]}$$

The aim is to select the best arm (ie, the arm with the highest response rate $p_{[K]}$) with high probability. Assuming a fixed advantage δ (> 0) for the best arm over all the other arms, the least favourable configuration is:

$$p_{[1]} = p_{[2]} = \cdots = p_{[K-1]} = p_0, p_{[K]} = p_0 + \delta$$

In the following, we consider designs with $K = 2, 3, or\ 4$ arms and assume arm K is the best arm to be selected.

Let the correct selection probability be denoted as CSP, which has a lower bound P. We consider the cases where $P = 0.85, 0.90, 0.95$. The sample size determination uses the following criteria: for fixed p_0, δ, and P,

CSP ≥ P for selecting the superior arm (Arm K).

There are two scenarios of observed outcomes for correct selection of Arm K:

1) The observed response rate in Arm K is the highest among all arms, ie,

$$p_{[K]} > p_{[K-1]} \ge \cdots = p_{[1]}$$

Then the correct selection probability is the outcome probability.

2) The observed response rate in Arm K and the another J arms are the highest, ie,

$$p_{[K]} = p_{[K-1]} = \cdots = p_{[K-J+1]} > p_{[K-J]} \ge \cdots \ge p_{[1]}$$

Then the correct selection probability is $1/(J+1)$ of the outcome probability.

Simon et al. (1985) calculated sample sizes for $\delta = 0.15$ and $CSP \ge P = 0.90$ (Table 1) and Liu et al. (2006) made an arguable comment on Simon's result. Liu et al. thought Simon et al. used normal approximation and the real CSP ranged from 0.86 to 0.88. We verified that the CSP range from 0.86 to 0.88 used only cases in Scenario 1). If cases in both Scenario 1) and 2) are used as in Simon et al. (1985), the CSP will be above 0.90. For an example in two arm selection design, we plot correct selection probability (CSP) with sample size per arm (Figure 1). A SAS program for sample size calculation is attached in Appendix I.

Table 1: Sample Size per Arm for Binary Endpoint with K Arm Selection Designs and Correct Selection Probability 0.90 with Response Rate Difference 0.15

Response Rates		Sample Size per Arm (n)		
$p_1, \cdots, p_{K-1}$	p_K	$K = 2$	$K = 3$	$K = 4$
0.10	0.25	21	31	37
0.20	0.35	29	44	52
0.30	0.45	35	52	62
0.40	0.55	37	55	67
0.50	0.65	36	54	65
0.60	0.75	32	49	59
0.70	0.85	26	39	47

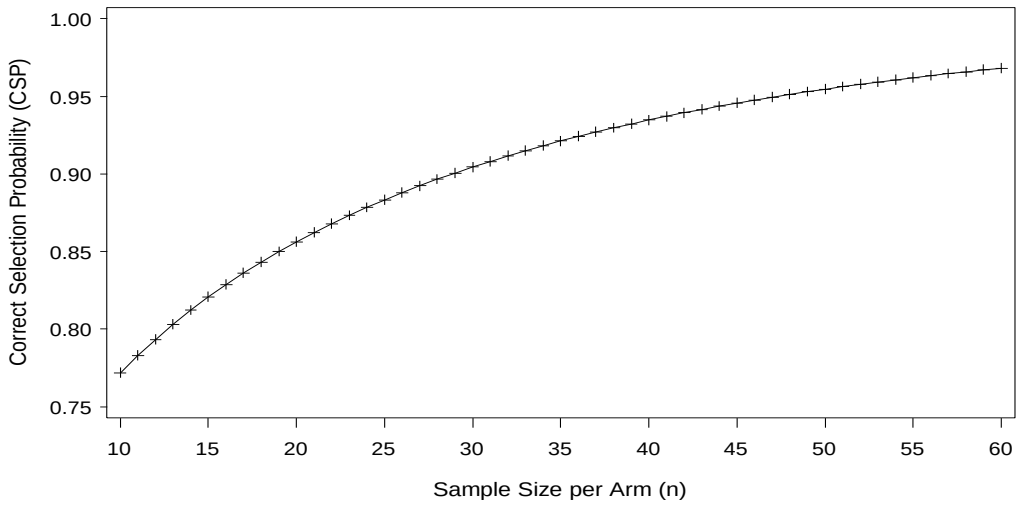

Figure 1: Plot of Correct Selection Probability (CSP) over Sample Size per Arm for Two Arm Selection Designs with Binary Endpoints and Response Rate 0.20 and 0.35

3. Selection Designs with Normal Endpoint

Bechhofer (1954) pioneered the research on the selection designs with normal endpoint. Gibbons et al. (1977) summarized results on selection designs. Let the clinical trial have K treatment arms and each arm enrol n subjects. The outcome X_{ij} for each subject be normally and independently distributed with mean μ_i and variance σ^2, ie

$$X_{i,j} \sim N(\mu_i, \sigma^2), \text{ where } i = 1, 2, \cdots, K;\ j = 1, 2, \cdots, n.$$

Let the means be ordered as $\mu_{[1]} \leq \mu_{[2]} \leq \cdots \leq \mu_{[K-1]} \leq \mu_{[K]}$. The selection design intended to select the largest mean $\mu_{[K]}$ with high correct selection probability (CSP) assuming its superior to other means by at least δ. The least favourable configuration is

$$\mu_{[1]} = \mu_{[2]} = \cdots = \mu_{[K-1]} = \mu_0, \mu_{[K]} = \mu_0 + \delta.$$

For fixed K (number of treatment arms) and δ (mean difference), the sample size can be determined exactly to achieve the desired correct selection probability (CSP). Bechhofer (1954) gave a formula for the per arm sample size by

$$\tau = \sqrt{n}\ {}^{\delta}/_{\sigma}$$

where τ is the constant determined by the following equation:

$$CSP = \int_{-\infty}^{+\infty} F(y+\tau)^{K-1} f(y)\, dy$$

with $F(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} \exp(-x^2/2)\, dx$ as the cumulative distribution function of the standard normal distribution and $f(y) = \frac{1}{\sqrt{2\pi}} \exp(-y^2/2)$ its density function. The constant τ can be calculated using numerical integration (eg, SAS or R). In Table 2, we list constant τ for $CSP = 0.95, 0.90, 0.85,$ and 0.80. We copied constant τ for $CSP = 0.95, 0.90, 0.85,$ and 0.80 from Bechhofer (1954) and calculated it for $CSP = 0.85$ using a SAS program. The sample size per arm can be calculated using the constants. Eg, in case of $K = 3$ arms, $CSP = 0.90$ and $\delta/\sigma = 0.3$, the sample size per arm is $n = 56$.

Table 2: Constant for Selecting the Normal Distribution with the Largest Mean

Correct Selection Probability (CSR)	Constant $\tau = \sqrt{n}\ {}^{\delta}/_{\sigma}$		
	$K = 2$	$K = 3$	$K = 4$
0.95	2.3262	2.7101	2.9162
0.90	1.8124	2.2302	2.4516
0.85	1.4658	1.9079	2.1394
0.80	1.1902	1.6524	1.8932

4. Selection Designs with Time-to-Event Endpoint

Liu et al. (1993, 1995, 2006) studied selection designs in Phase II trials with time-to-event endpoint and calculated sample sizes by fitting Cox proportional hazards model with exponential survival distribution and uniform censoring in studies with fixed enrolment and follow-up period. In this note, we provided an alternative design with fixed number of events using an asymptotic method for sample size estimation with constants calculated in the previous section for normal endpoint.

For a selection trial with K arms and a time-to-event (survival) endpoint, we assume the endpoints in arm $1, 2, \cdots, K$ follow exponential distributions with constant parameters (hazard rates) $\lambda_1, \cdots, \lambda_{K-1}, \lambda_K$ respectively. The selection trial is to be designed with high probability to select the best arm, ie, the arm with the smallest hazard rate. Our aim is to estimate sample size based on the total number of events from all arms combined (Note: this is different from binomial endpoint). For the sample size estimation, we assume all subjects will be followed to their events (eg, no censoring allowed). We will estimate sample size first and subsequently use simulations to demonstrate that the designs have the specified correct selection probability (CSP).

For designs with Karms, assume the superior arm have a fixed advantage over all other arms, ie, the hazard ratios of the superior arm over other arms are at most r $(0 < r < 1)$. The least favourable configuration is:

$$\lambda_{[1]} = \cdots = \lambda_{[K-1]} = \lambda_0, \lambda_{[K]} = r\lambda_0.$$

Let n be the number of subjects in each arm. Then the estimation of log hazard for each arm will act like mean estimation for the normal case, which follows normal distribution asymptotically:

$$\log(\widehat{\lambda_k}) \sim N\left(\log(\lambda_{[k]}),\ \frac{1}{n}\right), k = 1, \cdots, K.$$

For a given correct selection probability (CSR), let τ be the constant as defined in the previous section for the normal endpoint. The parameters satisfy equation $\tau = \sqrt{n}\ \delta/\sigma$, where $\delta = \log(\lambda_{[0]}) - \log(\lambda_{[K]}) = -\log(r)$ as one sample mean difference and $\sigma = 1$ as one sample variance respectively. The estimated total number of events for all Karms satisfying the equation

$$Kn = K(\ \tau/\log(r))^2.$$

Using the above formula, we calculate total number of event for $K = 2, 3,$ and 4 arms and for hazard ratios between the superior arm and other arms range from 0.05 to 0.80 (Table 3). In general, sample sizes using our method are consistent with Liu et al. (1993, 1995, and 2006). The case with a hazard ratio of 0.667 in Table 3 below is identical to the case with a hazard ratio of 1.5 from Liu et al. (2006). For $K = 2, 3,$ and 4 arms, they provided number of events for the worst arm as 24, 36 and 43 respectively; while our calculation provide the total number of events for all arms as 40, 91 and 147 respectively.

Table 3: Total Number of Events for Selection Designs with Time-to-Event Endpoints and Correct Selection Probability (CSP) 0.90 and $K = 2, 3,$ and 4 Arms

No. of Arms	Total Number of Events for Hazard Ratio r (Arm K vs. Arm $1, \cdots, K-1$)					
	$r = 0.8$	$r = 0.75$	$r = 0.7$	$r = 0.667$	$r = 0.6$	$r = 0.5$
$K = 2$	132	80	52	40	26	14
$K = 3$	300	181	118	91	58	32
$K = 4$	483	291	189	147	93	51

Using simulations, we demonstrated that the Correct Selection Probability (CSP) for selection designs with sample size in Table 3 ranges from 0.89 to 0.91. We assume that Arm 1 to $(K-1)$ have median survival 6 months, enrollment evenly in 12 months, about 25% more patients enrolled than the total number of events, and the trial is followed to the required total number of events.

In Table 4, we present estimated CSP from simulation for hazard ratio $r = 0.667$ to 0.8. For each case, we performed 5000 simulated trials with follow-up to the required number of events for estimating CSR.

We also estimate total number of events for selection designs with correct selection probability 0.95 and 0.85. The sample sizes are presented in Table 5 and 6.

Table 4: Correct Selection Probability (CSP) by Simulation for Designs in Table 3

No. of Arms	Hazard Ratio r (Arm K vs. Arm $1, \cdots, K-1$)							
	$r = 0.8$		$r = 0.75$		$r = 0.7$		$r = 0.667$	
	Ne/Np	CSP	Ne/Np	CSP	Ne/Np	CSP	Ne/Np	CSP
$K = 2$	132/166	0.90	80/100	0.91	52/66	0.90	40/50	0.90
$K = 3$	300/375	0.90	181/228	0.90	118/150	0.91	91/114	0.89
$K = 4$	483/604	0.89	291/364	0.91	189/240	0.89	147/184	0.90

Ne/Np = total number of events/total number of subjects.

Table 5: Total Number of Events for Selection Designs with Time-to-Event Endpoints and Correct Selection Probability (CSP) 0.95 and $K = 2, 3,$ and 4 Arms

No. of Arms	Total Number of Events for Hazard Ratio r (Arm K vs. Arm $1, \cdots, K-1$)					
	$r = 0.8$	$r = 0.75$	$r = 0.7$	$r = 0.667$	$r = 0.6$	$r = 0.5$
$K = 2$	218	131	86	65	42	23
$K = 3$	443	267	174	135	85	46
$K = 4$	684	412	268	207	131	71

Table 6: Total Number of Events for Selection Designs with Time-to-Event Endpoints and Correct Selection Probability (CSP) 0.85 and $K = 2, 3,$ and 4 Arms

No. of Arms	Total Number of Events for Hazard Ratio r (Arm K vs. Arm $1, \cdots, K-1$)					
	$r = 0.8$	$r = 0.75$	$r = 0.7$	$r = 0.667$	$r = 0.6$	$r = 0.5$
$K = 2$	87	52	34	27	17	9
$K = 3$	220	132	86	67	42	23
$K = 4$	368	222	144	112	71	39

5. Selection Designs for Binary Endpoint with Minimum Advantage Requirement

Sargent & Goldberg (2001) proposed a new type of selection designs which selects the best arm only if its observed response rate over other arms is above a fixed threshold *d*, and if the advantage $\leq d$, other factors will be considered for the selection. They called those designs as flexible designs and considered $d = 0, 0.025, 0.05$. Such selection designs allow flexibility in decision making. Liu et al. (2006) renamed them as designs with minimum advantage requirements.

Our calculations indicate that correct selection probability (CSP) does not monotonously increase with sample size when the threshold d is based on response rates. Figure 2 plots the case for $K = 2$ arms, response rates 0.20 and 0.35 ($\delta = 0.15$, and the minimum advantage $d = 0.05$. When sample size increases from 19 to 20, 39 to 40, 59 to 60, CSP decreases.

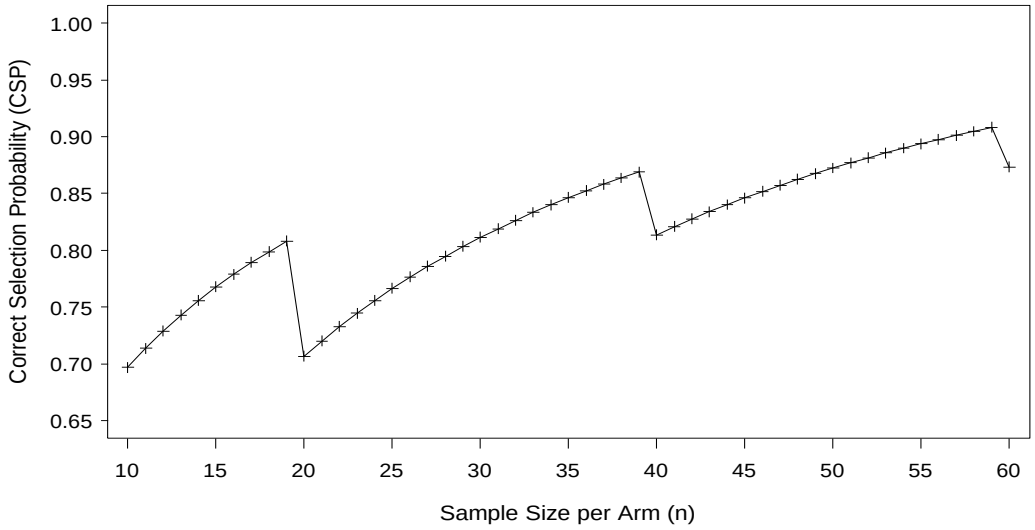

Figure 2: Plot of correct selection probability (CSP) over sample size per arm for two arm selection designs with binary endpoint and response rate 0.35 and 0.20, the minimum advantage requirement in difference of the observed response rate > 0.05 (d)

We present an alternative selection design where "Minimum Advantage Requirements" is based on number of observed responses rather than based on response rate. With the specified correct selection probability, we give an exact method to calculate the sample sizes based on binomial distributions. We suggest selecting the observed best arm if the number of observed response is more than a pre-fixed non-negative integer m. For the minimum advantage requirement $m = 2, \delta = 0.15$, and $K = 2$ arms, Tables 7 shows the sample sizes per arm. Figure 3 displays relationship between CSP and sample sizes for $m = 0, 1, 2$. All calculations are based on the exact binomial distribution implemented in a SAS program (Appendix II).

Table 7: Sample Size per Arm for Response Rate Difference 0.15 and the Minimum Advantage Requirement > 2 (m) in the Observed Response Between the two Arms

Response Rates		n per Arm		
		Correct Selection Probability		
p_1	p_2	0.90	0.85	0.80
0.10	0.25	48	40	34
0.20	0.35	57	46	39
0.30	0.45	63	50	41
0.40	0.55	65	52	43

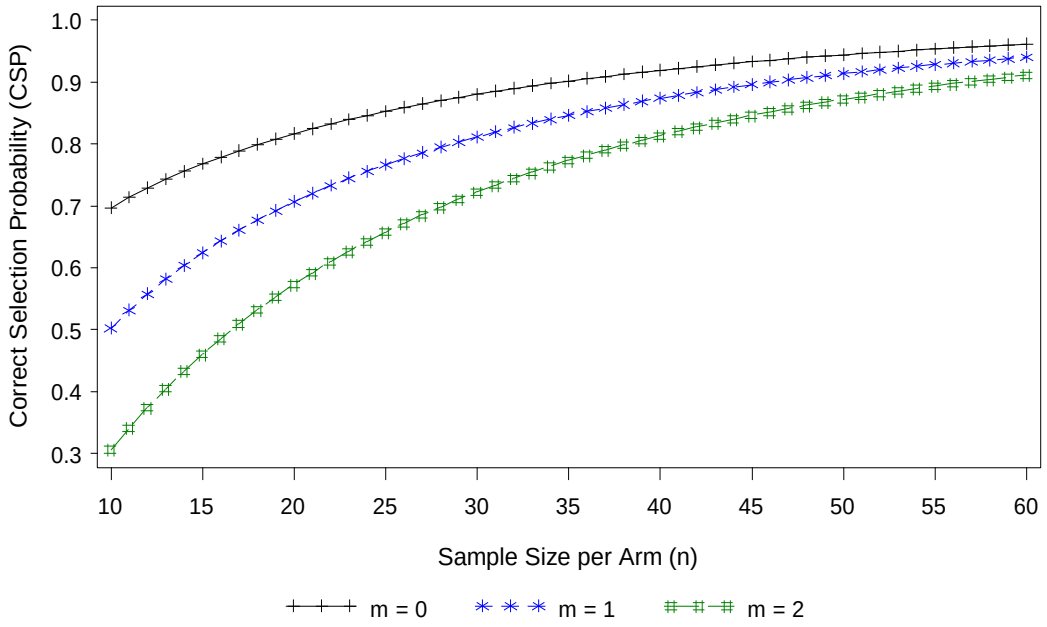

Figure 3: Plot of correct selection probability (CSP) over sample size per arm for two arm selection designs with binary endpoints and response rate 0.35 and 0.20 and the minimum advantage requirement in the difference of observed responses $> m$ for selection

6. Discussion and Conclusion

Randomized selection designs are useful when there are more than one treatment regimens to select. If there is a superior arm, the design provides high probability for its selection. In this note, we make additional sample size calculations using SAS program with exact binomial distribution for classical selection designs to supplement the work by Simon et al. For selection designs with survival endpoints, it is generally desirable to follow the trial to a fixed number of events for all arms. By assuming exponential

distribution of survival time and adapting Bechhofer's method on selection designs with normal endpoint, we develop a method to estimate sample size of total number of events for all arms and further verify the design have desirable correct selection probability using simulations. We propose a new class of flexible designs with binary endpoints and give an exact method calculating sample size with specified correct selection probability based on binomial distributions.

Acknowledgements

Nguyen V. Dat and Gordon Sun reviewed an earlier version of the paper and provided valuable and insightful comments, which was presented in the 7th International Conference on Multiple Comparison Procedures, MCP 2011 in Washington DC.

References

Simon RM, Wittes RE, Ellenberg S (1985). Randomized phase II clinical trials. *Cancer Treatment Reports* 69: 1375−1381.

Liu PY, Moon J, LeBlanc M (2006). Phase II selection designs. *Handbook of Statistics in Clinical Oncology*, Second Edition, (J Crowley, DP Ankerst eds), Chapman & Hall/CRC: 155−164.

Liu PY, Dahlberg S, Crowley, J (1993). Selection designs for pilot studies based on survival. *Biometrics* 49: 853−862.

Liu PY, Dahlberg S (1995). Design and analysis of multiarm clinical trials with survival endpoints. *Controlled Clinical Trials* 16: 119−130.

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

Bechhofer R (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *The Annals of Mathematical Statistics* 25: 16−39.

Gibbons JD, Olkin I, Sobel M (1977). *Selecting and Ordering Populations: A New Statistical Methodology*. Wiley: New York.

Sargent DJ, Goldberg RM (2001). A flexible design for multiple armed screening trials. *Statist. Med.* 20: 1051−1060.

Lui KJ (2002). Letter to the editor: A flexible design for multiple armed screening trials. *Statist. Med.* 21: 625−627.

Sargent DJ (2002). Author's reply to letter to the editor. *Statist. Med.* 21: 628.

Steinberg SM, Venzon, DJ (2002). Early selection in a randomized phase II clinical trials. *Statist. Med.* 21: 1711−1726.

Thall PF, Simon R, Ellenberg SS (1989). A two-stage design for choosing among several experimental treatments and a control in clinical Trials, *Biometrics* 45: 537-547.

Cheung YK (2009). Selecting promising treatments in randomized phase II cancer trials with an active control, *J. Biopharm. Statistics* 19: 494-508.

Wason JMS, Jaki T (2012). Optimal design of multi-arm multi-stage trials, *Statist. Med.* 31: 4269-4279.

Wu Y, Zhao PL (2013). Interim treatment selection with a flexible selection margin in clinical trials, *Statist. Med.* 32: 2529-2543.

Jung SH (2013). *Randomized Phase II Cancer Clinical Trials*. Chapman & Hall/CRC: Boca Raton, FL.

Appendix I

A SAS program for calculating sample size per arm in selection designs with binary endpoint in Section 1. The parameter values assigned in the macro variables can be changed for other cases.

```
%let K = 3;              * Number of treatment arms;
%let p0 = 0.20;          * Response rate in the (K-1) inferior arms (Arm 1 to K-1);
%let delta = 0.15;       * Response rate for the superior Arm K is p0 + delta;
%let cspmin = 0.90;      * Lower bound of correct selection probability (CSP);
%let nmax = 100;         * Upper bound of sample size per arm;

data dd1;
  do n=1 to &nmax;
    csp1=0;
    do i=0 to n;
      if i=0 then fi=cdf('BINOMIAL',i,&p0,n)**(&K-1);
      else if i>0 then fi=cdf('BINOMIAL',i,&p0,n)**(&K-1) –
          cdf('BINOMIAL',I 1,&p0,n)**(&K-1);
            csp1 = csp1 + fi*(1 - cdf('BINOMIAL',i,&p0+&delta,n));
   end;
   output;
  end;
run;

data dd2;
  do n=1 to &nmax;
    csp2=0;
    do i=0 to n; do j=1 to &K-1;
      if i>0 then g = fact(&K-1)/(fact(j)*fact(&K-1-j))*(cdf('BINOMIAL', i-1,&p0,n)
       **(&K-1*(pdf('BINOMIAL',i,&p0,n)**j)*pdf('BINOMIAL',i,&p0+&delta,n)/(j+1);
      else if i=0 and j=(&K-1) then g = (pdf('BINOMIAL',i,&p0,n)**j)
        *pdf('BINOMIAL',i,&p0+&delta,n)/(j+1);
       else if i=0 and j<(&K-1) then g = 0;
             csp2 = csp2 + g;
   end; end;
   output;
 end;
run;

data dd3(keep=n csp1 csp2 csp);
 merge dd1 dd2;
 by n;
 csp = csp1 + csp2;
 if csp>=&cspmin then output;
run;

proc print; run;
```

Appendix II

A SAS program for calculating sample size of selection designs with minimum advantage requirement for binary endpoint and 2 arms in Section 5. The parameter values assigned in the macro variables can be changed for other cases.

```
%let K=2;          * Number of treatment arms (K=2);
%let p0=0.20;      * Response rate in the inferior arm;
%let delta=0.15; * Response rate for the superior is p0 + delta;
%let cspmin=0.90; * Lower bound of correct selection probability (CSP);
%let m=2;          * Selecion the superior arm if the response difference > m;
%let pho=0;        * Weight for ambiguous region probability (0<pho<1, default value 0);
%let nmax=100; * Upper bound of sample size per arm;

data dd1;
  do n=1 to &nmax;
    do x=0 to n; do y=0 to n;
      pxy = pdf('BINOMIAL',x,&p0 + &delta,n)*pdf('BINOMIAL',y,&p0,n);
      output;
    end; end;
  end;
run;

data dd2;
  set dd1;
  if ((x-y) > &m) then corr = pxy;
  else if (-&m =< (x-y) =< &m) then ambg = pxy;
run;

proc sql;
  create table dd3 as select *,
    sum(corr) as csp1,
    sum(ambg) as csp2
  from dd2
  group by n
  order by n;
quit;

proc sort data=dd3 out=dd4(keep = n csp1 csp2) nodupkey;
  by n;
run;

data dd5;
  set dd4;
  csp = csp1 + csp2*&pho;
  if csp >= &cspmin then output;
run;

proc print; run;
```