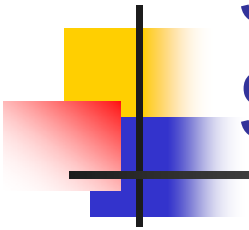
Selection of Endpoints and Sample Size Estimation in Clinical Trials

Jen-pei Liu, Ph.D.
National Taiwan University
&
National Health Research Institutes
at
Department of Family Medicine
National Cheng-Kung University
May 7, 2005

Selection of Endpoints and Sample Size Estimation in Clinical Trials



- Selection of Endpoints
- Sample Size Estimation
- Adjustment of Baseline

Types of Data

- Continuous Endpoints
- Numerical discrete data
  - Heart beats per minutes
  - Total NINSS
  - Total Hamilton Rating Scale for Depression
  - Total Alzheimer's Disease Assessment Scale
- Numerical continuous data
  - Age
  - Weight
  - ALT
  - Peak flow rate (liters per minute)
  - FEV_1 (% of predicted value)

Types of Data

- Categorical Endpoints
- Nominal scale data

  Classification of patients according to their attributes

  - Gender
  - Race
  - Occurrence of a particular adverse reaction
  - Occurrence of ALT>3 times upper normal limit

Types of Data

- Ordinal (Ordered) categorical data

  A certain order among different categories

  - Symptom score

    0 = no symptom, 1 = mild, 2 = moderate, 3 = severe

  - Severity of adverse reactions

- Censored Endpoints

  - Time to the occurrence of a pre-defined event.
  - The occurrence of the event may not observed for some patients. Then the time to the occurrence of the event for these subjects is censored

Types of Data

Cross-sectional vs. longitudinal data

- Cross-sectional data (snap shot at one time point)

  Clinical data are collected and evaluated at a particular time point during the trial

- Longitudinal data (snap shots at several time points)

  Clinical data collected and evaluated over a series of time points during the trial

Example

Knapp et al (JAMA 1994; 271: 985-991)

- A multi-center trial with 33 centers
- Double-blind, randomized, 4 parallel groups
- Forced escalation

  30 weeks of randomized treatment
- 6 visits

  The start of randomized treatment(baseline)6,12,18,24, and 30 weeks
- Cross-sectional data

  CIBI and ADAS-cog evaluated at the start of randomized treatment
- Longitudinal

  A series of CIBI and ADAS-cog evaluated at the start of the study, the start of randomized treatment,6,12,18,24, and 30 weeks

Types of Comparison

- Within-group (patient) comparison

  Comparison of the changes within the same patients at different time points during the trial.

- Between-group (patient) comparison

  Comparison between groups of patients under different treatments.

Example: Major depression disorder

Stark and Hardison (VCP, 1985;46,53-58)

Cohn and Wilcox (JCP,1985:46,21-31)

- Double-blind, randomized, three parallel groups
- One-week placebo washout period
- Fluoxetine vs. imipramine vs. placebo
- 6 weeks of randomized treatments
- Primary efficacy endpoint

  HAM-D score at the last follow-up visit
- Within each group

  Change from baseline in HAM-D score
- Between groups

  Comparison of the change from baseline in HAM-D score between groups

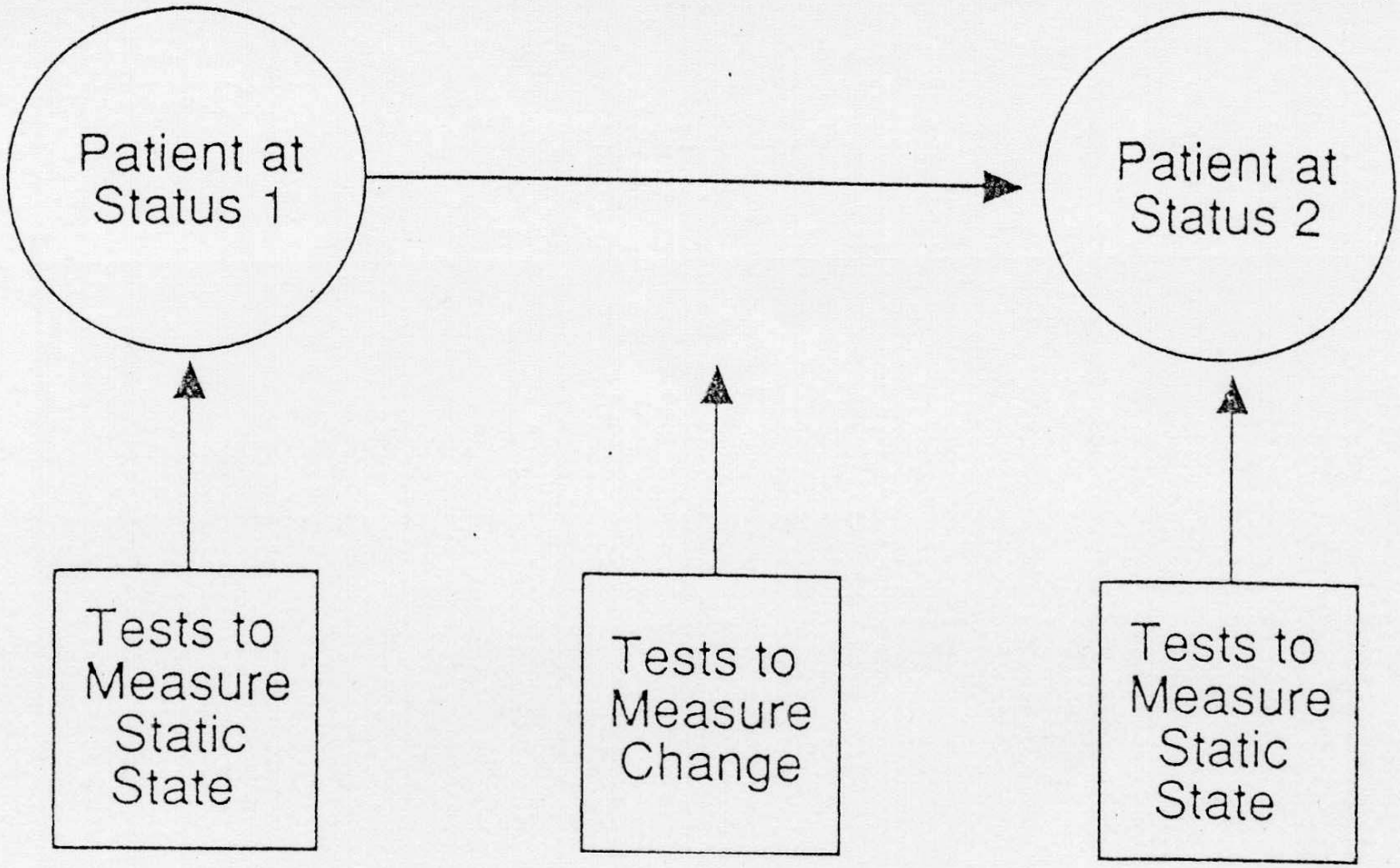


FIG. 22.1 Differences between tests that measure a patient's static or dynamic state. Tests may also measure patients at two or more static states to demonstrate (i.e., measure) change. Some types of states are not sensitive to changes caused by medicines (e.g., intelligence quotient).

Endpoints

- Raw measurements at a time point.
- Change at a time point from baseline.
- Percent change at a time point from baseline.
- Clinically meaningful targeted value attained at a time point, i.e. sitting DBP <= 85 mm Hg
- Selection of time points should be able to measure the effect of the intervention.

Selection of Endpoints

- Endpoints should reflect the change of clinical status caused by the intervention.
- Endpoints should be sensitive to the change of clinical status caused by the intervention.
- Endpoints should be validated.
- Raw measurements at a time point can only measure the static clinical status.
- Change at a time point from baseline can measure the magnitude of the change of clinical status caused by the intervention.
- Change from baseline has the same unit as the raw measurement

Selection of Endpoints

- Percent change at a time point from baseline measures the relative magnitude of the change of clinical status caused by the intervention.
- Percent change from baseline is unitless.
- The same percent change may reflect different magnitudes of change
- 20/100 = 2/10 = 200/1000 = 20%

Selection of Endpoints

- One of the key inclusion criteria for clinical trial in treatment of mild to moderate essential hypertension is sitting DBP being between 95-115 mm Hg.
- Three changes from baseline: 115 $\rightarrow$ 105, 105 $\rightarrow$ 95, 95 $\rightarrow$ 85.
- 95 Changes from baseline: 8.7%, 9.5%, 10.5%
- Only 95 $\rightarrow$ 85 reaches the clinically meaningful targeted value.

Selection of Endpoints

- Endpoints should reflect clinically meaningful interpretation and applicability.
- Clinically meaningful targeted value > change from baseline > percent change from baseline.
- Clinical investigators should have responsibility for determination of the efficacy endpoints used in the clinical trials.

Selection of Endpoints

	LDL	HDL	TG
Targeted Value	< 100mg/dL	40-60 mg/dL	< 150 mg/dL
Bile acid Binding Resin	↓15-30%	↑3-5%	no change
Nicotinic acid	↓ 5-25%	↑15-35%	↓15-25%
Fibric acid	↓ 5-20%	↑10-20%	↓ 20-50%
HMG-CoA Inhibitor	↓18-55%	↑ 3-5%	↓ 7-30%

Measures for Comparison in Proportions between Groups

Difference in proportions

Relative risk

The ratio of the proportions of the test group to the control.

Odds ratio

The ratio of the odds of the test group to the control.

Odds

The number of patients with the attribute to that without the attribute.

The US Physicians' Health Study (NEJM 1989; 321: 129-35)

	Aspirin	Placebo
N	11037	11034
MI	139 (1.26%)	239 (2.17%)
No MI	10898 (98.74%)	10795 (97.83%)

Difference in proportion of MI = 1.26% - 2.17% = -0.91%
(average of fewer 91 MIs per 10,000)

Relative risk of MI for aspirin = 1.26% / 2.17% = 0.581
(the risk of MI in aspirin reduces 42%)

Odds ratio of MI for aspirin = (139 / 10898) / (239 / 10798)
= 1.275% / 2.214% = 0.576
(the odds of MI in aspirin reduces 42%)

Difference in proportions and relative risk can only be used in prospective studies while odds can be used in both prospective as well as retrospective studies.

Categorical Endpoints

- Difference in proportions provides the absolute magnitude of difference.
- Both relative risk and odds ratio gives the relative magnitude of difference.
- 50% $\rightarrow$ 25% and 0.05% $\rightarrow$ 0.025% both yield a relative risk of 50% but differences in proportion are 25% and 0.025% respectively.
- Relative risk and odd ratio are appropriate when the proportion of the event for control group is small (<5%).
- When the proportion of the event is small (<5%), the relative risk $\approx$ Odds ratio.

Censored Data

- Median survival

  The time to the pre-defined event (e.g. death) occurring in 50% of the patients.

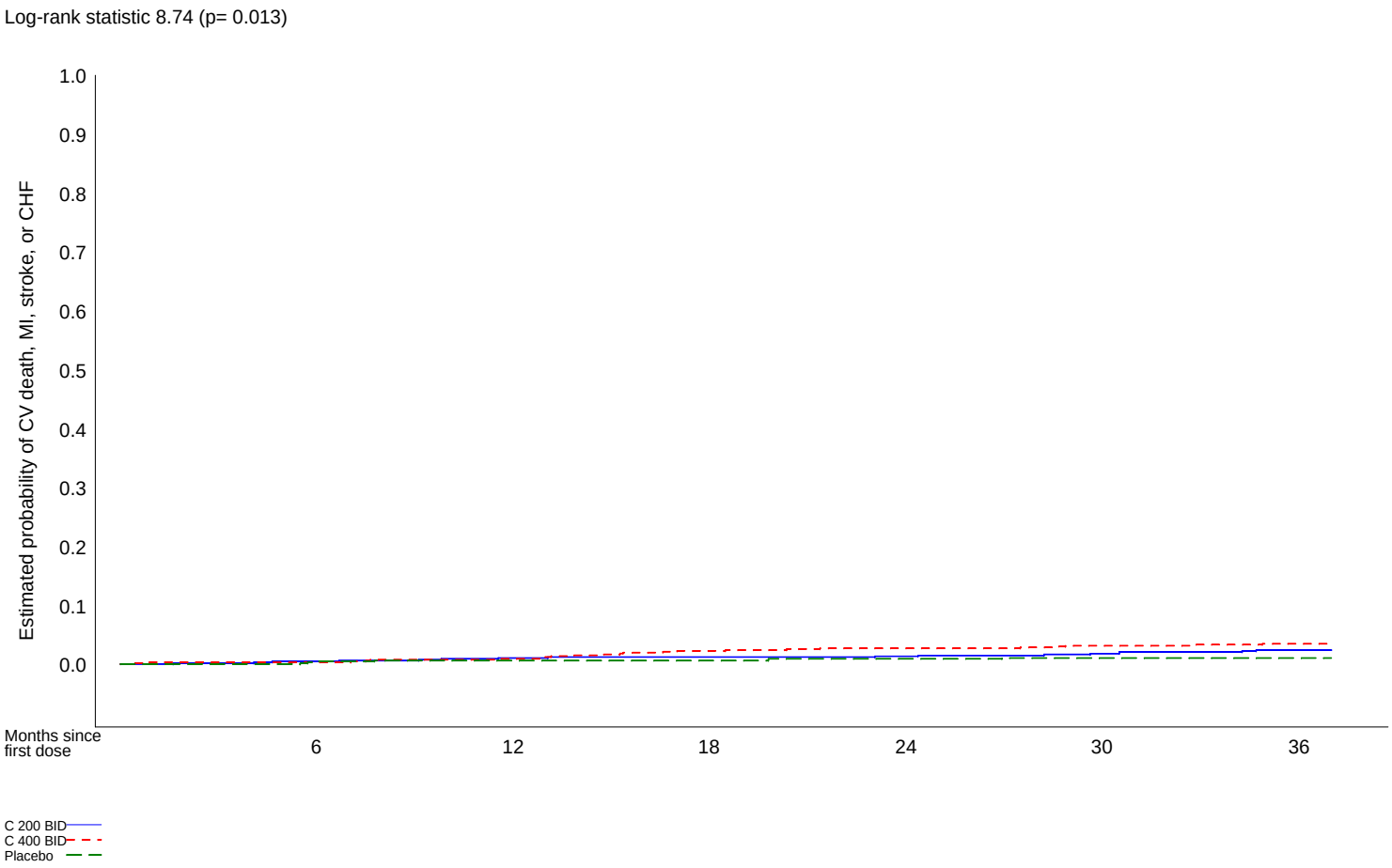
- Survival rate at a particular time point
- Hazard ratio

  The hazard of the occurrence of a pre-defined event of the test group to the control group

- Survival rate and hazard ratio can be

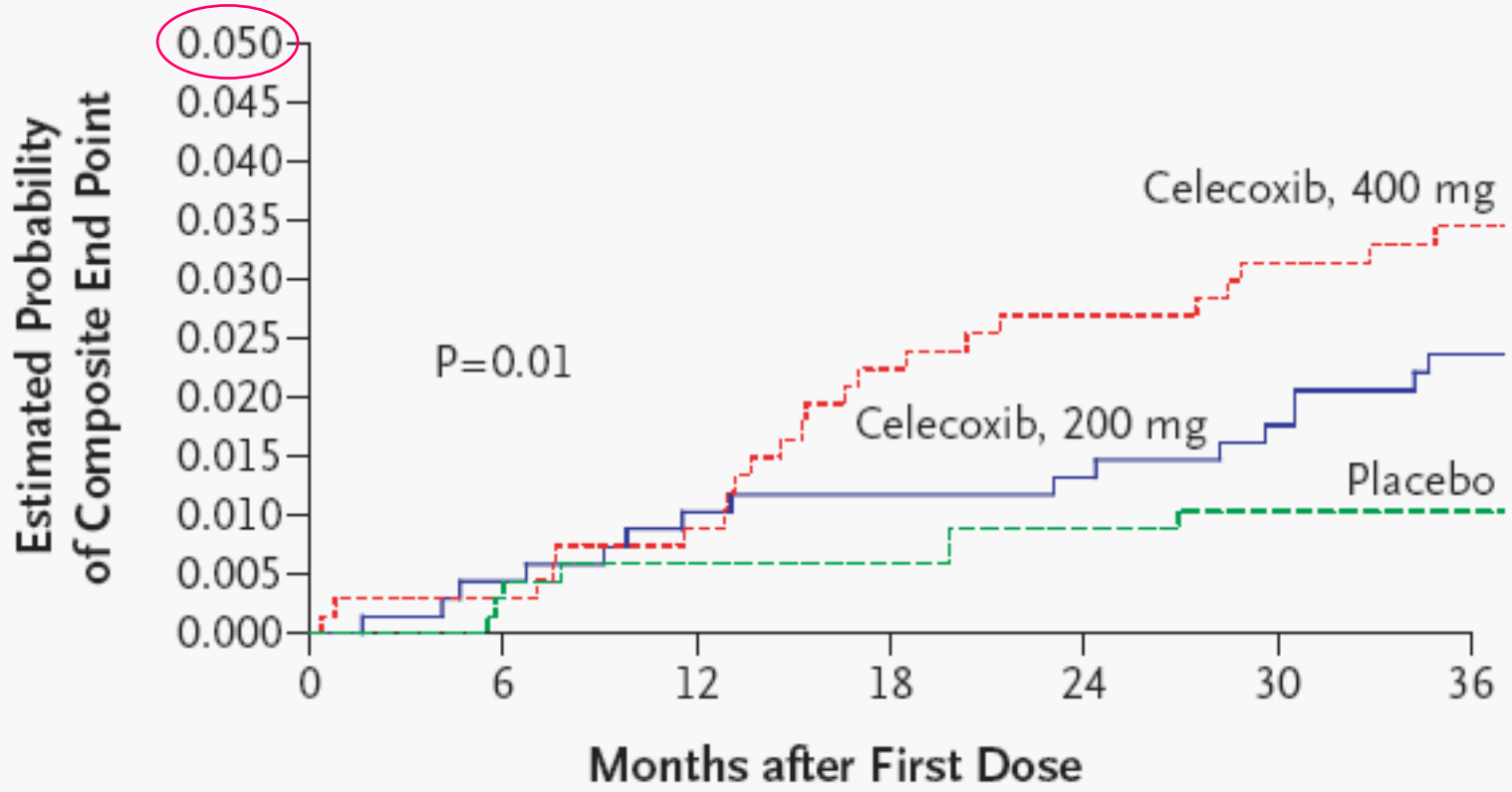
  exchanged under statistical models

Kaplan-Meier Estimates of the Risk of Serious CV Events in the APC Trial by Treatment Arm*



CEC adjudication data (received 02/03/2005) RAND data (received 11/23/2004) CRF data (received 02/03/2005)
E:\Proj\APC Cox2\Programs\f_KMplot.sas v.010 death_MI_stroke_CHF_005_01.com (last run: 02/14/2005, 10:46)

Kaplan-Meier Estimates of the Risk of Serious CV Events in the APC Trial by Treatment Arm*



No. at Risk

Celecoxib, 400 mg	671	669	665	655	651	648	576
Celecoxib, 200 mg	685	681	676	675	673	670	595
Placebo	679	677	675	672	668	667	585

Sample Size Estimation

- Investigators must determine the expected difference you want to detect in the trial and variability associated with clinical endpoints
- Expected difference should be clinically meaningful.
- Expected difference should not be over-exaggerated nor over-conservative.
- Variability should be realistic.

Sample Size Estimation

- All formulas for sample size estimation provide the minimal number of patients required for trials.
- Sample size increases in square as the expected difference decreases or the standard deviation increases.
- Sample size increase four times as the expected difference decreases 50% or standard deviation doubles.
- Sample size increases if the significance level decreases or power increases.

Sample Size Estimation

Introduction

- E = experimental treatment group
- C = control treatment group

We consider

- Mean difference
- Difference in proportions
- Relative risk
- Equivalence trial
- Time to event

Notation

- α = type Ⅰ error
- β = type Ⅱ error
- Z_{α} = the α th normal quartile: Level of Significance

  $P\{ Z > Z_{\alpha} \} = \alpha$ one sided

  $P\{|Z| > Z_{\alpha/2}\} = \alpha$ two sided
- Z_{β} = the β th normal quartile: Level of Power

  $P\{Z < Z_{\beta}\} = 1 - \beta$

References

- Donner, A. (1984) Approaches to sample size estimation in the design y clinical trials. Statistic in Medicine, vol. 3, 198-214
- Lachin, J.M. (1981) Introduction to sample size determination and power analysis for clinical trials, Controlled Clinical Trials, vol. 2, 93-114
- Chow, S. C., and Liu, J.P. (2004) Design and Analysis of Clinical Trials: Concepts and Methodologies, 2nd Ed. Chapter 11, Wiley, New York, New York.

Assume Equal Allocation

- Mean difference: $\mu_E - \mu_C$

$$Assume \quad N(\mu_C, \sigma_C^2) \; and \; N(\mu_E, \sigma_E^2)$$

$$\sigma_C^2 \; and \; \sigma_E^2 \; are \; known$$

$$H_0 : \mu_C - \mu_E = 0$$

$$H_A : \mu_C - \mu_E = \Delta$$

$$n = \frac{(\sigma_C^2 + \sigma_E^2)[Z_{\alpha/2} + Z_\beta]^2}{\Delta^2} \quad per \; group$$

$$= \frac{2\sigma^2[Z_{\alpha/2} + Z_\beta]^2}{\Delta^2} \quad if \; \sigma_C^2 = \sigma_E^2 = \sigma^2$$

Example

$$H_0 : \mu_C - \mu_E = 0$$

$$H_A : \mu_C - \mu_E = 8$$

$$\sigma^2 = 15^2$$

$$\alpha = 0.05 \ (two\ sided), Z_\alpha = 1.96$$

$$1 - \beta = 0.8, Z_\beta = 0.84$$

$$n = \frac{2(15)^2[1.96 + 0.84]^2}{8^2} = 55.2 \ or \ 56$$

Risk difference: P_E - P_C

$$P_C = event\ \ rate\ \ of\ control\ \ group$$

$$P_E = event\ \ rate\ \ of\ \text{exp}erimental\ \ group$$

$$H_0 : P_E = P_C$$

$$H_A : P_E - P_C = \Delta$$

$$n = \left[Z_\alpha \sqrt{2\bar{P}(1-\bar{P})} + Z_\beta \sqrt{P_E(1-P_E) + P_C(1-P_C)} \right]^2 \Big/ \Delta^2 \qquad per\ group$$

$$where \quad \bar{P} = \frac{P_E + P_C}{2}$$

Example

$$H_0 : P_C = P_E$$

$$H_A : P_C = 0.6,\ P_E = 0.4$$

$$\alpha = 0.05\ (two\ sided),\ Z_\alpha = 1.96$$

$$1 - \beta = 0.8,\ Z_\beta = 0.84$$

$$n = \frac{\left[1.96\sqrt{2(0.5)(0.5)} + 0.84\sqrt{(0.4)(0.6) + 0.6(0.4)}\right]^2}{(0.6 - 0.4)^2} = 96.8\ or\ 97$$

Unequal Allocation

n $\rightarrow$ experimental group

sn: control group

$$n = \left[Z_{\alpha} \sqrt{(s+1)\overline{P}_s(1-\overline{P}_s)} + Z_{\beta} \sqrt{sP_E(1-P_E) + P_C(1-P_C)} \right]^2 \Big/ (s\sigma^2)$$

$$\textit{where} \quad \overline{P}_s = \frac{(P_E + sP_C)}{(s+1)}$$

Relative Risk: P_E / P_C

P_C, P_E *and* $\bar{P}$ *are defined as before*

$$R = P_E / P_C$$

$$H_0 : R = 1$$

$$H_A : R = P_E / P_C = r$$

$$n = \left[Z_\alpha \sqrt{2\bar{P}_R(1-\bar{P})} + Z_\beta \sqrt{P_C[1+r-P_C(1+r^2)]} \right]^2 \Big/ [P_c(1-r)]^2$$

where $\bar{P}_R = \dfrac{P_C(1+r)}{2}$

Compare with results for risk difference

Example

$$H_0 : R = 1$$

$$H_A : R = \frac{0.4}{0.6} = \frac{2}{3}$$

$$\alpha = 0.05 \ (two \ sided), \ Z_\alpha = 1.96$$

$$1 - \beta = 0.8, \ Z_\beta = 0.84$$

$$n = 96.8 \ or \ 97$$

Non-interiority Trial

$$H_0 : P_C \geq P_E + \theta$$

$$H_A : P_C < P_E + \theta, \quad \theta > 0$$

$$n = \frac{\left[Z_\alpha \sqrt{2\bar{P}(1-\bar{P})} + Z_\beta \sqrt{P_E(1-P_E) + P_C(1-P_C)} \right]^2}{(P_C - P_E - \theta)^2}$$

where $P_C < P_E + \theta$ *and* $\theta > 0$

In practice, set $P_E = P_C$ and

θ = the difference in treatment efficacy

Example

$P_E = 0.75 \quad P_C = 0.8$

$H_0 : P_C \geq P_E + \theta$

$H_A : P_C < P_E + \theta, \quad \theta = 0.1$

$\alpha = 0.1\,(one-sided), Z_\alpha = 1.282$

$1-\beta = 0.8 \quad Z_\beta = 0.84$

$$n = \frac{\left[1.28\sqrt{2(0.775)(0.225)} + 0.84\sqrt{(0.8)(0.2)+(0.75)(0.25)}\right]^2}{(0.75-0.8-0.1)^2} = 624$$

where $P_C < P_E + \theta$ *and* $\theta > 0$

When P_E is assumed 0.8, then n reduces to 145

Unequal allocation

$$n = \frac{\left[Z_\alpha\sqrt{(s+1)\overline{P}_s(1-\overline{P}_s)} + Z_\beta\sqrt{sP_E(1-P_E)+P_C(1-P_C)}\right]^2}{s(P_C-P_E-\theta)^2}$$

Time to Event ---
Assume exponential Distribution

$$Survival\ curve\ S_e(t) = e^{-t/ue},\ S_c(t) = e^{-t/uc}$$

$$H_0 : \mu_E / \mu_C = 1 \qquad H_A : \mu_E / \mu_C = \theta$$

$$Follow\ up\ to\ \ failure$$

$$n = \frac{2(Z_\alpha + Z_\beta)^2}{\log_e(\theta)}$$

Time to Event ---
Assume exponential Distribution

Patient's enter the trial at a uniform rate over a T-year period. If the trial terminates at a time T, then

$$n = \frac{(Z_\alpha + Z_\beta)^2[\phi(\mu_E) + \phi(\mu_C)]}{(\mu_E^{-1} - \mu_C^{-1})^2}$$

$$where \quad \phi(\mu_i) = \frac{T}{\mu_i^3} \Big/ \left[\frac{T}{\mu_i} - 1 + \exp(-\frac{T}{\mu_i})\right] \quad i = C, E$$

Time to Event ---
Assume exponential Distribution

Patients are recruited over the interval $(0, T_0)$, but with a follow up until T, then

$$\phi(\mu_i) = \frac{1}{\mu_i^2}\left[1 - \frac{\{\exp[-(T - T_0)/\mu_i] - \exp(-T/\mu_i)\}\mu_i}{T_0}\right]^{-1}$$

Example

$$H_A : \mu_E / \mu_C = 1.5$$

$$\alpha = 0.05\,(one\ sided), \quad Z_\alpha = 1.645$$

$$1 - \beta = 0.9,\ Z_\beta = 1.282$$

Follow up to failure

$$n = \frac{2(1.645 + 1.282)^2}{\log e^2 (1.5)} = 105$$

Example

Study over a $5-$ year period, $H_A: \mu_E/\mu_C = 1.5$, $\mu_C = 3$

$$\phi(\mu_C) = -\frac{5}{3^2}\left[\frac{5}{3} - 1 + e^{-\frac{5}{3}}\right]^{-1} = 0.217$$

$$\phi(\mu_E) = \frac{5}{4.5^3}\left[\frac{5}{4.5} - 1 + e^{-\frac{5}{4.5}}\right]^{-1} = 0.125$$

$$n = \frac{(1.645 + 1.282)^2 (0.217 + 0.125)}{(4.5^{-1} - 3^{-1})^2} = 238$$

recruited over a $4-$ year period but follow up for an addition 1 year

$$\phi(\mu_C) = 0.184$$

$$\phi(\mu_E) = 0.105$$

$$n = 207$$

Account for Patient dropout

Drop-out rate in group E

$$P_E^* = P_E(1-d) + P_C d$$

$$\Delta^* = P_E^* - P_C = (1-d)(P_E - P_C)$$

substitute Δ^* *for* Δ

$$n^* = \frac{n}{(1-d)^2}$$

Two-sided equivalence

- For two parallel groups

$$\text{Define } \mu_L - \mu_R = \theta \text{ and let } -\delta_L = \delta_U = \delta$$

$$\text{and } \sigma_0^2 \text{ and } \sigma_a^2 \text{ be the variance when}$$

$$\mu_L - \mu_R = \theta \text{ and } \mu_L - \mu_R = \delta \text{ ,}$$

$$\text{where } 0 \le \theta \le \delta$$

Two-sided equivalence

- General Formulas for sample size per group

$$For\ \mu_T - \mu_R = 0$$

$$n = \left[Z(\alpha)\sigma_0 + Z(\beta/2)\sigma_a\right]^2 / \delta^2$$

$$For\ \mu_T - \mu_R = \theta \le \delta$$

$$n = \left[Z(\alpha)\sigma_0 + Z(\beta/2)\sigma_a\right]^2 / [\delta - \theta]^2$$

Two-sided equivalence

For continuous endpoints, assume that

$\sigma_0^2 = \sigma_a^2 = 2\sigma^2$ (*Liu and Chow*, 1992; *Liu*, 1995)

For $\mu_T - \mu_R = 0$

$$n = 2\sigma^2 \left[Z(\alpha) + Z(\beta/2) \right]^2 / \delta^2$$

For $\mu_T - \mu_R = \theta \le \delta$, *approximate formula*

$$n = 2\sigma^2 \left[Z(\alpha) + Z(\beta) \right]^2 / (\delta - \theta)^2$$

Adjustment of Covariates

- Covariates are factors that affect the primary efficacy endpoints
  - prognostic, risk, or confounding factors
  - age, gender, race, disease severity, etc.
- Patient-specific covariates

  Covariates measured before randomization
  - Baseline FEV_1, FVC, etc.
- Time-dependent covariates

  Covariates measured after randomization
  - May be affected by the treatments
  - Cd_4 level during the trials

Adjustment of Covariates

- Stratification based on known covariates before randomization and conduct of the trials (pre-randomization adjustment).
- Adjustment of covariates in the analysis improvement of the precision of the estimated treatment effects (post-randomization adjustment).
- Adjustment of covariates reduces the variability associated with the estimated treatment effect.
- The estimated treatment effect is unbiased without adjustment of covariates as long as assignment of treatments is random (That is the rewards when you pay the price of randomization).

Adjustment of Covariates

- Need to check the treatment-by-covariate interaction.
- If the treatment-by-covariate interaction exists, generalizability of the results is limited.
- Avoid to adjust the primary endpoints for the covariates measured after randomization.
- Specify the covariates in the protocol.

Summary

- Efficacy endpoints should be clinically meaningful.
- Minimal sample size should be chosen to provide sufficient power to detect a clinically significant difference.
- A under-powered trial is unethical.
- Adjustment of covariate can reduce variability of the estimated treatment effect.
- For a randomized trial, unadjusted treatment effect is still unbiased.