Sample size determination in clinical trials with multiple co-primary endpoints

Takashi Sozu, Ph.D.

Kyoto University School of Public Health

Tomoyuki Sugimoto, Ph.D.

Hirosaki University Graduate School of Science and Technology

Toshimitsu Hamasaki, Ph.D.

Osaka University Graduate School of Medicine

December 16, 2011

Joint Meeting of the 2011 Taipei International Statistical Symposium
and 7th Conference of the Asian Regional Section of the IASC

Outline

- Background and objectives
- Statistical settings
- Derivation of power formula
- Behavior of sample sizes
- Conclusion

Background

- Clinical trials often employ two or more primary endpoints.
  - A major concern is whether or not clinical trials should achieve statistical significance on all of the multiple primary endpoints (i.e., co-primary endpoints).
- Statistical Principles for Clinical Trials ICH (1998)
- Multiple Endpoint Expert Team (PhRMA) listed 20 diseases where regulatory agencies have required co-primary endpoints.

Offen et al. (2007)

Common solutions for the multiplicity

- Composite variable ICH (1998)
  - ○ This approach addresses the multiplicity problem without requiring adjustment to multiplicity.
  - ✕ A clinically meaningful and validated variable is not always available.
  - ✕ Interpretation of the variable is not easy.

- Assuming the independency among endpoints and increasing the power for each endpoint
  - ○ The power is simply defined. Eaton and Muirhead (2006)
  - ✕ The power would be under-evaluated. (The sample size is over-sized.)

Existing approaches for power and sample size determination

Continuous (Normal)

Xiong et al. (2005)
Sozu et al. (2006, 2011)
Eaton, Muirhead (2006)
Sugimoto et al. (2011)

IBC 2010
(2010.12.6)
Sozu et al.
(submitted)

Binary

Song (2009)
Sozu et al. (2010, 2011)
Hamasaki et al.
(submitted)

MCP 2011 (2011.8.31)

MCP 2011 (2011.8.31)
Hamasaki et al.
(submitted)

Ordinal

Time-to-event

Objective

We discuss the power and sample size determination for superiority comparative clinical trials with multiple co-primary endpoints (for achieving statistical significance for all of the endpoints).

Continuous (Normal)

Cont. and Binary

Binary

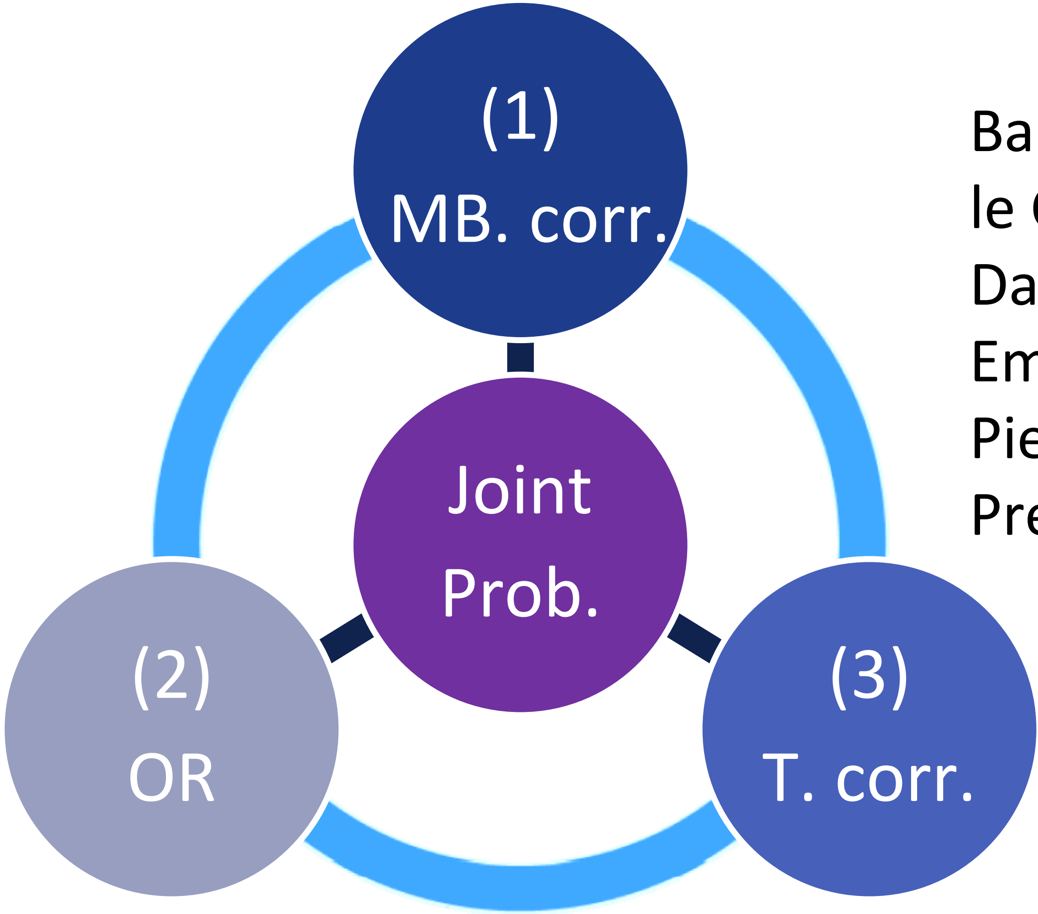
Common steps of our research

- Define the response variables and association measures (correlations) among them
- Calculate the correlation coefficients among the test statistics to derive a power formula
- Evaluate the behaviors of sample sizes

Association measures among endpoints

Scale	Distribution	Association measure(s)
Cont. (Normal)	Multivariate Normal (MN)	Correlation coefficient (r)
Binary	Multivariate Bernoulli (MB)	(1) r of MB dist. (MB. corr.) (2) Odd ratio (OR) (3) r of a latent MN dist. (Tetrachoric corr.)
Cont. and Binary	MN (latent distribution)	r of a latent MN dist. (Biserial corr.)

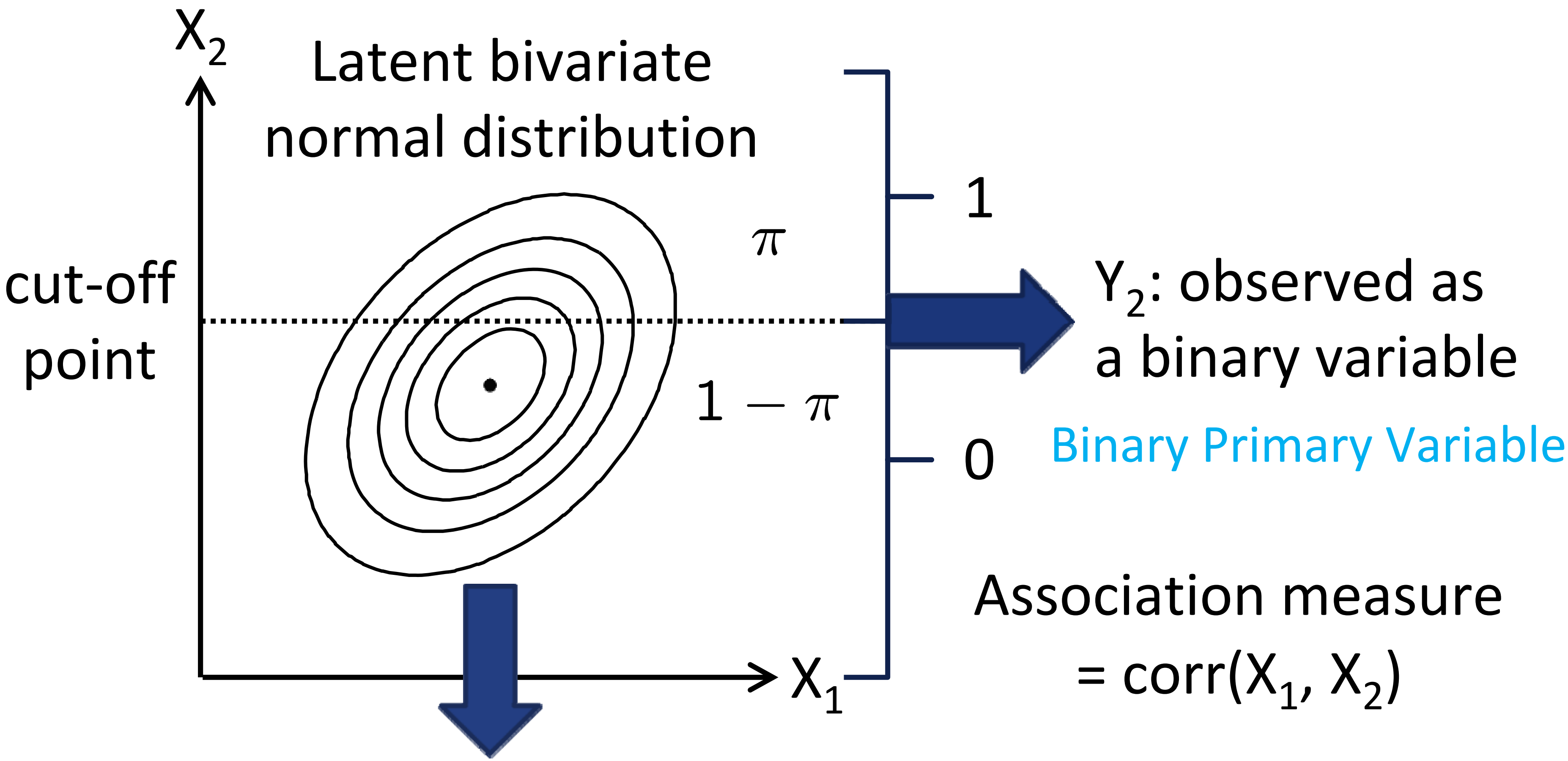
Relationships among association measures for binary variables



Bahadur (1961)
le Cessie (1994)
Dale (1986)
Emrich, Piedmonte (1991)
Prentice (1988)

A joint probability can be estimated
if the individual data of endpoints are available.

Biserial model for mixed cont. and binary variables



Y_1: observed as a normal variable
Normal Primary Variable

Pearson (1909)
Tate (1955a, 1955b)

Hypothesis testing

- $\delta_k = \begin{cases} \text{Difference in means for Normal Primary Variable} \\ \text{Difference in success probabilities for Binary Primary Variable} \end{cases}$

  $(k = 1, \ldots, K)$

  - $\delta_k > 0$ means an improvement

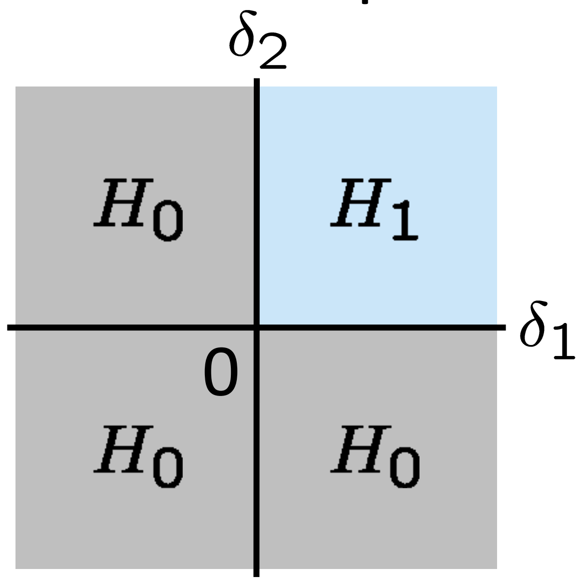
- $H_0 : \delta_k \leq 0$ for at lest one $\boldsymbol{k}$

  $(\iff \max \delta_k \leq 0)$

- $H_1 : \delta_k > 0$ for all $\boldsymbol{k}$

  $(\iff \min \delta_k > 0)$

Parameter space for two endpoints



Testing methods

- Normal Primary Variable

(1) Z-test: Known variance — Xiong et al. (2005)

(2) T-test: Unknown variance — Sozu et al. (2006, 2011)

- Binary Primary Variable

(1)(2) Chi-square test without/with cc

(3)(4) Arcsine transformation without/with cc

(5) Fisher's exact test (cc: continuity correction)

(6) Test based on log-transformed relative risks

(1-5) Sozu et al. (2010) (6) Hamasaki et al. (submitted)

Power formula:
Asymptotic normal test (except for T-test)

- Overall power = $1 - \beta = \Pr\left(\bigcap_{k=1}^{K} \{Z_k > z_\alpha\}\right)$
  - Z_k : Test statistic
  - z_α : Critical value (α : Significance level)
- Transform Z_k into the standardized statistics Z_k^* .
  - $(Z_1^*, \ldots, Z_K^*)^\top \sim N_K(\mathbf{0}, \mathbf{\Sigma})$
  - $\mathbf{\Sigma}$: diagonal = 1, off diagonal is given by a function of the association measure of the response variables.
- The power can be calculated from the CDF of $N_K(\mathbf{0}, \mathbf{\Sigma})$.

Sozu et al. (2010, 2011)

Power formula: T-test

- Overall power = $1 - \beta = \Pr\left(\bigcap_{k=1}^{K} \{T_k > t_\alpha\}\right)$
  - T_k : Test statistic
  - t_α : Critical value
- Calculate the power using a Monte Carlo integration
  - Generate Wishart random numbers for variance-covariance matrix
  - Calculate the conditional power
  - Repeat the above steps and calculate the mean of the conditional power

Sozu et al. (2006, 2011)

Power formula: Fisher's exact test

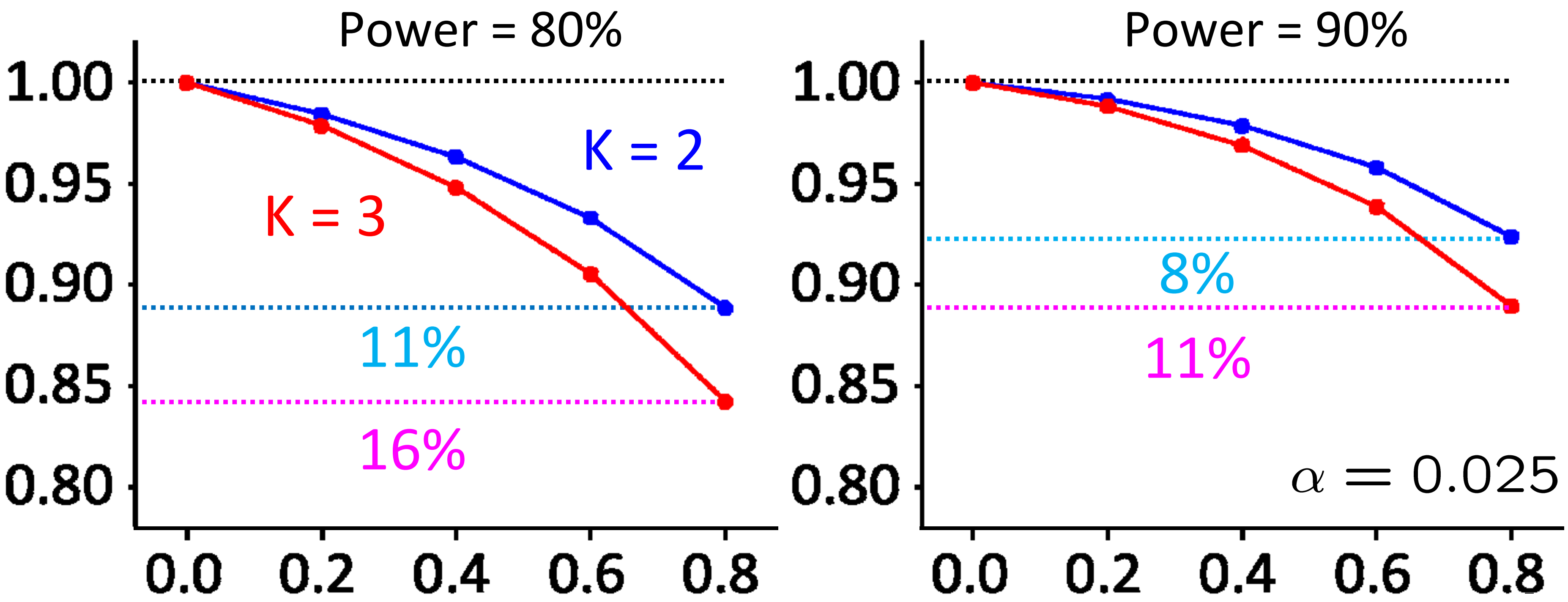
- Overall power = $1 - \beta = \Pr\left(\bigcap_{k=1}^{K} \{P_k < \alpha\}\right)$
  - P_k : one-sided p-value

- Calculate the power using a Monte Carlo integration.
  - Generate random numbers for the response variables from the assumed distribution

Sozu et al. (2010, 2011)

Behavior of sample sizes for $\delta_1 = \delta_2 (= \delta_3)$ Normal Primary Variables: Z-test

$$\text{Decrease in sample size} = \frac{\text{sample size at each r}}{\text{sample size at r = 0.0}}$$



r : correlation coefficient (compound symmetry)

Behavior of sample sizes for Normal Primary Variables

Sozu et al. (2006, 2011)

- The sample size ↘ the correlation ↗
  - when the effect sizes are approximately equal among endpoints
- The decrease in SS is determined by the following four design parameters:

  (1) Significance level of alpha (α)

  (2) Target power $(1-\beta)$

  (3) Effect size ratios of endpoints

  (4) Correlations among endpoints

  Sugimoto et al. (2011)
- The sample size (SS) based on T-test (unknown variance) + 1 $\simeq$ the SS based on Z-test (known variance).

Behavior of sample sizes for other Primary Variables

- **Binary Primary Variables:** Sozu et al. (2010, 2011)
  - The features of the testing methods are similar to the case of single endpoint (K=1).
    - e.g. The Arcsine transformation with CC provides sample sizes approximately equal to those obtained by Fisher's exact test.
- **Mixed Normal and Binary Variables:**
  - The decrease in the sample size is relatively small as compared to the case of Normal Primary Variable.

Convenient formula for sample size

- Sample size formula for Normal Primary Variables

$$n = \frac{(C_K + z_\alpha)^2}{\kappa_p \cdot \min(\delta_k)^2} \xrightarrow[\text{endpoints (K=1)}]{\text{Single}} \frac{(z_\beta + z_\alpha)^2}{\kappa_p \cdot \delta_1^2}$$

  - C_K is the function of the four design parameters
    - A Value can be obtained from the numerical table.
  - $\kappa_p = p/(1+p),\ \ p = n_C/n_T$
    - n_T, n_C : The number of subjects of each group
- A required sample can be calculated without using a statistical software.

Sugimoto et al. (2011)

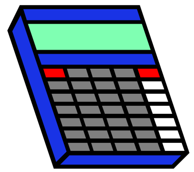
An numerical example: K = 2

- $\gamma_1 = \delta_1/\delta_2 = 0.5/0.4 = 1.25$
- $\rho_{12} = 0.5$ (Correlation between two endpoints)
- $\alpha = 0.025$
- $\beta = 0.2$

Numerical table → $C_2 = 0.925$

$$n = \frac{(C_2 + z_\alpha)^2}{\kappa_p \cdot \delta_2^2} = \frac{(0.925 + 1.96)^2}{(1/2) \cdot 0.4^2} = 104.04 \to 105$$

$$n_T = n_C \to p = 1 \to \kappa_p = p/(1+p) = 1/2$$

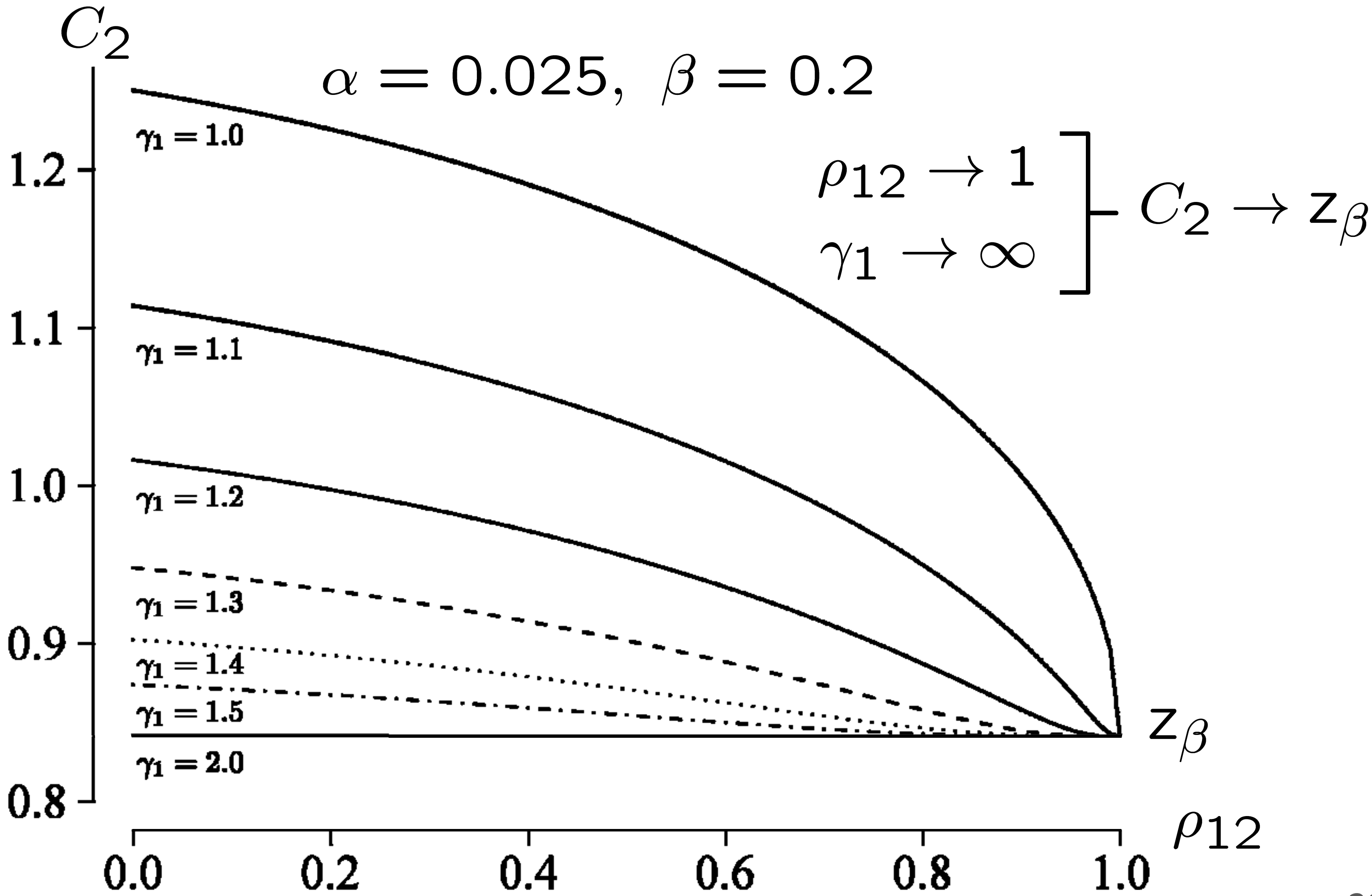


Example of numerical table for C_2

$$K = 2,\ \alpha = 0.025,\ \beta = 0.2$$

γ_1	ρ_{12} 0.0	0.2	0.3	0.5	0.7	0.8	0.95
1.00	1.250	1.226	1.210	1.168	1.109	1.066	0.961
1.02	1.219	1.195	1.179	1.138	1.079	1.038	0.934
⋮				⋮			
1.25	0.979	0.962	0.952	0.925	0.890	0.870	0.843
⋮				⋮			
2.00	0.842	0.842	0.842	0.842	0.842	0.842	0.842

Curve of C_2



C_2
$\alpha = 0.025,\ \beta = 0.2$
$\rho_{12} \to 1$
$\gamma_1 \to \infty$
$C_2 \to z_\beta$
$\gamma_1 = 1.0$
$\gamma_1 = 1.1$
$\gamma_1 = 1.2$
$\gamma_1 = 1.3$
$\gamma_1 = 1.4$
$\gamma_1 = 1.5$
$\gamma_1 = 2.0$
z_β
ρ_{12}
1.2
1.1
1.0
0.9
0.8
0.0
0.2
0.4
0.6
0.8
1.0

Conclusions

- We introduced the method of power and sample size calculations for multiple co-primary endpoints for achieving statistical significance for all of the endpoints.
- It is important to consider associations among endpoints into sample size calculation when
  - the endpoints are (positively) correlated and
  - the effect sizes (i.e., the corresponding individual powers) are approximately equal among the endpoints.

Thank you very much for your attention

Search IBC2012

XXVI[th] International Biometric Conference

Organised by the Biometric Society of Japan,
Japanese Region of the International Biometric Society

This research is partially supported by the Ministry of Education, Culture, Sports, Science and Technology, Grant-in-Aid for Young Scientists (B), No. 22700290.

References 1

- Bahadur RR. A representation of the joint distribution of responses to *n* dichotomous items. In *Studies in Item Analysis and Prediction, Vol. VI, Stanford Mathematical Studies in the Social Sciences*, Solomon H (ed.). Stanford University Press: Stanford, CA, 1961; 158-168.
- Bartlett MS. The use of transformations. *Biometrics* 1947; **3**: 39-52.
- Bloch DA, Lai TL, Su Z, Tubert-Bitter P. A combined superiority and non-inferiority approach to multiple endpoints in clinical trials. *Statistics in Medicine* 2007; **26**: 1193-1207.
- Breedveld FC, Weisman MH, Kavanaugh AF, Cohen SB, Pavelka K, van Vollenhoven R, Sharp J, Perez JL, Spencer-Green GT. The PREMIER study: A multicenter, randomized, double-blind clinical trial of combination therapy with adalimumab plus methotrexate versus methotrexate alone or adalimumab alone in patients with early, aggressive rheumatoid arthritis who had not had previous methotrexate treatment. *Arthritis & Rheumatism* 2006; **51**, 26-37.
- Capizzi T, Zhang J. Testing the hypothesis that matters for multiple primary endpoints. *Drug Information Journal* 1996; **30**: 949-956.
- Casagrande JT, Pike MC, Smith PG. An improved approximate formula for calculating sample sizes for comparing two binomial distributions. *Biometrics* 1978; **34**:483-486.
- Chow SC, Shao J, Wang H. *Sample Size Calculations in Clinical Research* (2nd~edn). Chapman & Hall/CRC, Boca Raton, FL, 2007.

References 2

- Chuang-Stein C, Stryszak P, Dmitrienko A, Offen W. Challenge of multiple co-primary endpoints: A new approach. *Statistics in Medicine* 2007; **26**:1181-1192.
- Committee for Medicinal Products for Human Use (CHMP). *Guideline on Medicinal Products for the Treatment of Alzheimer's Disease and other Dementias* (CPMP/EWP/553/95). EMEA: London, 2008.
- Committee for Proprietary Medicinal Products (CPMP). *Points to Consider on Multiplicity Issues in Clinical Trials* (CPMP/EWP/908/99). EMEA: London, 2002.
- le Cessie S, van Houwelingen JC. Logistic regression for correlated binary data. *Applied Statistics* 1994; **43**: 95-108.
- Chow SC, Shao J, Wang H. *Sample Size Calculations in Clinical Research* (2nd edn). Chapman and Hall/CRC: Boca Raton, FL, 2007.
- Dale JR. Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics* 1986; **42**: 909-917.
- Eaton ML, Muirhead RJ. On a multiple endpoints testing problem. *Journal of Statistical Planning and Inference* 2007; **137**: 3416-3429.
- Emrich LJ, Piedmonte MR. A method for generating high-dimensional multivariate binary variates. *The American Statistician* 1991; **45**: 302-304.

References 3

- Food and Drug Administration. *Daft Guidance for Industry. Irritable Bowel Syndrome*: Clinical Evaluation of Products for Treatment. Center for Drug Evaluation and Research, Food and Drug Administration, Rockville, MD, March 2010.
- Genz A. Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics* 1992; **1**: 141-150.
- Ho TW, Ferrari MD, Dodick DW, Galet V, Kost J, Fan X, Leibensperger H, Froman S, Assaid C, Lines C, Koppen H, Winner PK. Efficiency and tolerability of MK-0974 (telcagepant) a new oral antagonist of calcitonin gene-related peptide receptor, compared with Zolmitriptan for acute migraine: a randomised, placebo-controlled, parallel-treatment trial. *The Lancet* 2008; **37**: 2115-2123.
- Hung HMJ, Wang SJ. Some controversial multiple testing problems in regulatory applications. *Journal of Biopharmaceutical Statistics* 2009; **19**: 1-11.
- *International Conference on Harmonization (ICH) of Technical Requirements for Regulations of Pharmaceuticals for Human use*. ICH Tripartite Guideline E-9 Documents, Statistical Principles for Clinical Trials, 5 February 1998.
- Johnson ME. Multivariate Statistical Simulation. New York: John Wiley & Sons, 1987.

References 4

- Johnson NL, Kotz S. Distributions in Statistics: Continuous Multivariate Distributions. New York: John Wiley & Sons, 1972.
- Julious SA. Designing clinical trials with uncertain estimates of variability. *Pharmaceutical Statistics* 2004; 3: 261-268.
- Julious SA. Sample Sizes for Clinical Trials. Boca Raton, FL: Chapman & Hall, 2009.
- Kordzakhia G, Siddiqui O, Huque MF. Method of balanced adjustment in testing co-primary endpoints. *Statistics in Medicine* 2010; **29**: 2055-2066.
- Lev J. The point biserial coefficient of correlation. *The Annals of Mathematical Statistics* 1949; 20: 125-126.
- Machin D, Campbell M, Tan SB, Tan SH. Sample Size Tables for Clinical Studies. 3rd ed. Chichester, UK: Wily-Blackwell, 2009.
- Miwa A, Hayter J, Kuriki S. The evaluation of general non-centred orthant probabilities. *Journal of the Royal Statistical Society, Series B* 2003; **65**: 223-234.
- Molenberghs G, Geys H, Buyse M. Evaluation of surrogate endpoints in randomized experiments with mixed discrete and continuous outcomes. *Statistics in Medicine* 2001; **20**, 3023-3038.

References 5

- 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. Multiple co-primary endpoints: medical and statistical solutions. *Drug Information Journal* 2007; **41**: 31--46.
- Patel HI. Comparison of treatments in a combination therapy trial. *Journal of Biopharmaceutical Statistics* 1991; **1**: 171-183.
- Pearson K. Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society*, *Series A* 1900; **195**: 1-47.
- Pearson K. On a new method for determining the correlation between a measured character A and a character B. *Biometrika* 1909; **7**: 96-105.
- Peskind ER, Potkin SG, Pomara N, Ott BR, Graham SM, Olin JT, McDonald S. Memantine treatment in mild to moderate Alzheimer disease: A 24-week randomized, controlled trial. *American Journal of Geriatric Psychiatry* 2006; **14**: 704–715.
- Phillips A, Ebbutt A, France L, Morgan D. The International conference on harmonization guideline ‘Statistical Principles for Clinical Trials’: Issue in applying the guideline in practice. *Drug Information Journal* 2000; **34**: 337-348.

References 6

- Plackett RL. A class of bivariate distributions. Journal of the American Statistical Association 1965; 60: 516-522.
- Pocock SJ. Clinical trials with multiple outcomes: a statistical perspective on their design, analysis and interpretation. Controlled Clinical Trials 1997; 18: 530-545.
- Pong A, Shein-Chung C. Statistical/practical issues in clinical trials. Drug Information Journal 1997; 31: 1167-1174.
- Prentice RL. Correlated binary regression with covariates specific to each binary observation. Biometrics 1988; 44: 1033-1048.
- Rögers SL, Farlow MR, Doody RS, Mohs R, Friedhoff LT, The Donepezil Study Group. A 24-week, double-blind, placebo-controlled trial of donepezil in patients with Alzheimer's disease. Neurology 1998; 50: 136-145.
- Rosler M, Anand R, Cicin-Sain A, Gauthier S, Agid Y, Dal-Bianco P, Stahelin HB, Hartman R, Gharabawi M. Efficacy and safety of rivastigmine in patients with Alzheimer's disease: International randomised controlled trial. British Medical Journal 1999; 318: 633-640.

References 7

- Rudnick MR, Davidson C, Laskey W, Stafford JL, Sherwin PF. VALOR Trial Investigators. Nephrotoxicity of iodixanol versus ioversol in patients with chronic kidney disease: The Visipaque Angiography/Interventions with Laboratory Outcomes in Renal Insufficiency (VALOR) Trial. American Heart Journal 2008; 156: 776-782.
- Sahai H, Khurshid A. Formulae and tables for the determination of sample sizes and power in clinical trials for testing differences in proportions for the two-sample design: a review. Statistics in Medicine 1996; 15: 1-21.
- Sankoh AJ, D'Agostino RB, Huque MF. Efficacy endpoint selection and multiplicity adjustment method in clinical trials with inherent multiple endpoint issues. Statistics in Medicine 2003; 22: 3133-3150.
- SAS Institute, Inc., SAS Online Doc 9.1.3. SAS Institute, Inc.: Cary, NC, 2006.
- Song JM. Sample size for simultaneous testing of rate differences in non-inferiority trials with multiple endpoints. Computational Statistics & Data Analysis 2009; 53: 1201-1207.
- Sozu T, Kanou T, Hamada C, Yoshimura I. Power and sample size calculations in clinical trials with multiple primary variables. Japanese Journal of Biometrics 2006; 27: 83-96.

References 8

- Sozu T, Sugimoto T, Hamasaki T. Sample size determination in clinical trials with multiple co-primary binary endpoints. Statistics in Medicine 2010; 29: 2169-2179.
- Sugimoto T, Sozu T, Hamasaki T. A convenient formula for sample size calculations in clinical trials with multiple co-primary continuous endpoints. Pharmaceutical Statistics, to appear.
- Tariot PN, Solomon PR, Morris JC, Kershaw P, Lilienfeld S, Ding C, The Galantamine USA Study Group. A 5-month, randomized, placebo-controlled trial of galantamine in AD. Neurology 2000; 54: 2269-2276.
- Tate RF. Correlation between a discrete and a continuous variable. Point-biserial correlation. The Annals of Mathematical Statistics 1954; 25: 603-607.
- Tate RF. The theory of correlation between two continuous variable when one is dichotomized. Biometrika 1955a. 42: 205-216.
- Tate RF. Applications of correlation models for biserial data. Journal of the American Statistical Association 1955b. 50: 1078-1095.
- Ury HK. Continuity-corrected approximations to sample size or power when comparing two proportions: Chi-squared or arc sine? The Statistician 1981; 30: 199-203.
- Walters DE. In defence of the arc sine approximation. The Statistician 1979; 28: 219-222.

References 9

- Winblad B, Kilander L, Eriksson S, Minthon L, Batsman S, Wetterholmf AL, Jansson-Blixt C, Haglund A. Donepezil in patients with severe Alzheimer's disease: double-blind, parallel-group, placebo-controlled study. The Lancet 2006; 367: 1057-1065.
- Yates F. Contingency tables involving small numbers and the 2 test. Journal of the Royal Statistical Society, Supplement 1934; 1: 217-235.
- Xiong C, Yu K, Gao F, Yan Y, Zhang Z. Power and sample size for clinical trials when efficacy is required in multiple endpoints: application to an Alzheimer’s treatment trial. Clinical Trials 2005; 2: 387-393.

Backup

Why multiple endpoints are required?

(1) Lack of a consensus on a single most important variable from the medical perspective

(2) No clear aetiology of diseases

(3) A disease condition is characterized in multi-dimensional ways

Pocock (1997)

Pong and Shein-Chung (1997)

Sankoh et al. (2003)

Chuang-Stein et al. (2007)

Statistical Principles for Clinical Trials

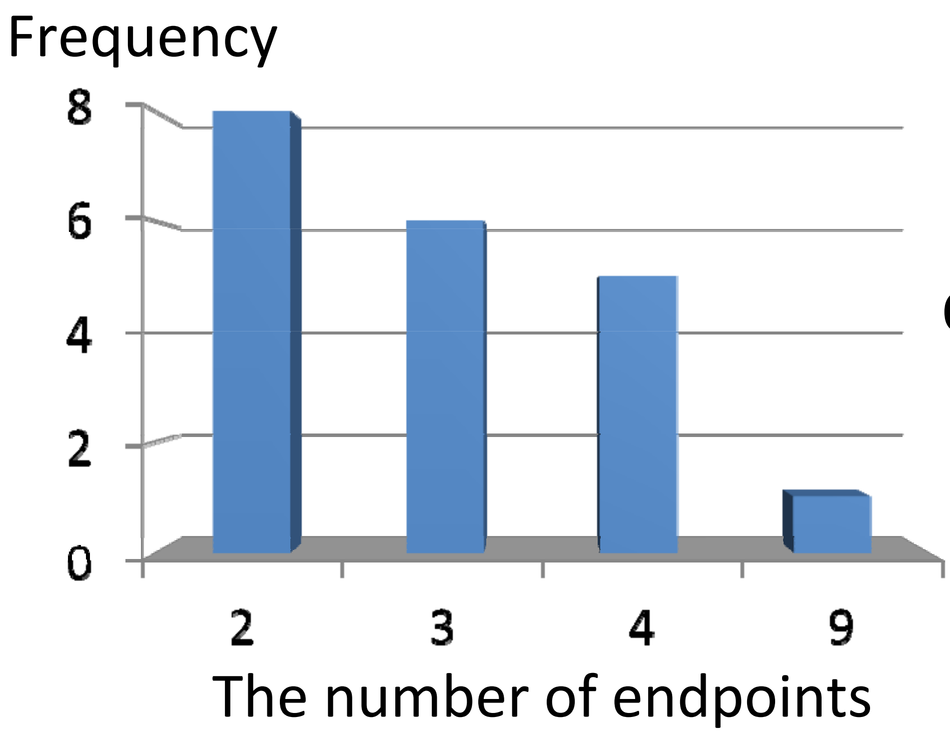
- 2.2.2 Primary and Secondary Variables
  - There should generally be only one primary variable.

- 2.2.5 Multiple Primary Variables
  - It may sometimes be desirable to use more than one primary variable, each of which (or a subset of which) could be sufficient to cover the range of effects of the therapies.

ICH (1998)

Examples of co-primary endpoints

- Multiple Endpoint Expert Team (PhRMA) listed 20 diseases where regulatory agencies have required co-primary endpoints.



Offen et al. (2007)

Example of a clinical trial

- A randomized, parallel-treatment, placebo-controlled, double-blind trial
- Acute migraine
- Five co-primary endpoints are considered.

(1) pain freedom

(2) pain relief

(3) phonophobia

(4) photophobia

(5) nausea

Ho et al. (2008)

Example of guidance

- Irritable Bowel Syndrome
- Primary endpoint

FDA (2010)

(1) Abdominal pain (11-points: 0 to 10)

(2) IBS-C (Constipation): stool frequency

(2) IBS-D (Diarrhea): stool consistency

- The Bristol stool Form Scale (seven levels)

http://www.e-chiken.com/shikkan/kabinsei.htm

Rösler et al. (1999; BMJ)

Statistical methods

The study sample population of about 200 in each group was planned to enable achievement of 90% power with $\alpha = 0.05$ for detecting at least a 3.0 point improvement on the Alzheimer's disease assessment scale and an increase from 15-30% among patients scoring < 4 on the clinician impression of change scale.

Three possible clinical scenarios

- Showing significance for

Sankoh et al. (2003)

(1) all primary variables

- co-primary endpoints
  - reverse multiplicity problem

Offen et al. (2007)

(2) majority of the primary variables

(The constitution of the majority is defined in the protocol)

(3) one or more of the primary variables

- alternative primary endpoints

Offen et al. (2007)

Feature of three association measures

	Pros	Cons
B. corr	▪ Parameter of the assumed distribution	▪ Restricted range
OR	▪ No restricted range from 0 to infinity ▪ A direct extension to a global cross-ratio	▪ A scale depends strongly on response probabilities
L. corr	▪ No restricted range from -1 to 1	▪ Iterative calculations are necessary to specify the value from a value of other measure

Assumptions for response variables

- Biserial model
  - NPV (X) is observed as X (X = X)
  - BPV (Y) is obtained by a dichotomized of X (X -> Y)

Pearson (1903)

- Point-biserial model
  - NPV (X) is distributed as a mixed normal distribution
  - BPV (Y) is distributed as a Bernoulli distribution

Standardized test statistics

$$1-\beta = \mathrm{P}\left(\bigcap_{k=1}^{K}\{Z_k > z_\alpha\}\right) \simeq \mathrm{P}\left(\bigcap_{k=1}^{K}\{Z_k^* > c_k^*\}\right)$$

$$Z_k^* = \begin{cases} \dfrac{\bar{Y}_{\mathrm{T}k} - \bar{Y}_{\mathrm{C}k} - \delta_k}{\sigma_k\sqrt{\dfrac{1+\kappa}{\kappa n}}}, & k \le k_m \\ \dfrac{p_{\mathrm{T}k} - p_{\mathrm{C}k} - \delta_k}{\sqrt{\dfrac{\kappa\pi_{\mathrm{T}k}\theta_{\mathrm{T}k} + \pi_{\mathrm{C}k}\theta_{\mathrm{C}k}}{\kappa n}}}, & k > k_m, \end{cases},$$

$$c_k^* = \begin{cases} z_\alpha - \dfrac{\delta_k}{\sigma_k}\sqrt{\dfrac{\kappa n}{1+\kappa}}, & k \le k_m \\ \dfrac{1}{\sqrt{\kappa\pi_{\mathrm{T}k}\theta_{\mathrm{T}k} + \pi_{\mathrm{C}k}\theta_{\mathrm{C}k}}}\left\{\sqrt{\dfrac{(\pi_{\mathrm{T}k} + \kappa\pi_{\mathrm{C}k})(\theta_{\mathrm{T}k} + \kappa\theta_{\mathrm{C}k})}{1+\kappa}}\, z_\alpha - \sqrt{\kappa n}\,\delta_k\right\}, & k > k_m. \end{cases}$$

Correlation between standardized test statistics

$$
\gamma^{kk'} = \begin{cases}
\text{two continuous variables } (k, k' \leq k_m): \\
\dfrac{\kappa \operatorname{Corr}(Y_{\mathrm{T}jk}, Y_{\mathrm{T}jk'}) + \operatorname{Corr}(Y_{\mathrm{C}jk}, Y_{\mathrm{C}jk'})}{1+\kappa}, \\
\text{continuous and binary variables } (k \leq k_m, \text{ and } k' > k_m): \\
\dfrac{\kappa \operatorname{Corr}(Y_{\mathrm{T}jk}, Y_{\mathrm{T}jk'})\sqrt{\pi_{\mathrm{T}k'}\theta_{\mathrm{T}k'}} + \operatorname{Corr}(Y_{\mathrm{C}jk}, Y_{\mathrm{C}jk'})\sqrt{\pi_{\mathrm{C}k'}\theta_{\mathrm{C}k'}}}{\sqrt{1+\kappa}\sqrt{\kappa\pi_{\mathrm{T}k'}\theta_{\mathrm{T}k'} + \pi_{\mathrm{C}k'}\theta_{\mathrm{C}k'}}}, \\
\text{two binary variables } (k, k' > k_m): \\
\dfrac{\kappa \operatorname{Corr}(Y_{\mathrm{T}jk}, Y_{\mathrm{T}jk'})\sqrt{\pi_{\mathrm{T}k}\theta_{\mathrm{T}k}\pi_{\mathrm{T}k'}\theta_{\mathrm{T}k'}} + \operatorname{Corr}(Y_{\mathrm{C}jk}, Y_{\mathrm{C}jk'})\sqrt{\pi_{\mathrm{C}k}\theta_{\mathrm{C}k}\pi_{\mathrm{C}k'}\theta_{\mathrm{C}k'}}}{\sqrt{\kappa\pi_{\mathrm{T}k}\theta_{\mathrm{T}k} + \pi_{\mathrm{C}k}\theta_{\mathrm{C}k}}\sqrt{\kappa\pi_{\mathrm{T}k'}\theta_{\mathrm{T}k'} + \pi_{\mathrm{C}k'}\theta_{\mathrm{C}k'}}}.
\end{cases}
$$

Sample size calculation

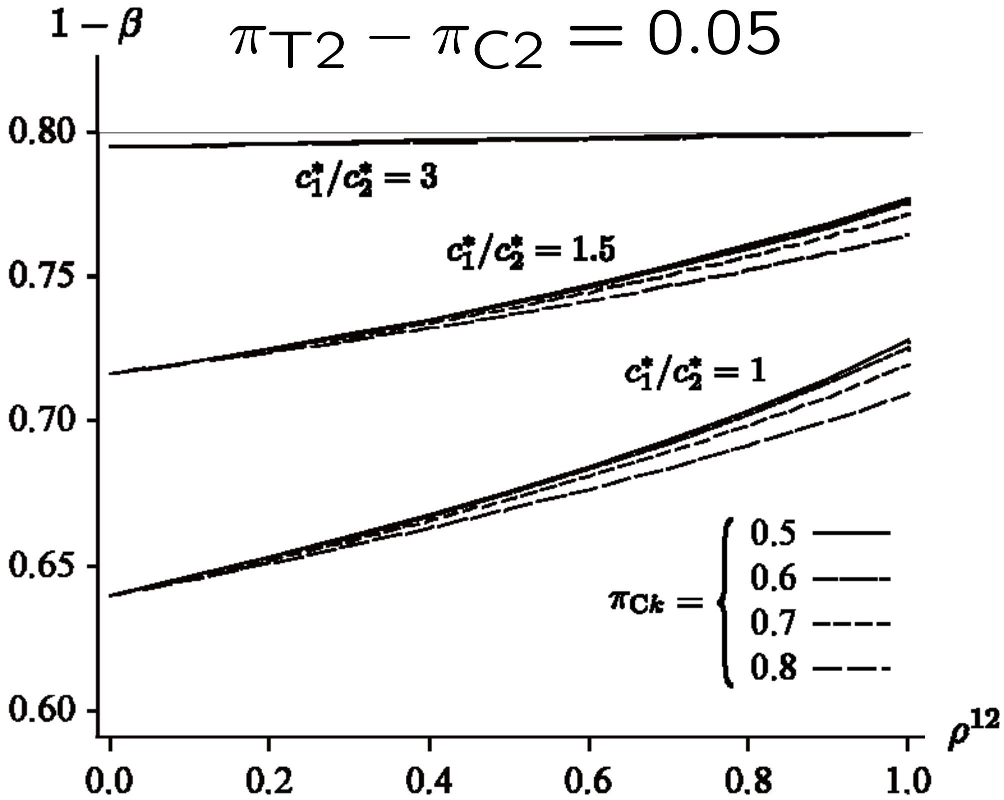
- Specify the value of parameters.
  - NPV: mean μ_{ik} and variance σ_k^2
  - BPV: success probability π_{ik}
- Specify the value of $\rho_{kk'} = \text{corr}(X_{ijk}, X_{ijk'})$.

  (Consider equal sample sizes: $n_1 = n_2 = n$)
- Choose a starting value of n and calculate the power.
- Repeat the above steps by gradually increasing n .
  - End the operation when the calculated power exceeds the desired $1 - \beta$, and select n as the minimum value of sample size.

Behavior of overall power (mixed)

one continuous and one binary variable (K=2)

$\pi_{T2} - \pi_{C2} = 0.05$



individual power = 0.8

Behaviors of sample sizes

c_1^*/c_2^*	δ_1^*	Proportions $(\pi_{T2}\ \pi_{C2})$	Correlation ρ^{12} 0.0	0.3	0.5	0.8	E_1	E_2
1	0.100	(0.55 0.50)	2055 (80.1)	2016 (79.9)	1980 (80.0)	1904 (80.0)	1565	1565
	0.103	(0.65 0.60)	1931 (80.0)	1895 (80.1)	1862 (80.1)	1793 (80.0)	1471	1471
	0.112	(0.75 0.70)	1643 (80.1)	1614 (80.1)	1588 (80.0)	1534 (80.0)	1251	1251
	0.132	(0.85 0.80)	1189 (80.0)	1171 (80.1)	1154 (80.1)	1122 (80.0)	906	906
	0.201	(0.60 0.50)	509 (79.8)	499 (80.3)	490 (80.9)	472 (80.2)	388	388
	0.210	(0.70 0.60)	468 (80.1)	459 (80.0)	451 (80.0)	435 (80.1)	356	356
	0.231	(0.80 0.70)	385 (80.1)	379 (80.1)	373 (80.2)	361 (80.2)	294	294
	0.281	(0.90 0.80)	262 (80.7)	258 (80.6)	254 (80.5)	248 (80.4)	199	199
1.5	0.115	(0.55 0.50)	1819 (79.8)	1789 (80.0)	1760 (79.8)	1702 (79.7)	1183	1565
	0.119	(0.65 0.60)	1710 (80.0)	1682 (80.1)	1656 (80.1)	1603 (80.1)	1112	1471
	0.129	(0.75 0.70)	1454 (80.0)	1432 (80.1)	1411 (80.1)	1370 (80.0)	946	1251
	0.151	(0.85 0.80)	1053 (80.1)	1038 (80.0)	1025 (80.1)	1000 (80.2)	685	906
	0.232	(0.60 0.50)	451 (79.7)	443 (79.9)	436 (80.4)	422 (79.8)	293	388
	0.242	(0.70 0.60)	414 (80.0)	407 (80.1)	401 (80.0)	389 (80.1)	270	356
	0.266	(0.80 0.70)	341 (80.2)	336 (80.0)	331 (80.1)	322 (80.2)	222	294
	0.323	(0.90 0.80)	232 (80.6)	229 (80.7)	226 (80.7)	221 (80.7)	151	99

Behaviors of sample sizes (cont.)

3	0.160	(0.55 0.50)	1583 (79.7)	1578 (80.1)	1574 (80.4)	1568 (80.3)	611	1565
	0.165	(0.65 0.60)	1487 (80.1)	1483 (80.2)	1479 (80.1)	1474 (80.1)	574	1471
	0.179	(0.75 0.70)	1265 (80.1)	1262 (80.1)	1259 (80.1)	1254 (80.0)	489	1251
	0.211	(0.85 0.80)	916 (80.1)	914 (80.1)	912 (80.2)	909 (80.2)	354	906
	0.322	(0.60 0.50)	392 (79.6)	391 (79.6)	390 (79.6)	388 (79.5)	152	388
	0.336	(0.70 0.60)	360 (80.0)	359 (80.2)	358 (80.1)	357 (80.0)	139	356
	0.370	(0.80 0.70)	297 (80.3)	296 (80.3)	295 (80.2)	294 (80.2)	115	294
	0.450	(0.90 0.80)	202 (80.7)	201 (80.6)	201 (80.6)	200 (80.9)	78	199

Summary of the results

- The behaviors of sample sizes are in agreement with the results in the continuous case.
- π_{ik} is not close to 1
  - AS = AN << ASc = Fi = ANc
- π_{ik} is close to 1
  - AS < AN << ASc = Fi < ANc
- The behaviors of achieved sample sizes for K=3 are similar to those for K=2.
- The empirical power attains a pre-specified power.
  - The performance of ASc and Fi is better.

An illustration: PREMIER study

- Early aggressive rheumatoid arthritis
- ACR50
  - The percentage of patients in whom an ACR50 response was achieved
- mTSS
  - The mean change from baseline in the modified total Sharp score.
- Adalimumab + Methotrexate v.s. Methotrexate

Breedveld (2006)

An illustration: PREMIER study

Sample sizes per group with $\alpha = 0.025, 1 - \beta = 0.8$

mTSS		ACR50	Correlation ρ^{12}						
δ_1	σ_1	$(\pi_{T2}\ \pi_{C2})$	0.0	0.3	0.5	0.8	E_1	E_2	c_1^*/c_2^*
4.4	19.0	(0.59 0.46)	346	340	334	323	294	231	0.72
4.4	20.0	(0.59 0.46)	369	363	358	347	326	231	0.63
4.4	21.0	(0.59 0.46)	394	389	384	374	359	231	0.56
4.4	22.0	(0.59 0.46)	422	417	413	404	394	231	0.50

E_1, E_2: Sample size separately calculated for each endpoint so that the individual power is at least 0.8.

Breedveld (2006)

Joint density function of test statistics