ST520 Statistical Principles of Clinical Trials

Chapter 1: Introduction - Summary

- **In prospective studies and case-control studies, what quantities can be estimated?**

1 Introduction

1.1 Brief Introduction to Epidemiology

- **Prevalence** of disease $P(D)$
- **Incidence** of disease: probability of getting disease during a certain time period
- **Relative risk** $\psi = \frac{P(D|E)}{P(D|\overline{E})}$ (easier interpretation, association not causation)
- **Odds ratio** $\theta = \frac{P(D|E)/\{1-P(D|E)\}}{P(D|\overline{E})/\{1-P(D|\overline{E})\}}$
- $\psi > 1 (= 1, < 1) \Longleftrightarrow \theta > 1 (= 1, < 1)$
- For rare disease, $P(D \mid E) \approx 0$ and $P(D \mid \overline{E}) \approx 0$, then $\theta \approx \psi$
- Estimates
  - $\hat{\theta} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$
  - $\hat{V}\{\log(\hat{\theta})\} = \frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}$
  - $\hat{V}(\hat{\theta}) = \hat{\theta}^2 \left(\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}\right)$

Cross-sectional study

- can estimate $P(D)$, $P(D \mid E)$ and the similar type (i.e., $P(\bar{D})$, $P(D \mid \bar{E})$, and so on), ψ, θ

Longitudinal study

- Prospective
  - can estimate $P(D \mid E)$ and the similar type, ψ, θ
- Retrospective: case-control
  - can estimate $P(E \mid D)$ and the similar type, θ (ψ for rare disease)

1.2 Brief Introduction and History of Clinical Trials

Chapter 2: Phase I and II clinical trials - Summary

- **Phases of clinical trials and major goals**
- **Phase II designs**
- **Gehan's two-stage designs (precision/Confidence interval/expected sample size)**
- **Simon's two-stage designs**

2 Phase I and II clinical trials

2.1 Phases of Clinical Trials

- Phase I (toxicity and dose-finding)
- Phase II (screening and feasibility)
- Phase III (comparative study)
- Phase IV (post marketing)

2.2 Phase II clinical trials

- surrogate markers
- Example (a binary endpoint): π =the response rate
  - X =the number of responses among a random sample of n
  - $X \sim b(n, \pi)$
  - Normal approximation $p = X/n \sim N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$
  - $(1-\alpha)$ confidence interval $p \pm z_{1-\alpha/2}\left\{\frac{p(1-p)}{n}\right\}^{1/2}$
- Exact confidence interval $C(X)$ such that $C(k) = (\pi_L(k), \pi_U(k))$

- $P_{\pi_L(k)}(X \geq k) = \sum_{j=k}^{n} \left(\begin{array}{c} n \\ j \end{array} \right) \pi_L(k)^j \{1 - \pi_L(k)\}^{n-j} = \alpha/2$
- $P_{\pi_U(k)}(X \leq k) = \sum_{j=0}^{k} \left(\begin{array}{c} n \\ j \end{array} \right) \pi_U(k)^j \{1 - \pi_U(k)\}^{n-j} = \alpha/2$

- **Gehan**'s two-stage design
  - Stage I. (start with n_0 and a minimum acceptable response rate π_0) if no one responds, declare a failure.
    - determine n_0 : $P_\pi(X = 0) \leq (1 - \pi_0)^{n_0} \leq \alpha$
  - Stage II. (add $n - n_0$) count the total number of responses, calculate p and construct a CI for π.
    - determine n : based on the precision of 95% CI using π_0.
  - The expected sample size is

$$\begin{aligned} & n_0 P(\text{stopping at Stage I}) + nP(\text{stopping at Stage II}) \\ = \; & n_0 P(X = 0 \mid \pi_0) + n\{1 - P(X = 0 \mid \pi_0)\} = (n_0 - n)(1 - \pi_0)^{n_0} + n \end{aligned}$$

- **Simon**'s two-stage design (n_1, n, r_1, r) with $\pi_0 < \pi_1$
  - Stage I. (start with n_1) $X_1 \sim b(n_1, \pi)$ if $X_1 \leq r_1$, declare a failure.
  - Stage II. (add $n_2 = n - n_1$) $X_2 \sim b(n_2, \pi)$ if $X_1 + X_2 \leq r$, declare a failure.

    constraints:

* $P(\text{success} \mid \pi \leq \pi_0) \leq \alpha \Leftrightarrow P\{(X_1 > r_1)\&(X_1 + X_2 > r) \mid \pi = \pi_0\} \leq \alpha$

* $P(\text{failure} \mid \pi \geq \pi_1) \leq \beta \Leftrightarrow P\{(X_1 > r_1)\&(X_1 + X_2 > r) \mid \pi = \pi_1\} \geq 1 - \beta$

– Optimal design : to minimize the expected sample size

$$\begin{aligned} & n_1 P(\text{stopping at Stage I}) + nP(\text{stopping at Stage II}) \\ = \; & n_1 P(X_1 \leq r_1 \text{or } X_1 > r \mid \pi_0) + nP(r_1 + 1 \leq X_1 \leq r \mid \pi_0). \end{aligned}$$

Chapter 3: Phase III clinical trials - Summary

- **Meta analysis**
- **Methods of Moment**

3 Phase III clinical trials

3.1 Why are clinical trials needed

3.2 Issues to consider before designing a clinical trial

3.3 Ethical issues

3.4 The randomized clinical trial - hierarchical models

- To address the question of whether the results from the different studies are random samples form underlying groups with a common response rate or from groups with different underlying response rates.
- Hierarchical models for (n_i, X_i, π_i) $i = 1, \ldots, N$
  - (1st) $\pi_1, \ldots, \pi_N \overset{iid}{\sim} (\mu_\pi, \sigma_\pi^2)$
  - (2rd) $X_i \mid n_i, \pi_i \sim b(n_i, \pi_i)$, $i = 1, \ldots, N$
- Goal is to estimate σ_π^2
- **Law of conditional expectation**: $E(X) = E\{E(X \mid Y)\}$
- **Law of conditional variance**: $V(X) = E\{V(X \mid Y)\} + V\{E(X \mid Y)\}$
- $p_i = X_i/n_i$

- $E(p_i \mid \pi_i, n_i) = \pi_i$
- $V(p_i \mid \pi_i, n_i) = \pi_i(1-\pi_i)/n_i$
- $E(p_i) = E\{E(p_i \mid \pi_i, n_i)\} = \mu_\pi$
- $V(p_i) = E\{V(p_i \mid \pi_i, n_i)\} + V\{E(p_i \mid \pi_i, n_i)\} = E\{\pi_i(1-\pi_i)/n_i\} + \sigma_\pi^2$

- $\bar{p} = N^{-1}\sum_{i=1}^N p_i$
- $s_p^2 = \frac{\sum_{i=1}^N (p_i - \bar{p})^2}{N-1}$: $E(s_p^2) = V(p_i)$
- $N^{-1}\sum_{i=1}^N \frac{p_i(1-p_i)}{n_i - 1}$: $E\left\{N^{-1}\sum_{i=1}^N \frac{p_i(1-p_i)}{n_i-1}\right\} = E\left\{\frac{\pi_i(1-\pi_i)}{n_i}\right\}$
- $\hat{\sigma}_\pi^2 = \left\{\frac{\sum_{i=1}^N (p_i - \bar{p})^2}{N-1}\right\} - \left\{N^{-1}\sum_{i=1}^N \frac{p_i(1-p_i)}{n_i-1}\right\}$ is unbiased for σ_π^2.
- $\hat{\mu}_\pi = \frac{p_1 + \ldots + p_N}{N}$

Chapter 4: Randomization - Summary

- **Roles of randomization**
- **Design-based inference vs model-based inference**
- **Schemes of randomization and goals**

4 Randomization

- Advantage of randomization: eliminate conscious and unconscious biases
- Disadvantage of randomization:
  - interference with physician patient relationship
  - resources expended in the control group

4.1 Inference

4.1.1 Design-based inference

- Sharp null hypothesis: A & B (two treatments) are exactly the same for each patient
- Test statistics: T =difference of the sample means
- The distribution of test statistics is induced by the randomization $\implies$ the permutation distribution of T
- One-sided p-value (the alternative is "A is better than B") $P(T \geq t_{obs} \mid \text{sharp } H_0) = \frac{\#\text{of } t_i \geq t_{obs}}{\#\text{of } t_i \text{ under permutation}}$
- Two-sided p-value (the alternative is "A is different than B") $P(|T| \geq |t_{obs}| \mid \text{sharp } H_0) = \frac{\#\text{of } |t_i| \geq |t_{obs}|}{\#\text{of } t_i}$
- Remark: In the permutational distribution, we treat each individual's response as fixed. Randomness is induced by the treatment assignment mechanism.

4.1.2 Model-based inference

- The distribution of test statistics is induced by assumptions about a super-population and a probability model.
- Statistical model:
- $Y_1, Y_2 \overset{iid}{\sim} N(\mu_A, \sigma^2)$
- $Y_3, Y_4 \overset{iid}{\sim} N(\mu_B, \sigma^2)$
- Null hypothesis: $H_0 : \mu_A = \mu_B$
- $T = \frac{\bar{Y}_A - \bar{Y}_B}{s_p(n_A^{-1} + n_B^{-1})^{1/2}} \overset{H_0}{\sim} t_{n_A+n_B-2}$
- p-value$= P(t_{n_A+n_B-2} \geq t_{obs} \mid H_0)$

4.1.3 Causal inference from an experimental sample to a larger population

- Potential outcomes $Y^*(1)$ and $Y^*(2)$
- The average treatment effect is ATE$= E\{Y^*(2) - Y^*(1)\}$
- Treatment indicator $A \in \{1, 2\}$
- Observed outcome $Y = Y^*(1)I(A = 1) + Y^*(2)I(A = 2) = Y^*(A)$ under SUTVA
- Randomization: $A \perp\!\!\!\perp \{Y^*(1), Y^*(2)\}$

- implication $E\{Y^*(a)\} = E\{Y^*(a) \mid A = a\} = E\{Y \mid A = a\} \equiv \mu_a$ for $a = 1, 2$
- $\mu_2 - \mu_1$ is the same as ATE
- estimator: $\bar{Y}_2 - \bar{Y}_1$

4.2 Fixed allocation randomization

1. Simple randomization
   - Two treatments A & B; $\Delta = \mu_2 - \mu_1$; $\hat{\Delta} = \bar{Y}_2 - \bar{Y}_1$
   - Treatment allocation with $P(A) = \pi$
   - $V(\hat{\Delta}) = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$
   - Advantage: easy; impossible to guess the next treatment assignment; iid
   - Disadvantage: possibility of severe treatment imbalance
2. Permuted block randomization
   - try to balance A & B
   - fixed block size
   - varying block size
3. Stratified randomization (permuted block randomization): see homework 3 additional question

- The maximum imbalance between A & B is # of strata$\times$(block size)/2
- Advantage: treatment groups are similar; more precise estimates of treatment difference
- Disadvantage: too many prognostic factors to form strata, very few (or even zero) patient in some strata, back to simple randomization

- Linear combination of random variables
  - Let $Z_1, \ldots, Z_m$ be random variables, and let $a_0, \ldots, a_m$ be constants
  - $E(a_0 + a_1 Z_1 + \ldots + a_m Z_m) = a_0 + a_1 E(Z_1) + \ldots + a_m E(Z_m)$
  - In addition, $Z_1, \ldots, Z_m$ are mutually independent, then $V(a_0 + a_1 Z_1 + \ldots + a_m Z_m) = a_1^2 V(Z_1) + \ldots + a_m^2 V(Z_m)$
  - Eg. Let $X_1, \ldots, X_n \overset{iid}{\sim}$ be random variables, and let $a_0, \ldots, a_m$ be constants
  - $E(a_0 + a_1 Z_1 + \ldots + a_m Z_m) = a_0 + a_1 E(Z_1) + \ldots + a_m E(Z_m)$
  - In addition, $Z_1, \ldots, Z_m$ are mutually independent, then $V(a_0 + a_1 Z_1 + \ldots + a_m Z_m) = a_1^2 V(Z_1) + \ldots + a_m^2 V(Z_m)$

- Effect of blocking within strata on the precision of estimators
  - $Y_i = \mu + \alpha S_i + \beta X_i + \epsilon_i$, strata indicator $S_i \in \{0, 1\}$, treatment indicator $X_i \in \{0, 1\}$,$\epsilon_i \overset{iid}{\sim} (0, \sigma^2)$
  - $\hat{\Delta} = \bar{Y}_A - \bar{Y}_B$
  - Under stratified randomization: $V(\hat{\Delta}) = \frac{4\sigma^2}{n}$
  - Under permuted block randomization: $V(\hat{\Delta}) = \sigma^2 \frac{4}{n}$

- Under simple randomization: $V(\hat{\Delta}) = \{\sigma^2 + \alpha^2\theta(1-\theta)\}E\left(\frac{1}{n_A} + \frac{1}{n-n_A}\right)$
- The test statistics $T = \frac{\bar{Y}_A - \bar{Y}_B}{s_p(\frac{1}{n_A}+\frac{1}{n_B})^{1/2}}$ is conservative in a stratified random design with $\alpha \neq 0$ (i.e. some strata effect)

4.3 Baseline adaptive randomization

- Efron biased coin design
- Urn Model (L.J. Wei)
- Minimization method of Pocock and Simon

4.4 Response adaptive randomization

- Play-the-Winner Rule (Zelen)
- Urn Model (L.J. Wei)

5 Some additional issues in Phase III clinical trails

- Blinding and Placebos
- Ethics
- The protocol document

Chapters 6 and 7: Sample Size Calculations - Summary

- **Types of outcomes: continuous/binary/survival outcomes**
- 2 **vs** K **treatment comparisons**
- **Power and sample size calculation under clinically important alternatives (least favorable configuration)**
- **Sample size calculation: need to specify nuisance parameters**
- **Equivalency trials**

6 Sample Size Calculations

- Generally, $\begin{cases} T_n \overset{\Delta=0}{\sim} N(0,1) & \Rightarrow \text{one-sided level-}\alpha \text{ test} \Rightarrow \text{rejection region: } T_n \geq z_\alpha \\ T_n \overset{\Delta=\Delta_A}{\sim} N(\phi(n,\Delta_A,\theta), \sigma_*^2(\Delta_A,\theta)) & \Rightarrow \text{power: } P_{\Delta=\Delta_A}(T_n \geq z_\alpha) = 1-\beta \end{cases}$
- Key formula for sample size calculation:

$$\phi(n,\Delta_A,\theta) = \begin{cases} Z_\alpha + Z_\beta \sigma_*(\Delta_A,\theta) \text{ for a one-sided test} \\ Z_{\alpha/2} + Z_\beta \sigma_*(\Delta_A,\theta) \text{ for a two-sided test} \end{cases}$$

6.1 Comparison of two means

- $H_0 : \Delta = \mu_1 - \mu_2 \leq 0$ vs $H_A : \Delta > 0$
- $Y_i \mid A_i = 1 \sim (\mu_1, \sigma^2), Y_i \mid A_i = 2 \sim (\mu_2, \sigma^2)$
- $T_n = \dfrac{\bar{Y}_1 - \bar{Y}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$
- $\begin{cases} T_n \overset{\Delta=0}{\sim} N(0,1) \\ T_n \overset{\Delta=\Delta_A}{\sim} N\left(\frac{\Delta_A}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}, 1\right) \end{cases}$

6.2 Comparison of two proportions

- Hypothesis testing: $H_0 : \pi_1 \leq \pi_2$ $(\Delta = \pi_1 - \pi_2 \leq 0)$ vs $H_A : \pi_1 > \pi_2 (\Delta > 0)$

- Data: $X_1 \sim b(n_1, \pi_1)$, $X_2 \sim b(n_2, \pi_2)$, $p_1 = \frac{X_1}{n_1}$ and $p_2 = \frac{X_2}{n_2}$ $(n_1 = n_2 = \frac{n}{2})$
- $T_1 = \frac{p_1 - p_2}{\sqrt{\bar{p}(1-\bar{p})(\frac{1}{n_1}+\frac{1}{n_2})}}, \bar{p} = \frac{X_1+X_2}{n_1+n_2}$
- $$\begin{cases} T_1 \overset{\Delta=0}{\sim} N(0,1) \\ T_1 \overset{\Delta=\Delta_A}{\sim} N\left(\frac{\Delta_A}{\left\{\bar{\pi}(1-\bar{\pi})(\frac{1}{n_1}+\frac{1}{n_2})\right\}^{1/2}}, \frac{\pi_1(1-\pi_1)+\pi_2(1-\pi_2)}{2\bar{\pi}(1-\bar{\pi})} \right) \bar{\pi} = \frac{\pi_1+\pi_2}{2} \end{cases}$$
- $T_2 = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}}$
- $$\begin{cases} T_2 \overset{\Delta=0}{\sim} N(0,1) \\ T_2 \overset{\Delta=\Delta_A}{\sim} N\left(\frac{\Delta_A}{\left\{\frac{\pi_1(1-\pi_1)}{n_1}+\frac{\pi_2(1-\pi_2)}{n_2}\right\}^{1/2}}, 1 \right) \end{cases}$$

6.3 Arcsin square root transformation for proportions

- $X \sim b(n, \pi)$, $p = \frac{X}{n} \implies \sin^{-1}(p^{1/2}) \sim N\left(\sin^{-1}(\pi^{1/2}), \frac{1}{4n}\right)$
- $T_3 = \frac{\sin^{-1}(p_1^{1/2}) - \sin^{-1}(p_2^{1/2})}{\left(\frac{1}{4n_1}+\frac{1}{4n_2}\right)^{1/2}}$
- $$\begin{cases} T_3 \overset{\Delta=0}{\sim} N(0,1) \\ T_3 \overset{\Delta=\Delta_A}{\sim} N\left(\frac{\sin^{-1}(\pi_1^{1/2}) - \sin^{-1}(\pi_2^{1/2})}{\left(\frac{1}{4n_1}+\frac{1}{4n_2}\right)^{1/2}}, 1 \right) \end{cases}$$

7 Comparing more than two treatments

7.1 A general result combining K estimators

- Assume that $\hat{\theta}_i \sim N(\theta_i, \sigma_i^2), i = 1, \ldots, K$
- $Z = \sum_{i=1}^{K} w_i(\hat{\theta}_i - \hat{\theta})^2,$ where $\hat{\theta} = \frac{\sum_{i=1}^{K} w_i \hat{\theta}_i}{\sum_{i=1}^{K} w_i}$ and $w_i = \frac{1}{\sigma_i^2}$
- Under $H_0 : \theta_1 = \ldots = \theta_K = \theta$, $Z \sim \chi^2_{K-1}$
- Under H_A : not all equal, $Z \sim \chi^2_{K-1,\phi^2}$ where $\phi^2 = \sum_{i=1}^{K} w_i(\theta_i - \bar{\theta})^2$ and $\bar{\theta} = \frac{\sum_{i=1}^{K} w_i \theta_i}{\sum_{i=1}^{K} w_i}$

7.2 Comparing multiple proportions using arcsine square root transformation

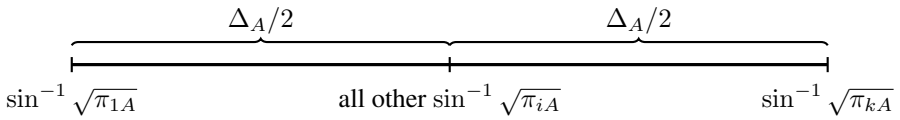
- $H_0 : \pi_1 = \pi_2 = \ldots = \pi_K$ vs H_A : not all equal
- $\sin^{-1}\sqrt{p_i} \sim N\left(\sin^{-1}\sqrt{\pi_i}, \frac{1}{4n_i}\right)$
- $T_n = \sum_{i=1}^{K} 4n_i(\sin^{-1}\sqrt{p_i} - \hat{A}_p)^2$ where $\hat{A}_p = \frac{\sum_{i=1}^{K} 4n_i \sin^{-1}\sqrt{p_i}}{\sum_{i=1}^{K} 4n_i} = \frac{\sum_{i=1}^{K} n_i \sin^{-1}\sqrt{p_i}}{\sum_{i=1}^{K} n_i}$
- Under $H_0,\ T_n \sim \chi^2_{K-1} \Rightarrow$ level $\alpha \Rightarrow T_n \geq \chi^2_{K-1,\alpha}$
- Under $H_A,\ T_n \sim \chi^2_{K-1,\phi^2}$, where

$$\phi^2 = \phi^2(n_1, \ldots, n_K; \pi_{1A}, \ldots, \pi_{KA}) = \sum_{i=1}^{K} 4n_i(\sin^{-1}\sqrt{\pi_{iA}} - \bar{A}_{\pi A})^2, \bar{A}_{\pi A} = \frac{\sum_{i=1}^{K} n_i \sin^{-1}\sqrt{\pi_{iA}}}{\sum_{i=1}^{K} n_i}. \quad (1)$$

- Desired power $1-\beta \Rightarrow \phi^2(\alpha, \beta, K-1)$ can be calculated by R $\Rightarrow$ sample size necessary
- If $n_1 = \ldots = n_K = n/K$
- Under $H_A : \pi_1 = \pi_{1A}, \ldots, \pi_K = \pi_{KA,}$ solve $\phi(\alpha, \beta, K-1) = \phi^2(\frac{n}{K}, \ldots, \frac{n}{K}; \pi_{1A}, \ldots, \pi_{KA})$ in (1) for n
- The above sample size calculation needs to determine $\pi_{1A}, \ldots, \pi_{KA}$

7.3 Choosing clinically important alternatives

- Under H_A: any of the alternatives differ from each other by Δ_A or more
- Least favorable configuration:



- Under the least favorable configuration, the smallest non-centrality parameter is $\phi^2 = \phi^2(n; \Delta_A) = \frac{4n}{K} \times \frac{\Delta_A^2}{2} = \frac{2n\Delta_A^2}{K}$
- Solve $\phi(\alpha, \beta, K-1) = \phi^2(n; \Delta_A)$ for n: $n = \frac{K\phi^2(\alpha,\beta,K-1)}{2\Delta_A^2}$
- The above sample size calculation needs to determine Δ_A

7.4 Multiple comparisons for arcsine square root transform

- $\binom{K}{2} = \frac{K(K-1)}{2}$ pairwise treatment comparisons

- $T_{nij} = \frac{2(\sin^{-1}\sqrt{p_i} - \sin^{-1}\sqrt{p_j})}{(\frac{1}{n_i}+\frac{1}{n_j})^{1/2}}$ for $i < j$ in $\{1, \ldots, K\}$
- **Bonferroni correction:** declare a pairwise treatment comparison significant if (two-sided test)
- $|T_{nij}| \geq Z_{\alpha/K(K-1)}$, as $\frac{\alpha/2}{K(K-1)/2} = \frac{\alpha}{K(K-1)}$

7.5 Chi-square tests

- $T_n = \sum_{\text{over } 2\times K \text{cells}} \frac{(O_j - E_j)^2}{E_j} \overset{H_0}{\sim} \chi^2_{K-1}$

7.6 K-sample tests for continuous response

- $\{Y_{ij} : i = 1, \ldots, n_j\} \sim N(\mu_j, \sigma^2)$ for $j = 1, \ldots, K$ (assuming equal variance)
- $\bar{Y}_j \sim N\left(\mu_j, \frac{s^2_{Y_j}}{n_j}\right), j = 1, \ldots, K$
- $T_n = \frac{\sum_{j=1}^K n_j(\bar{Y}_j - \bar{\bar{Y}})^2}{s^2_Y}$ where $\bar{\bar{Y}} = \frac{\sum_{j=1}^K n_j \bar{Y}_j}{n}$ and the pooled variance is $s^2_Y = \frac{\sum_{j=1}^K \sum_{i=1}^{n_j}(Y_{ij} - \bar{Y}_j)^2)}{n-K}$
- $H_0 : \mu_1 = \ldots = \mu_K, T_n \sim \chi^2_{K-1}$
- $H_A : \mu_1 = \mu_{1A}, \ldots, \mu_K = \mu_{KA}, T_n \sim \chi^2_{K-1,\phi^2}$, where

$$\phi^2 = \phi^2(n_1, \ldots, n_K; \mu_{1A}, \ldots, \mu_{KA}) = \frac{\sum_{j=1}^K n_j(\mu_{jA} - \bar{\mu}_A)^2}{\sigma^2_Y}, \bar{\mu}_A = \frac{\sum_{j=1}^K n_j \mu_{jA}}{n} \tag{2}$$

- Sample size calculations (assuming $n_1 = \ldots = n_K = n/K$)
- Under $H_A : \mu_1 = \mu_{1A}, \ldots, \mu_K = \mu_{KA,}$ solve $\phi^2(\alpha, \beta, K-1) = \phi^2(\frac{n}{K}, \ldots, \frac{n}{K}; \pi_{1A}, \ldots, \pi_{KA})$ in (2) for n
- The above sample size calculation needs to determine $\pi_{1A}, \ldots, \pi_{KA}$
- Under H_A: any of the alternatives differ from each other by Δ_A or more:
- Under the least favorable configuration, the smallest non-centrality parameter is $\phi^2 = \phi^2(n; \Delta_A) = \frac{n\Delta_A^2}{2K\sigma_Y^2}$
- Solve $\phi(\alpha, \beta, K-1) = \phi^2(n; \Delta_A)$ for n: $n = \frac{2K\sigma_Y^2\phi^2(\alpha,\beta,K-1)}{\Delta_A^2}$
- The above sample size calculation needs to determine Δ_A

7.7 Non-inferiority/Equivalency Trials

- Binary outcome:
- $H_0 : \pi_1 \leq \pi_2 - \Delta_A$ vs $H_A : \pi_1 > \pi_2 - \Delta_A$
- $T_n = \dfrac{p_1 - p_2 + \Delta_A}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$
- Continuous outcome:
- Similar

Chapter 8: Survival analysis - Summary

- **Parameter of interest: survival curves**
- **Kaplan Meier estimator**
- **Log-rank test: hypothesis, T, sample size calculation, design**

8 Survival Analysis

8.1 Notation

- CDF: $F(t) = P(T \leq t)$
- Survival function: $S(t) = P(T \geq t)$
- Density function: $f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt}$
- Mortality rate: $m(t) = P(t \leq T \leq t+1 | T \geq t)$
- Hazard rate: $\lambda(t) = \lim_{h \to 0} \frac{P(t \leq T \leq t+h | T \geq t)}{h}$
- Cumulative hazard function: $\Lambda(t) = \int_0^t \lambda(u) du$
- Other relations: $S(t) = \exp\left\{-\int_0^t \lambda(u) du\right\} \qquad \lambda(t) = -\frac{d \log\{S(t)\}}{dt}$

8.2 Parametric distributions

- Exponential distribution: $\lambda(t) = \lambda$, $S(t) = \exp(-\lambda t)$, Median survival time: $m = \frac{\log(2)}{\lambda}$ and $E(T) = \frac{1}{\lambda}$
- Weibull distribution: $\lambda(t) = \lambda t^{\gamma - 1}, \lambda, \gamma > 0$, $S(t) = \exp(-\frac{\lambda t^\gamma}{\gamma})$
- Gompertz-Makeham distribution: $\lambda(t) = \theta + \beta e^{\gamma t}$
- Log-normal distribution: $\log(T) \sim N(\mu, \sigma^2)$

- Gamma distribution: $f(t) \propto t^{\rho} e^{-\lambda t}, \rho > -1, \lambda > 0$

8.3 Life-table estimate $\hat{S}$

- n_r: # of individuals at risk at the beginning of the time interval
- d: # of deaths in the time interval
- w: # of individuals censored in the time interval
- $m = \frac{d}{n_r - w/2}$
- $\hat{S}(t) = \prod_{u \leq t} \{1 - m(u)\}$
- Greenwood's formula: $se\left\{\hat{S}(t)\right\} = \hat{S}(t)\left\{\sum_{j=1}^{t} \frac{d_j}{(n_{r_j} - w_j/2)(n_{r_j} - d_j - w_j/2)}\right\}^{1/2}$
- $1 - \alpha$ CI: $\hat{S}(t) \pm z_{\alpha/2} \times se\left\{\hat{S}(t)\right\}$

8.4 Kaplan-Meier (Product-limit) estimator under non-informative censoring

- T_i: the survival time if the individual was followed until death
- C_i: the censoring time in the hypothetical setting that the could not die
- $U_i = \min(T_i, C_i)$
- $\Delta_i = I(T_i \leq C_i)$

- Observed data: $(U_i, \Delta_i), i = 1, \ldots, n$
- # of individuals at risk at t: $n(t) = \sum_{i=1}^{n} I(U_i \geq t)$
- # of observed deaths at t: $d(t) = \sum_{i=1}^{n} I(U_i = t, \Delta_i = 1)$

$$KM(t) = \prod_{\text{death times } u \leq t} \left\{ 1 - \frac{d(u)}{n(u)} \right\}$$

$$se\{KM(t)\} = KM(t) \left\{ \sum_{\text{death times } u \leq t} \frac{d(u)}{n(u)\{n(u) - d(u)\}} \right\}^{1/2}, \tag{3}$$

where (3) follows by Greenwood's formula.

8.5 Two-sample Log Rank Test

- Treatment-specific survival functions: $S_1(t) = P(T \geq t | A = 1),\ S_0(t) = P(T \geq t | A = 0)$
- Treatment-specific hazard ratios: $\lambda_1(t)$ and $\lambda_0(t)$
- $H_0 : S_1(t) = S_0(t) = S(t), \forall t > 0$ (or $H_0 : \lambda_1(t) = \lambda_0(t) = \lambda(t), \forall t > 0$
- $H_A : S_1(t) \geq S_0(t), \forall t > 0$ with strict inequality for at least one t
- Log rank test:

- Given $(U_i, \Delta_i, A_i) : i = 1, \ldots, n$, for $j = 0, 1$, define

$$n_j = \sum_{i=1}^{n} I(A_i = j), \qquad n = n_0 + n_1$$

$$n_j(u) = \sum_{i=1}^{n} I(U_i \geq u, A_i = j), \qquad n(u) = n_0(u) + n_1(u)$$

$$d_j(u) = \sum_{i=1}^{n} I(U_i = u, \Delta_i = 1, A_i = j), \qquad d_j(u) = d_0(u) + d_1(u)$$

- Log rank statistic

$$T_n = \frac{\sum_{\text{all death times } u} \left\{ d_1(u) - \frac{n_1(u)}{n(u)} d(u) \right\}}{\left[\sum_{\text{all death times } u} \frac{n_1(u) n_0(u) d(u) \{n(u) - d(u)\}}{n^2(u) \{n(u) - 1\}} \right]^{1/2}},$$

  where $d_1(u)$ is the observed # of deaths from treatment 1, and $\frac{n_1(u)}{n(u)} d(u)$ is expected # of deaths from treatment 1 under H_0.

- Under $H_0 : T_n \sim N(0, 1) \implies \begin{cases} \text{two-sided level } \alpha \Rightarrow \text{rejection region: } |T_n| \geq Z_{\alpha/2} \\ \text{one-sided level } \alpha \Rightarrow \text{rejection region: } T_n \geq Z_\alpha \text{ or } -T_n \geq Z_\alpha \end{cases}$

- The log rank test is non-parametric.

	Treatments				
	1	2	...	K	Total
No. of deaths	$d_1(u)$	$d_2(u)$	...	$d_K(u)$	$d(u)$
No. alive	$n_1(u) - d_1(u)$	$n_2(u) - d_2(u)$	...	$n_K(u) - d_K(u)$	$n(u) - d(u)$
No. at risk	$n_1(u)$	$n_2(u)$	...	$n_K(u)$	$n(u)$

8.6 Power and sample size

- $H_A : \frac{\lambda_1(t)}{\lambda_0(t)} = \exp(\gamma_A), t \geq 0$
- $T_n \overset{H_A}{\sim} N\left(\{d\theta(1-\theta)\}^{1/2}\gamma_A, 1\right)$, where d is total # of deaths, and θ is proportion randomized to treatment 1.
- For a two-sided level α test having power $1-\beta$ to detect H_A and equal allocation, we have

$$\gamma_A d^{1/2}/2 = Z_{\alpha/2} + Z_\beta \Rightarrow d = \frac{4(Z_{\alpha/2} + Z_\beta)^2}{\gamma_A^2}$$

- Design specifications: Accrual rate $a(u) = a$ overall, or $a/2$ for each treatment
- $F_j(u) = P(T \leq u | A = j) = 1 - S_j(u), j = 0, 1$
- $d_j = \int_0^{Acc} \frac{a(u)}{2} F_j(L-u)du, j = 0, 1, d_j = d_j(Acc, L)$
- $d_1(Acc, L) + d_0(Acc, L) = \frac{4(Z_{\alpha/2}+Z_\beta)^2}{\gamma_A^2}$

8.7 K-sample Tests

- Given $(U_i, \Delta_i, A_i), i = 1, \ldots, n, A_i \in \{1, \ldots, K\}$
- $H_0 : S_1(t) = \ldots = S_K(t), t \geq 0$ (or $H_0 : \lambda_0(t) = \ldots = \lambda_K(t), t \geq 0$)
- Define

$$\tilde{T}_n = \begin{pmatrix} \sum_{\text{death times } u} \left\{ d_1(u) - \frac{n_1(u)}{n(u)} d(u) \right\} \\ \ldots \\ \sum_{\text{death times } u} \left\{ d_{K-1}(u) - \frac{n_{K-1}(u)}{n(u)} d(u) \right\} \end{pmatrix}$$

- $\tilde{V}_n = (v_{nij})$, where

$$v_{njj} = \sum_u \frac{d(u)\{n(u) - d(u)\} n_j(u)\{n(u) - n_j(u)\}}{n^2(u)\{n(u) - 1\}}$$
$$v_{njj'} = -\sum_u \frac{d(u)\{n(u) - d(u)\} n_j(u) n_{j'}(u)}{n^2(u)\{n(u) - 1\}}, j \neq j'$$

- K-sample log-rank test statistics:

$$T_n = \tilde{T}_n^T \tilde{V}_n^{-1} \tilde{T}_n \overset{H_0}{\sim} \chi^2_{K-1}$$

- Sample size for K-sample log-rank test:

- Assume equal sample sizes.

- Scenario 1: $H_A : \lambda_1(t) = \lambda_{1A}, \ldots, \lambda_K(t) = \lambda_{KA}$, then

$$T_n \overset{H_A}{\sim} \chi^2_{\phi^2, K-1},$$

  where

$$\phi^2 = \phi^2(d; \lambda_{1A}, \ldots, \lambda_{KA}) = \frac{d}{K} \sum_{j=1}^{K} \left\{ \log(\lambda_{jA}) - \frac{\sum_{j=1}^{K} \log(\lambda_{jA}\}}{K} \right\}^2$$

- Equaling $\phi^2(d; \lambda_{1A}, \ldots, \lambda_{KA}) = \phi^2(\alpha, \beta, K-1)$ leads to

$$d = \frac{K\phi^2(\alpha, \beta, K-1)}{\sum_{j=1}^{K} \left\{ \log(\lambda_{jA}) - \frac{\sum_{j=1}^{K} \log(\lambda_{jA}\}}{K} \right\}^2}$$

- Scenario 2: $H_A :$ hazard ratio between any two treatments is $\geq \exp(\lambda_A)$

- $\lambda_A = \log\left(\frac{\lambda_{\max,A}}{\lambda_{\min,A}}\right) = \log\left(\lambda_{\max,A}\right) - \log\left(\lambda_{\min,A}\right)$

- The least favorable configuration leads to

$$\phi^2 = \phi^2(d; \lambda_A) = \frac{d}{K} \frac{\lambda_A^2}{2}$$

- Equaling $\phi^2(d; \lambda_A) = \phi^2(\alpha, \beta, K-1)$ leads to

$$d = \frac{2K\phi^2(\alpha, \beta, K-1)}{\gamma_A^2}$$

Example: Binary outcome and survival outcome

- A 6-year study recruits n patients right after the trial starts. The patients are randomized to two treatments (0 vs 1).
- The median survival for patients under treatment 0/1 is $4/5$ years and the survival distribution follows an exponential distribution.
- You are asked to design a two-arm randomized clinical trial with equal allocation to both arms (the new treatment and the control treatment) that will enable you to detect the clinically important difference given above with 90% power using a test at the $.05$ (one-sided) level of significance.
- If use the survival outcome: Where is the hypothesis? Test statistics?

How to design the trial? No accrual. So, sample size?

- **If use a binary outcome: Y=indicator of survival at the end of the study. What is the hypothesis? Test statistics? How to design the trial; that is sample size calculation?**

Chapter 9: Early Stopping of Clinical Trials - Summary

- **Group sequential test**
- **Different boundaries and properties**
- **Information based monitoring**

9 Early Stopping of Clinical Trials

9.1 General issues in monitoring clinical trials

9.2 Information based design and monitoring

- Two-sided: $H_0 : \Delta = 0$ vs $H_A : \Delta \neq 0$ (One-sided: $H_0 : \Delta \leq 0$ vs $H_A : \Delta > 0$)
- K-look group sequential test: $T(t) = \frac{\hat{\Delta}(t)}{se\{\hat{\Delta}(t)\}}$, for $t = t_1, \ldots, t_K$
- Efficient-based test

$$\begin{aligned} T(t) &= \frac{\hat{\Delta}(t)}{se\{\hat{\Delta}(t)\}} \overset{\Delta=\Delta^*}{\sim} N(\Delta^* I^{1/2}(t, \Delta^*), 1) \\ & \quad I(t, \Delta^*) \approx \left[se\{\hat{\Delta}(t)\} \right]^{-2} \end{aligned}$$

- Group sequential test strategy: reject H_0 if $|T(t)| \geq b(t)$ for the first time t.

9.3 Type I error

- For efficient based test,

$$\begin{pmatrix} T(t_1) \\ \vdots \\ T(t_K) \end{pmatrix} \overset{\Delta=\Delta^*}{\sim} N\left(\begin{pmatrix} \Delta^* I^{1/2}(t_1, \Delta^*) \\ \vdots \\ \Delta^* I^{1/2}(t_K, \Delta^*) \end{pmatrix}, V_T = \left(\sqrt{\frac{I(t_j, \Delta^*)}{I(t_l, \Delta^*)}}, j \leq l \right) \right)$$

- For equal increments of information, $I(t_1, \Delta^*) = I,\ I(t_j, \Delta^*) = j \times I,\ \ldots,\ I(t_K, \Delta^*) = K \times I$,

$$\begin{pmatrix} T(t_1) \\ \vdots \\ T(t_K) \end{pmatrix} \overset{\Delta=\Delta^*}{\sim} N\left(\begin{pmatrix} \Delta^* I^{1/2} \times \sqrt{1} \\ \vdots \\ \Delta^* I^{1/2} \times \sqrt{K} \end{pmatrix}, V_T = \left(\sqrt{\frac{j}{l}}, j \leq l \right) \right)$$

- α-level:

$$P_{\Delta=0}\{|T(t_1)| \leq b(t_1), \ldots, |T(t_K)| \leq b(t_K)\} = 1 - \alpha, \tag{4}$$

9.4 Choice of boundaries

- Let $b_j = b(t_j) = c \cdot j^{\Phi - 0.5}$. Solving (4) for $c \implies c = c(\alpha, K, \Phi)$.
- Pocock boundary: $\Phi = 0.5$: stringent
- O'Brien-Fleming boundary: $\Phi = 0$: conservative

9.5 Power and sample size in terms of statistical information

- Under $H_A : \Delta = \Delta_A$:

$$T(t) \overset{\Delta=\Delta_A}{\sim} N(\Delta_A I^{1/2}(t, \Delta_A), 1)$$

- Set

$$\Delta_A I^{1/2}(t^F, \Delta_A) = Z_{\alpha/2} + Z_\beta,$$

which implies that approximately

$$se\{\hat{\Delta}(t^F)\} = \left(\frac{Z_{\alpha/2} + Z_\beta}{\Delta_A}\right)^2.$$

- Define MI to be the **maximum information** at the final analysis

- K-look group sequential test: information at t_j is $I(t_j, \Delta_A) = \frac{j}{K} \times MI$. Then,

$$\begin{pmatrix} T(t_1) \\ \vdots \\ T(t_K) \end{pmatrix} \overset{\Delta=\Delta_A}{\sim} N\left(\begin{pmatrix} \delta\sqrt{\frac{1}{K}} \\ \vdots \\ \delta\sqrt{\frac{K}{K}} \end{pmatrix}, V_T = \left(\sqrt{\frac{j}{l}}, j \leq l\right)\right)$$

  where $\delta = \Delta_A\sqrt{MI}$.

- Power is $1 - P_\delta\left[\bigcup_{j=1}^{K}\left\{|T(t_j) < c(\alpha, K, \Phi)j^{\Phi-0.5}\right\}\right] = 1 - \beta$. The solution is $\delta(\alpha, K, \Phi, \beta)$.

- Set $\delta(\alpha, k, \Phi, \beta) = \Delta_A\sqrt{MI}$. The required maximum information becomes $MI = \left\{\frac{\delta(\alpha,k,\Phi,\beta)}{\Delta_A}\right\}^2$.

- Fixed sample design requires $I^{FS} = \left(\frac{Z_{\alpha/2}+Z_\beta}{\Delta_A}\right)^2$.

- K-look group sequential test requires

$$MI = I^{FS} \times \left\{\frac{\delta(\alpha, K, \Phi, \beta)}{Z_{\alpha/2} + Z_\beta}\right\}^2 := I^{FS} \times IF(\alpha, K, \Phi, \beta).$$

Define $IF(\alpha, k, \Phi, \beta) = \left\{ \frac{\delta(\alpha,K,\Phi,\beta)}{Z_{\alpha/2}+Z_\beta} \right\}^2$ as inflation factor. (This is how the fixed sample information translates to the maximum information for a K-look group sequential test.)

- **Average information**

- Let $\{V = j : \text{ stopping at } t_j\}$, then

$$E_{\Delta^*}(V) = \sum_{j=1}^{K} j \times P_{\Delta^*}(V = j)$$

$$AI(\Delta^*) = \frac{MI}{K} E_{\Delta^*}(V) = I^{FS} \left\{ \frac{IF(\alpha, K, \Phi, \beta)}{K} \right\} E_{\Delta^*}(V)$$

- Pocock designs require smaller sample sizes on average if H_A is true, but require larger sample sizes on average if H_0 is true.

Chapter 10: Integrative analysis of Trial and Real-word evidence studies

- **Pros and cons of different studies**