Sunday, March 19th, 2023

ENAR 2023 Spring Meeting @ Nashville, TN

Supplementary Materials

Tasuku Inao
Dr. Isao Yokota

Department of Biostatistics
Hokkaido University, Japan

Recent papers in Phase II design

The number of papers in 2005, 2010, 2014 Ivanova et al. J Biopharm Stat 2016

JCO, Ann Oncol, Cancer, CCR, Lancet Onc, NEJM, Lancet

	2005 (n=141)	2010 (n=150)	2014 (n=56)	Total (n=347)
Multi-arm	50 (35%)	51 (34%)	28 (50%)	129 (37%)
Single arm	91 (65%)	99 (66%)	28 (50%)	218 (63%)
Single stage	41 (45%)	44 (44%)	10 (36%)	95 (44%)
Simon	39 (43%)	35 (35%)	13 (46%)	87 (40%)
Fleming	2 (2%)	7 (7%)	3 (11%)	12 (6%)
Bayesian	1 (1%)	6 (6%)	0 (0%)	7 (3%)
Gehan	5 (6%)	1 (1%)	1 (3%)	7 (3%)
Flexible	0 (0%)	3 (3%)	0 (0%)	7 (3%)

How to calculate thresholds

$$\alpha = \sum_{i=u}^{K} NB(u,\ i;\ p_0) \quad \mathrm{pw} = \sum_{i=u}^{K} NB(u,\ i;\ p_1) \quad ※\ NB(S_k,\ k;\ p) = {}_{k-1}C_{S_k-1}\ p^{S_k}\ (1-p)^{k-S_k}$$

e.g. $(p_0, p_1) = (0.1,0.35)$, nominal α error rate : 0.025, nominal power: 0.8

		u					
		4		5		6	
K	19	α=0.11	pw=×	α=0.035	pw=×	α=0.009	pw=0.70
	20	α=0.13	pw=×	α=0.04	pw=×	α=0.011	pw=0.75
	21	α=0.15	pw=×	α=0.05	pw=×	α=0.014	pw=0.799
	22	α=0.17	pw=×	α=0.06	pw=×	α=0.018	pw=0.83
	23	α=0.19	pw=×	α=0.07	pw=×	α=0.023	pw=0.86

Find the pair (u, K) with smallest K with pw $\geq$ 0.8 under $\alpha \leq 0.025$.

$(u, K) = (4,9)$ in this setting.

So, threshold for futility stopping is $\{l_{17}, l_{18}, l_{19}, l_{20}, l_{21}, l_{22}\} = \{0,1,2,3,4,5\}$.

Simon's 2-stage design

2-stage design for futility stopping Simon, R. Control Clin Trials 1989

Stage1: Check the response for N_1 subjects ➡ Total of responders $\leq r_1$

→ Futility stopping

Stage2: Check the response for N subjects ➡ Total of responders $\leq r$

→ Not reject H_0

⬇

Reject H_0

○ (r_1, N_1, r, N)

Many pairs are satisfied in Simon's design ➡

Minimax design Min N

Optimal design Min $E[N|p_0]$

The formulation of Simon's 2-stage design

N_i : sample size by i th interim analysis

r_i : threshold by i th interim analysis

$X_1 \sim Bin(N_1, p)$, $X_2 \sim Bin(N_2, p)$, where X_i is the number of responders

The probability at 1st interim analysis

$$P_1(p) = P(X_1 \le r_1) = Bin(r_1 \mid N_1, p) = \sum_{i=0}^{r_1} {}_{N_1}\mathrm{C}_i \; p^i (1-p)^{N_1 - i}$$

The probability at final analysis

$$P_2(p) = P(X_1 \le r_1 \; or \; X_1 + X_2 \le r)$$

$$= Bin(r_1 \mid N_1, p) + \sum_{j=r_1+1}^{min(N_1, r)} {}_{N_1}\mathrm{C}_j \; p^j (1-p)^{N_1 - j} Bin(r - j \mid N_2, p)$$

The condition satisfying with design is $[(P_2(p_0) \ge 1 - \alpha) \wedge (P_2(p_1) \le \beta)]$

Multiplicity

Multiplicity : potential inflation of type I error rate
as a result of multiple testing

e.g. Nominal α error rate is 0.025

As using nominal level in each test,
H_0 may be rejected by mistake.

Test	Overall α error rate
1	0.025
2	0.049
3	0.073

○ Lan-DeMets' α-spending function Lan, K. K. G. et al. Biometrika 1983

To keep overall α error rate nominal level,
nominal α error rate is divided into the number of analysis.

e.g.

Test	Requirement of trials	Overall α error rate
1	Reject in Test 1	α_1
2	Reject in Test 2, not in Test 1	$\alpha_1 + \alpha_2$
⋮	⋮	⋮
K	Reject in Test K, not from 1 to K-1	$\alpha_1 + \alpha_2 + \ldots + \alpha_K = 0.025$

Lan-DeMets' α-spending function

O'Brien-Fleming design for efficacy stopping

Keeping the test statistic constant in a trial

O'Brien, P. C. et al. Biometrics 1979
Lan, K. K. G. et al. Biometrika 1983

Stage 1: Check the response for N_1 subjects ➡ $|Z_1| > Z_{OF_1}$ Efficacy stopping

⋮

Stage K: Check the response for $(N_1 + N_2 + \ldots N_K)$ subjects ➡ $|Z_K| > Z_{OF_K}$

⬇

Inability to obtain enough evidence to reject H_0

Z_k : Test statistic in k th interim analysis calculating from the data

Z_{OF_k} : Test statistic by O'Brien-Fleming design in k th interim analysis

Fixed sample size in O'Brien-Fleming design

2-type setting for fixed sample size

1. Wald test statistic (OF-Wald)

$$N = \frac{(z_\alpha + z_\beta)^2 p_1(1-p_1)}{(p_1 - p_0)^2} \qquad \left(\because \quad Z = \frac{\sqrt{N}(\hat{p} - p_0)}{\sqrt{\hat{p}(1-\hat{p})}} \right)$$

2. Score test statistic (OF-Score)

Fleming, T. R. Biometrics 1979

$$N = \left(\frac{z_\alpha \sqrt{p_0(1-p_0)} + z_\beta \sqrt{p_1(1-p_1)}}{p_1 - p_0} \right)^2 \qquad \left(\because \quad Z = \frac{\sqrt{N}(\hat{p} - p_0)}{\sqrt{p_0(1-p_0)}} \right)$$

z_α : $100(1-\alpha)$th percentile of the standardized normal distribution

Sample size calculation from test statistic

e.g. Wald Statistics

$$Z_0^{\text{Wald}} \mid H_1 \sim N\left(\frac{\sqrt{N}(p_1 - p_0)}{\sqrt{p_1(1-p_1)}}, 1\right), \text{ where } Z_0^{\text{Wald}} = \frac{(\hat{p} - p_0)}{\sqrt{\hat{p}(1-\hat{p})}/\sqrt{N}}$$

$$1 - \beta = P(Z_0^{\text{Wald}} > z_\alpha \mid H_1) = 1 - P(Z_0^{\text{Wald}} \leq z_\alpha \mid H_1)$$

$$= 1 - P\left(Z + \frac{\sqrt{N}(p_1 - p_0)}{\sqrt{p_1(1-p_1)}} \leq z_\alpha\right) = 1 - P\left(Z \leq z_\alpha - \frac{\sqrt{N}(p_1 - p_0)}{\sqrt{p_1(1-p_1)}}\right)$$

$$= 1 - \Phi\left(z_\alpha - \frac{\sqrt{N}(p_1 - p_0)}{\sqrt{p_1(1-p_1)}}\right)$$

z_α : $100(1-\alpha)$th percentile of the standardized normal distribution

The test statistics of O'Brien-Fleming design

1. α calculation in each test

α-spending function : $\alpha(t) = \min\left[2 - 2\Phi\left(\frac{Z_{\frac{\alpha}{2}}}{\sqrt{t}}\right), \alpha\right], \quad t \in [0,1]$

$\boldsymbol{t} = (t_1, t_2, t_3, \ldots, t_K)$ Information time

$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_K)$ α_k : The additive α by test k

The spending α in each test: $\left(\alpha_1, \alpha_2 - \alpha_1, \ldots, \alpha_K - \sum_{i=1}^{K-1} \alpha_i\right) = (\alpha_1', \alpha_2', \ldots, \alpha_K')$

2. The test statistics calculation in each test

① $P(U_1 > Z_{OF_1} \mid H_0) = \alpha_1' \Rightarrow Z_{OF_1} = \Phi\left(1 - \frac{\alpha_1'}{2}\right)$

② $P(U_1 \le Z_{OF_1} \text{ and } U_2 > Z_{OF_2} \mid H_0) = \alpha_2' \Rightarrow \alpha_2' = \int_{Z_{OF_2}}^{\infty} \int_{-\infty}^{Z_{OF_1}} f(U_1, U_2)\, dU_1\, dU_2$

f : bivariate normal distribution

mean vector (0,0) , Covariance Matrix $\left(1, \sqrt{\frac{t_i}{t_j}}, \sqrt{\frac{t_i}{t_j}}, 1\right)$

The thresholds of O'Brien-Fleming design

e.g. Wald Statistics ($p_0 = 0.1$, $p_1 = 0.35$)

Lachin. Control Clin Trials 1981

$$Z = \frac{\sqrt{N}(\hat{p} - p_0)}{\sqrt{\hat{p}(1-\hat{p})}}$$

※ Diagonal value are zero

→ Substitute a test statistic based on the mean weighted midpoint in the score interval

		N (sample size)						
		4	5	6	7	8	9	…
Responder	4	4.26	3.91	2.94	2.52	2.26	2.07	…
	5		4.92	4.81	3.59	3.06	2.75	…
	6			5.57	5.72	4.24	3.60	…
	7				6.23	6.62	4.89	…
	8					6.88	7.53	…
Test statistic		6.13	5.46	4.97	4.59	4.28	4.03	…
Threshold		×	×	6	6	6	7	…

Thresholds: The minimum number of responders for which $|Z| >$ (test staistic)

Estimation (Point Estimator)

Motivation

- Reference to design the future phase III trial and other phase II trial
  - The naive point estimator which is not taken into account of an interim monitoring may be biased. Guo, H.Y., and Liu, A.J. Biopharm Statist 2005.

Notation

- (M, S): Random variables that terminate the trial at the M th interim analysis with the number of response S.

Mass probability function with support S Jung, S.-H., and Kim, K.M. Stat Med 2004.

$$f(m,s) = c_{m,s} p^s (1-p)^{m-s} \quad \text{※ } S = \bigcup_{m=1}^{K} S_m = \bigcup_{m=1}^{K} \{(m,s) \mid l_{m-1} + 1 \le s \le l_m \text{ or } s = u\}$$

$$\text{※ } c_{m,s} = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_m} \binom{1}{x_1} \binom{2}{x_2} \cdots \binom{m}{x_m}$$

Naive estimator : $\hat{\pi}_n = \frac{s}{m}$

Guo, H.Y., and Liu, A.J. Biopharm Statist 2005.

Bias-adjusted estimator : $\hat{\pi}_g = \hat{\pi}_n - B(\hat{\pi}_n) \quad B(p) = E(\hat{p} \mid p) - p = \sum_{(m,s) \in \mathcal{S}} \hat{p}\, f(m,s \mid p) - p$

Point Estimator

The reasons to take up another estimator

- The undesirable cases because of bias-variance trade-off
- Flexible setting Robertson et al. arXiv 2021.

Median unbiased estimator : $\hat{\pi}_h = \frac{1}{2}(\pi_h^- + \pi_h^+)$ Jovic, G., and Whitehead, J. Stat Med 2010.

δ: step size, 0.0001, for example

※ $\pi_h^- = \begin{cases} \pi_{h_0}^-, & \mathrm{abs}(0.5 - P(\pi_{h_0}^-)) \leq \mathrm{abs}(P(\pi_{h_1}^-) - 0.5) \\ \pi_{h_1}^-, & \mathrm{abs}(0.5 - P(\pi_{h_0}^-)) \geq \mathrm{abs}(P(\pi_{h_1}^-) - 0.5) \end{cases}$ $\pi_{h_0}^- = \left\{\max\ \pi_h \,|\, P(\pi_h) \leq 0.5\right\} \quad \pi_{h_1}^- = \pi_{h_1}^- + \delta$

※ $\pi_h^+ = \begin{cases} \pi_{h_0}^+, & \mathrm{abs}(0.5 - P(\pi_{h_0}^+)) \leq \mathrm{abs}(P(\pi_{h_1}^+) - 0.5) \\ \pi_{h_1}^+, & \mathrm{abs}(0.5 - P(\pi_{h_0}^+)) \geq \mathrm{abs}(P(\pi_{h_1}^+) - 0.5) \end{cases}$ $\pi_{h_0}^+ = \left\{\max\ \pi_h \,|\, Q(\pi_h) \leq 0.5\right\} \quad \pi_{h_1}^+ = \pi_{h_1}^+ + \delta$

2-type P-value functions with stage-wise ordering

$$P(p) = \begin{cases} Pr((M,S) \succeq (m,s)\,|\,p) & \text{if } s \geq u \\ 1 - Pr((M,S) \prec (m,s)\,|\,p) & \text{if } s \leq l_k \end{cases} \qquad Q(p) = \begin{cases} Pr((M,S) \preceq (m,s)\,|\,p) & \text{if } s \geq u \\ 1 - Pr((M,S) \succ (m,s)\,|\,p) & \text{if } s \leq l_k \end{cases}$$

※ $A \succeq B$ means that A is latter or at the same stage as B in this ordering.

e.g. $(u, K) = (6,22)$: $(17,0) \preceq (18,1) \preceq \ldots \preceq (22,5) \preceq (22,6) \preceq (21,6) \preceq \ldots \preceq (6,6)$

Estimation (Confidence Interval)

Two-sided (1-2α) confidence interval $(p_L, p_U) = (p_L(S), p_U(S))$

Clopper-Pearson type:

$$Pr[S \leq s \,|\, p = p_U(s)] = \sum_{i=0}^{S} \binom{m}{i} p_U^i (1-p_U)^k - i = \alpha$$

$$Pr[S \geq s \,|\, p = p_L(s)] = \sum_{i=S}^{m} \binom{m}{i} p_L^i (1-p_L)^k - i = \alpha$$

Clopper, C.J., and Pearson, E.S. Biometrika 1934.

Jennison-Turnbull type:

$$Pr[(M,S) \succeq (m,s) \,|\, p = p_U(s)] = \alpha$$

$$Pr[(M,S) \preceq (m,s) \,|\, p = p_L(s)] = \alpha$$

Jennison, C., and Turnbull, B.W. Technometrics 1993.

○ mid-p approach

Berry, G., and Armitage, P. J R Stat Soc Ser D 1995.
Porcher, R., and Desseaux, K. BMC Med Res Methodol 2012.

Clopper-Pearson type using mid-p approach:

$$Pr[S < s \,|\, p = p_U(s)] + \frac{1}{2} Pr[S = s \,|\, p = p_U(s)] = \alpha$$

$$Pr[S > s \,|\, p = p_L(s)] + \frac{1}{2} Pr[S = s \,|\, p = p_L(s)] = \alpha$$

Jennison-Turnbull type using mid-p approach:

$$Pr[(M,S) \succ (m,s) \,|\, p = p_U(s)] + \frac{1}{2} Pr[(M,S) = (m,s) \,|\, p = p_U(s)] = \alpha$$

$$Pr[(M,S) \prec (m,s) \,|\, p = p_L(s)] + \frac{1}{2} Pr[(M,S) = (m,s) \,|\, p = p_L(s)] = \alpha$$

Confidence Interval

○ Interval from the perspective on acceptance region

Fixed Design	Group Sequential Design
Sterne's interval	Multistage Modified Sterne interval by Duffy and Santner
・Crow's interval ・Blyth and Still's interval	Multistage Crow-Blyth-Still interval by Duffy and Santner

Sterne, T.E. Biometrika 1954. Blyth, C.R., and Still, H.A. Am Stat 1983.

Multiple hypotheses

$$H_0^* : p = p_0, \quad H_1^* : p \neq p_0$$

Base method by Sterne(1954)

A confidence set for p by inverting the family of acceptance regions $A(p_0)$.

$A(p_0)$: Set of minimum cardinality as it contains the most likely outcomes under p_0.

Pointing out and modification

・Crow(1956) ⇒ Sterne's interval

1. Some lower limits $L(p_0)$ and upper limits $U(p_0)$ in $A(p_0)$ cannot be decreasing in p_0.
2. There are a value s not to be included in any $A(p_0)$.

・Blyth and Still (1983) ⇒ Crow's interval

It is not satisfied the assumption of monotonicity like the first criticism to Sterne's method.

Confidence Interval

Duffy and Santner's method

1. A linear ordering of outcomes (M, S) is performed by ordering by the success proportion (S/M). If the outcomes have the same ratio order, the tie is broken by stage-wise ordering. This ordering is denoted as follows: $(g,s)_1 \prec (g,s)_2 \prec \ldots$

2. Set a partition $0 < p_1 < \ldots < p_r < 1$, where the mesh is sufficiently small, like $r = 10{,}000$.

3. Set $A(p_1)$ is initialized to be $\{(g,s)_1\}$; equivalently, $L(p_1) = U(p_1) = \{(g,s)_1\}$.

4. Subsequent $L(p_i)$ and $U(p_i)$ are determined from previous $L(\,.\,)$ and $U(\,.\,)$ to satisfy:

   (i): $L(p_i) \geq L(p_{i-1})$.

   (ii): $U(p_i) \geq U(p_{i-1})$.

   (iii): $A(p_i) = \{L(p_i), \ldots, U(p_i)\}$ has coverage probability $1 - \alpha$.

   (iv): $A(p_i)$ is as small as possible in the sense of cardinality.

   ※ For property (iv), we try to do next 3 steps:

   (iv)-1: **Elimination:** Eliminate one or more outcomes from $A(p_{i-1})$ beginning with $L(p_{i-1})$.

   (iv)-2: **Substitution:** If (iv-1) is not possible, substitute $U(p_{i-1}) + 1$ for $L(p_{i-1})$.

   (iv)-3: **Adds:** If both are not possible, add points to $A(p_{i-1})$., beginning with $U(p_{i-1}) + 1$.

5. For the observed (g,s), we set confidence limits p_l and p_u as follows:

$$p_l = \min_{p_i} U(p_i) \geq (g,s) \qquad p_u = \max_{p_i} L(p_i) \leq (g,s)$$

Confidence Interval

Numerical example of Duffy and Santner's method

Reiczigel, J. Stat Med 2003.

$(p_0, p_1) = (0.1, 0.35)$, nominal one-sided α level = 0.025, nominal power = 0.8, $r = 9$.

	p_i						
(m, s)	0	0.1	0.2	0.3	⋯	0.8	0.9	1.0
(17,0)	1	0.17	0.02	0.002	⋯	1.31×10^{-12}	1.00×10^{-17}	0
(18,1)	0	0.28	0.08	0.019	⋯	1.78×10^{-11}	1.53×10^{-16}	0
(19,2)	0	0.26	0.14	0.032	⋯	1.28×10^{-10}	1.24×10^{-15}	0
⋮	⋮	⋮	⋮	⋮	⋯	⋮	⋮	⋮
(15,6)	0	7.76×10^{-4}	1.72×10^{-2}	5.89×10^{-2}	⋯	2.69×10^{-4}	1.06×10^{-6}	0
⋮	⋮	⋮	⋮	⋮	⋯	⋮	⋮	⋮
(11,6)	0	1.49×10^{-4}	5.28×10^{-3}	0.04	⋯	0.021	1.33×10^{-3}	0
⋮	⋮	⋮	⋮	⋮	⋯	⋮	⋮	⋮
(6,6)	0	1.00×10^{-6}	6.40×10^{-5}	7.30×10^{-4}	⋯	0.26	0.53	1

p_i	0.1	0.2	0.3	⋯	0.8	0.9
$L(p_i)$	(17,0)	(17,0)	(17,0)	⋯	(11,6)	(9,6)
$U(p_i)$	(17,0)	(15,6)	(11,6)	⋯	(6,6)	(6,6)

1. Calculate negative binomial probability.
2. Sort in ascending order these probability in p_i, and construct the region in the minimum cardinality.
3. For observed (m, s), set confidence interval as minimum and maximum p_i which (m, s) in these regions.

If we observed $(m, s) = (11,6)$, the 95% confidence interval is [0.3,0.8].

Comparison among different designs

Maximum sample size(MSS)

Maximum cost in a trial

e.g. α-spending function method with OF-Wald design ($p_0 = 0.1$, $p_1 = 0.35$)

Test	Sample size	Spending α error rate
29	29	0.0021
30	30	0.0022
31	31	0.0023
	MSS	**Overall α error rate**
	31	0.025

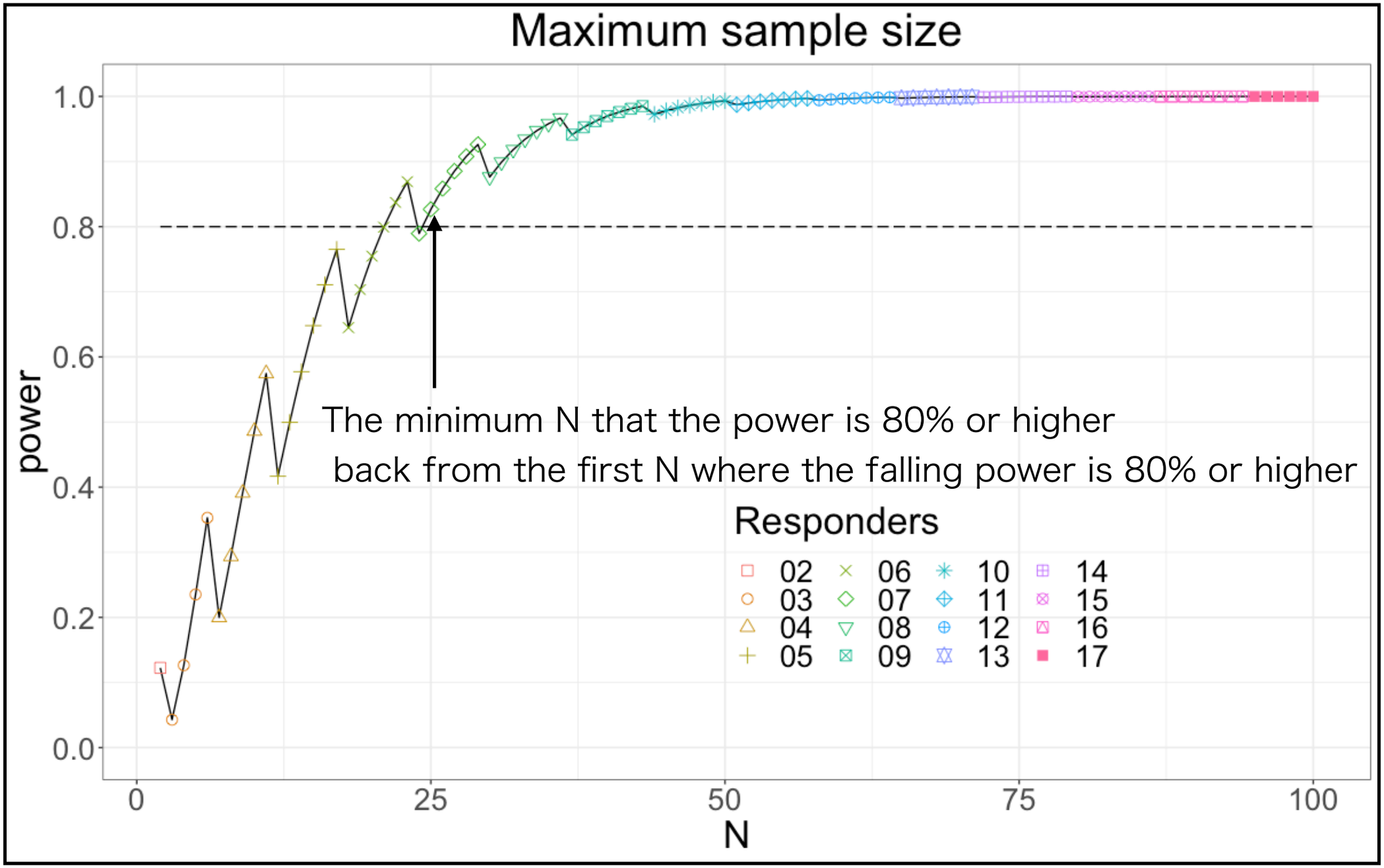
Average sample number(ASN)

Average cost by repeating the same trial

e.g. α-spending function method with OF-Wald design ($p_0 = 0.1$, $p_1 = 0.35$, $p = 0.55$)

Trial No.	Sample size	Termination
1	22	1
2	31	0
3	10	1
	ASN	**Power**
	21	0.666…

Maximum sample size in fixed design



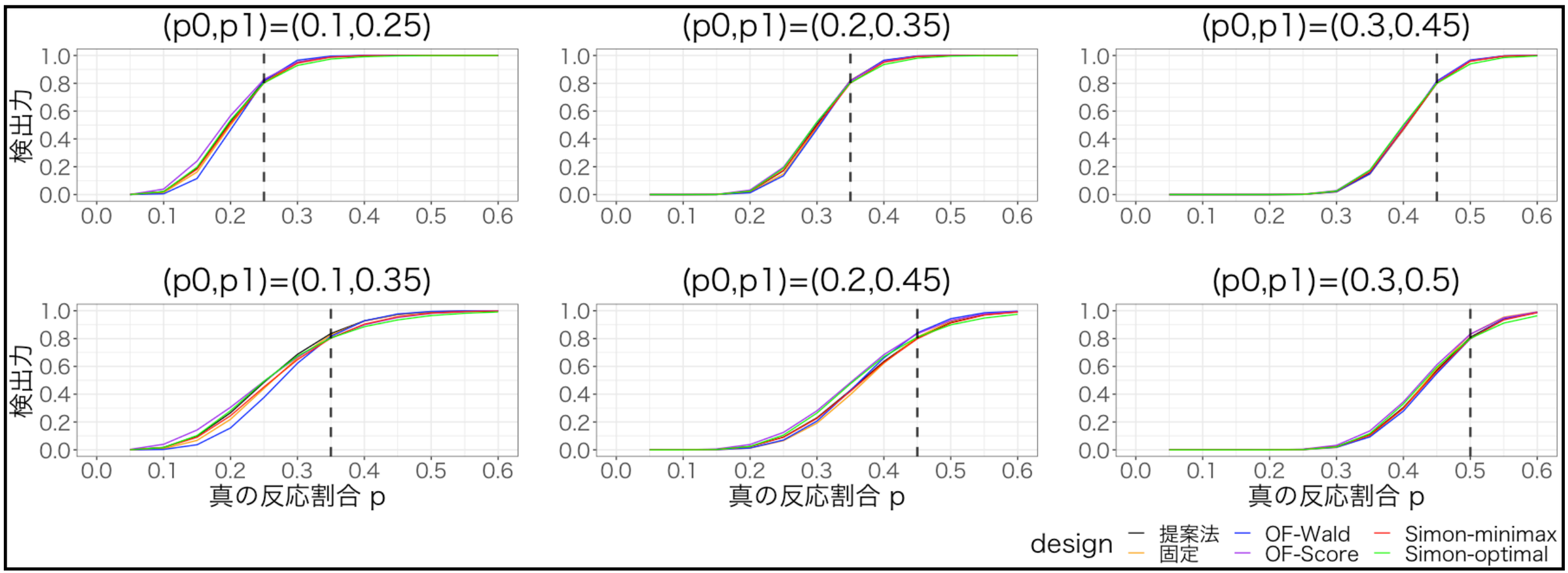
As increasing responders, power is decreasing

Maximum sample size is 25 in $(p_0, p_1) = (0.1, 0.35)$

Simulation Result: Maximum sample size

p_0	0.1						0.2				0.3	
p_1	0.25	0.3	0.35	0.4	0.45	0.5	0.35	0.4	0.45	0.5	0.45	0.5
Proposed	49	29	22	16	11	10	72	41	26	19	83	47
Fixed	53	33	25	19	14	10	78	44	31	24	88	54
OF												
Wald	70	45	31	23	17	14	85	51	34	24	92	52
Score	44	26	17	13	10	8	67	38	26	18	82	47
Simon												
Minimax	49	29	22	16	11	10	69	41	26	19	81	47
Optimal	58	38	30	18	12	11	83	55	35	23	100	65

Simulation Results: Power



Point

Actual power is 0.8 or over in all designs.

Power requirement is achieved for all designs.

References

Agresti, A. (2018). *An introduction to categorical data analysis* (3rd Edition). Hoboken, NJ: John Wiley & Sons.

Anderson, K. (2021). gsDesign: Group Sequential Design. R package version 3.2.1. https://CRAN.R-project.org/package=gsDesign

A'hern, R. P. (2001). Sample size tables for exact single-stage phase II designs. *Stat Med* **20**, 859-866.

Chang, M. N., Therneau, T. M., Wieand, H. S. and Cha, S. S. (1987). Designs for group sequential phase II clinical trials. *Biometrics* **43**, 865–874.

Chow, S. C., Shao, J., Wang, H. and Lokhnygina, Y. (2017). *Sample size calculations in clinical research*. Boca Raton, FL: Chapman and Hall/CRC.

Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, *26*(4), 404-413.

Conaway, M. R. and Petroni, G. R. (1995). Bivariate sequential designs for phase II trials. *Biometrics* **51**, 656–664.

References

Duffy, D. E., & Santner, T. J. (1987). Confidence intervals for a binomial parameter based on multistage tests. *Biometrics*, 81-93.

Fleming, T. R. (1982). One-sample multiple testing procedure for phase II clinical trials. *Biometrics* **38**, 143-151.

Green, S., Benedetti, J., Smith, A. and Crowley, J. (2012). *Clinical Trials in Oncology* (2nd Edition). Boca Raton, FL: CRC Press.

Guo, H. Y., & Liu, A. (2005). A simple and efficient bias-reduced estimator of response probability following a group sequential phase II trial. *J Biopharm Stat*, **15**(5), 773-781.

Hirji, K. F., Tsiatis, A. A. and Mehta, C. R. (1989). Median unbiased estimation for binary data. *Am Stat* **43**, 7–11.

Ivanova, A., Paul, B., Marchenko, O., Song, G., Patel, N., & Moschos, S. J. (2016). Nine-year change in statistical design, profile, and success rates of phase II oncology trials. *J Biopharm Stat* **26**(1), 141-149.

References

Jennison, C., & Turnbull, B. W. (1983). Confidence intervals for a binomial parameter following a multistage test with application to MIL-STD 105D and medical trials. *Technometrics*, **25**(1), 49-58.

Jovic, G., & Whitehead, J. (2010). An exact method for analysis following a two-stage phase II cancer clinical trial. *Stat Med*, **29**(30), 3118-3125.

Jung, S. H., & Kim, K. M. (2004). On the estimation of the binomial probability in multistage clinical trials. *Statistics in medicine*, **23**(6), 881-896.

Jung, S. H. (2013). *Randomized phase II cancer clinical trials.* Boca Raton, FL: CRC Press.

Kunz, C. U. and Kieser, M. (2012). Curtailment in single-arm two-stage phase II oncology trials. *Biom J* **54,** 445–456.

Lachin, J. M. (1981). Introduction to sample size determination and power analysis for clinical trials. *Control Clin Trials* **2**, 93–113.

Lan, K. K. G. and DeMets, D. L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika* **70**, 659–663.

References

O'Brien, P. C., and Fleming, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics* **35**(7) 549-556.

Porcher, R., & Desseaux, K. (2012). What inference for two-stage phase II trials?. *BMC medical research methodology*, **12**, 1-13.

Reiczigel, J. (2003). Confidence intervals for the binomial parameter: some new considerations. *Stat Med*, **22**(4), 611-621.

Robertson, D. S., Choodari-Oskooei, B., Dimairo, M., Flight, L., Pallmann, P., & Jaki, T. (2023). Point estimation for adaptive trial designs I: A methodological review. *Stat Med*, **42**(2), 122-145.

Seshan, V. E. (2018). clinfun: Clinical Trial Design and Data Analysis Functions. R package version 1.0.15. https://CRAN.R-project.org/package=clinfun

Simon, R. (1989). Optimal two-stage designs for phase II clinical trials. *Control clin trials*, **10**(1) 1-10.

Sterne, T. E. (1954). Some remarks on confidence or fiducial limits. *Biometrika*, **41**(1/2), 275-278.

References

Yin, G. (2012). *Clinical trial design: Bayesian and frequentist adaptive methods.* (Vol. 876). John Wiley & Sons.

Yin, H. (2020). Exact tests using binary data in adaptive two or multi-stage design. *Stat Methods Med Res,* **29**(8), 2269-2281.