2024 Summer Institute In Statistics for Clinical & Epidemiological Research

Module 3:

Design, Conduct, and Analysis of Randomized Clinical Trials with Time to Event Primary Endpoints

Lecture 18:

Sample Size Formula

Scott S. Emerson, M.D., Ph.D.
Professor Emeritus of Biostatistics
University of Washington

1

Sample Size Formula

Common Settings

Where am I going?

- The most common RCT designs can all use the same sample size formulat

2

2

Sample Size Calculation

- Traditional approach
  - Sample size to provide high power to “detect” a particular alternative
- Decision theoretic approach
  - Sample size to discriminate between hypotheses
    - “Discriminate” based on interval estimate
    - Standard for interval estimate: 95%
      - Equivalent to traditional approach with 97.5% power

3

3

Issues

- Summary measure
  - Mean, geometric mean, median, proportion, hazard…
- Structure of trial
  - One arm, two arms, k arms
  - Independent groups vs cross over
  - Cluster vs individual randomization
  - Randomization ratio
- Statistic
  - Parametric, semi-parametric, nonparametric
  - Adjustment for covariates

4

4

Measures of Precision

- Estimators are less variable across studies
  - Standard errors are smaller
- Estimators typical of fewer hypotheses
  - Confidence intervals are narrower
- Able to statistically reject false hypotheses
  - Z statistic is higher under alternatives

5

5

Criteria for Precision

- Standard error
- Width of confidence interval
- Statistical power
  - Probability of rejecting the null hypothesis
    - Select "design alternative"
    - Select desired power

6

6

Statistics to Address Variability

- At the end of the study:
  - Frequentist and/or Bayesian data analysis to assess the credibility of clinical trial results
    - Estimate of the treatment effect
      - Single best estimate
      - Precision of estimates
    - Decision for or against hypotheses
      - Binary decision
      - Quantification of strength of evidence

7

7

Sample Size Determination

- Based on sampling plan, statistical analysis plan, and estimates of variability, compute
  - Sample size that discriminates hypotheses with desired power, or
  - Hypothesis that is discriminated from null with desired power when sample size is as specified, or
  - Power to detect the specific alternative when sample size is as specified

8

8

Sample Size Computation

Standardized level α test (n = 1): $\delta_{\alpha\beta}$ detected with power β

Level of significance α when $\theta = \theta_0$

Design alternative $\theta = \theta_1$

Variability V within 1 sampling unit

Required sampling units: $$n = \frac{(\delta_{\alpha\beta})^2 V}{(\theta_1 - \theta_0)^2}$$

(Fixed sample test: $\delta_{\alpha\beta} = z_{1-\alpha/2} + z_{\beta}$)

9

9

When Sample Size Constrained

- Often (usually?) logistical constraints impose a maximal sample size
  - Compute power to detect specified alternative

Find β such that $$\delta_{\alpha\beta} = \sqrt{\frac{n}{V}}(\theta_1 - \theta_0)$$

  - Compute alternative detected with high power

$$\theta_1 = \theta_0 + \delta_{\alpha\beta}\sqrt{\frac{V}{n}}$$

10

10

Std Errors: Key to Precision

- Greater precision is achieved with smaller standard errors
- In a fixed sample study

Typically: $se\left(\hat{\theta}\right) = \sqrt{\frac{V}{n}}$

(V related to average "statistical information")

Width of CI: $2 \times \left(crit\ val\right) \times se\left(\hat{\theta}\right)$

Test statistic: $Z = \frac{\hat{\theta} - \theta_0}{se\left(\hat{\theta}\right)}$

11

11

Ex: One Sample Mean

- Data $Y_i \sim (\mu, \sigma^2),\ i = 1, \ldots, n \quad i.i.d$
- Summary measure $\theta = \mu$
- Estimator $\hat{\theta} = \bar{Y}$
- Sampling unit variability $V = \sigma^2$
- Standard error $se(\hat{\theta}) = \sqrt{V/n} = \sqrt{\frac{\sigma^2}{n}}$

12

12

Ex: Difference of Independent Means

- Data $Y_{ij} \sim \left(\mu_i, \sigma_i^2\right),\ i = 0{,}1, j = 1 \dots, n_j$ ind.
  $n = n_1 + n_0,\quad r = n_1/n_0$
- Summary measure $\theta = \mu_1 - \mu_0$
- Estimator $\hat{\theta} = \overline{Y}_1 - \overline{Y}_0$
- Sampling unit variability $V = (r+1)\left[\frac{1}{r}\sigma_1^2 + \sigma_0^2\right]$
- Standard error $se\left(\hat{\theta}\right) = \sqrt{V/n} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_0^2}{n_0}}$

13

13

Ex: Hazard Ratio

- Rt cens time to event $\left(Y_{ij}, \delta_{ij}\right) = 0{,}1,\ j = 1 \dots, n_j$ $indep$
  $n = n_1 + n_0,\quad r = n_1/n_0$
- Summary measure $\theta = \frac{h_1(t)}{h_2(t)}$ $hazard\ ratio$
- Estimator $\hat{\theta} = e^{\hat{\beta}}$ $\hat{\beta}\ from\ Cox\ PH$
- Sampling unit variability $V = \frac{(r+1)^2}{r\, Pr[\delta=1]}$
- Standard error $se\left(\log \hat{\theta}\right) = \sqrt{\frac{V}{n}} = \sqrt{\frac{(r+1)^2}{rD}}$ $D = \sum_i \sum_j \delta_{ij}$

14

14

Ex: Hazard Ratio

- Sample size formula will provide the number of events
- Accrual model to estimate the number of subjects:
  - Exponential survival times $T_{ij}^0 \sim \mathcal{E}(\lambda_i)$
  - Median survival in control group m_0 leading to hazard $\lambda_0 = \frac{\log 2}{m_0}$
  - Hypothesized hazard ratio $\theta = \frac{\lambda_1}{\lambda_0} \quad \Rightarrow \quad \lambda_1 = \theta\lambda_0$
  - Accrue subjects uniformly from time 0 to time a
  - Follow subjects up to time $\tau > a$
- Estimate probability of an observed event in combined sample

$$Pr\left[\delta_{ij} = 1\right] = 1 - \frac{\exp\{-\lambda_i(\tau - a)\}}{\lambda_i a} + \frac{\exp\{-\lambda_i \tau\}}{\lambda_i a}$$

- Accrue

$$n = n_0 + n_1 \qquad n_0 = \frac{D}{rPr\left[\delta_{1j}=1\right] + Pr\left[\delta_{0j}=1\right]} \qquad n_1 = rn_0$$

15

15

Modifications for Hazard Ratio

- Hypothesized hazard ratio might be
  - Under the null: 1.0
  - Under an alternative presuming adjusting for prognostic covariates
- Often sponsors will use simulations to explore
  - Non uniform accrual
  - Weibull time to event (or some other distribution)
  - Impact of intercurrent events
    - Competing risks
    - Withdrawal of consent
  - Attrition due to loss to follow-up

16

16

Ex: KM Estimate of Survival Probability

- If we wanted to use a Kaplan-Meier estimate of $S(\tau)$ as a primary endpoint, we can use the asymptotic normality of $\hat{S}(\tau)$
- The variance is computed using Greenwood's formula
  - Hazard estimate is a proportion: D_j / N_j
  - Variance of hazard estimate from theory about binomial proportions
  - Delta method to get variance of log (1 - D_j / N_j)
  - Then use properties of expectation to get variance of log S(t) = Σ log (1 - D_j / N_j)
    - Noninformative censoring leads to asymptotically uncorrelated hazard estimates
  - Use delta method to get variance of S(t)
  - Standard error is square root of variance of S(t)

17

17

Survival Probability Estimates

- Maximum likelihood estimates for
  - Conditional survival probability within intervals

$$\hat{p}_{j|j-1} = 1 - \frac{d_j}{n_j}$$

  - Unconditional survival probability

$$\hat{\pi}_{\mathrm{j}} = \prod_{i=1}^{j} \hat{p}_{i|i-1}$$

18

Notational convenience

For $0 = t_0 \le t_1 \le t_2 \le \cdots \le t_k$

$$n_j = \text{Number at risk in interval } \left(t_{j-1}, t_j\right]$$
$$d_j = \text{Number failed in interval } \left(t_{j-1}, t_j\right]$$

For $0 = t_0 \le t_1 \le t_2 \le \cdots \le t_k$

$$\begin{aligned}\pi_j &= \Pr\left(T > t_j\right) = \Pr\left(T > t_j \cap T > t_{j-1}\right) \\ &= \Pr\left(T > t_j \mid T > t_{j-1}\right)\Pr\left(T > t_{j-1}\right) \\ &= p_{j|j-1}\ \pi_{j-1} \\ &= p_{j|j-1} \times p_{j-1|j-2} \times \cdots \times p_{1|0} \\ &= \prod_{i=1}^{j} p_{i|i-1}\end{aligned}$$

19

Logarithmic Transformation

- Sums are easier to work with than products
  - The log transformed unconditional survival probability is the sum of log transformed conditional survival probabilities

$$\log\left(\pi_j\right) = \sum_{i=1}^{j} \log\left(p_{i|i-1}\right)$$

$$\log\left(\hat{\pi}_j\right) = \sum_{i=1}^{j} \log\left(\hat{p}_{i|i-1}\right)$$

20

Basic Approach

- We will find the standard error of the log transformed survival probabilities by
  - Estimating each conditional survival probability and finding the variance of the log transformed estimates
  - Invoking noninformative censoring to argue that the sum of our log transformed estimates must have the same distribution as the sum of log transformed independent estimates

21

Standard Error of Proportions

- From the laws of expectation, for the *j*th interval

$$\text{For } Y_1, Y_2, \ldots, Y_{nj} \overset{iid}{\sim} Bernoulli\left(p_{j|j-1}\right)$$

$$E\left[\overline{Y}\right] = p_{j|j-1}$$

$$Var\left(\overline{Y}\right) = \frac{p_{j|j-1}(1 - p_{j|j-1})}{n_j}$$

$$se\left(\overline{Y}\right) = \sqrt{\frac{p_{j|j-1}(1 - p_{j|j-1})}{n_j}}$$

22

Large Sample Approximation

- From the central limit theorem

$$\text{For } Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} Bernoulli(p)$$

$$\hat{p}_{j|j-1} = \overline{Y} \,\dot{\sim}\, N\left(p_{j|j-1}, \frac{p_{j|j-1}(1-p_{j|j-1})}{n_j} \right)$$

23

Logarithmic Transformation

- From the delta method

$$\text{For } \overline{Y} \,\dot{\sim}\, N\left(\theta, \frac{V}{n}\right)$$

$$g\left(\overline{Y}\right) \dot{\sim}\, N\left(g(\theta), [g'(\theta)]^2 \frac{V}{n} \right)$$

$$\text{So, for } Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} Bernoulli\left(p_{j|j-1}\right)$$

$$\log\left(\hat{p}_{j|j-1}\right) \dot{\sim}\, N\left(\log\left(p_{j|j-1}\right), \frac{(1-p_{j|j-1})}{n_j\, p_{j|j-1}} \right)$$

24

Noninformative Censoring

- In the presence of noninformative censoring, the risk set in any interval should look like a random sample of the population at risk
- Estimates of the conditional probability of survival for the intervals should be uncorrelated

$$\log\left(\hat{\pi}_j\right) = \sum_{i=1}^{j} \log\left(\hat{p}_{i|i-1}\right) \dot{\sim} N\left(\sum_{i=1}^{j} \log\left(p_{i|i-1}\right), \sum_{i=1}^{j} \frac{(1-p_{i|i-1})}{n_i p_{i|i-1}}\right)$$

25

Confidence Intervals

- Using the large sample approximation with plug-in estimates for standard errors

$100(1-\alpha)\%$ CI for $\log\left(\pi_j\right)$

$$\sum_{i=1}^{j} \log\left(\hat{p}_{i|i-1}\right) \pm z_{1-\alpha/2} \sqrt{\sum_{i=1}^{j} \frac{(1-\hat{p}_{i|i-1})}{n_i p_{i|i-1}}}$$

$100(1-\alpha)\%$ CI for π_j

$$\exp\left\{\sum_{i=1}^{j} \log\left(\hat{p}_{i|i-1}\right) \pm z_{1-\alpha/2} \sqrt{\sum_{i=1}^{j} \frac{(1-\hat{p}_{i|i-1})}{n_i \hat{p}_{i|i-1}}}\right\}$$

26

Greenwood's Formula

- SE for the survival probabilities by a second application of the delta method

$$\log\left(\hat{\pi}_j\right) = \sum_{i=1}^{j} \log\left(\hat{p}_{i|i-1}\right) \dot{\sim} N\left(\sum_{i=1}^{j} \log\left(p_{i|i-1}\right), \sum_{i=1}^{j} \frac{(1-p_{i|i-1})}{n_i p_{i|i-1}} \right)$$

$$\hat{\pi}_j = \exp\left(\sum_{i=1}^{j} \log\left(\hat{p}_{i|i-1}\right) \right) \dot{\sim} N\left(\pi_j, \pi_j^2 \sum_{i=1}^{j} \frac{(1-p_{i|i-1})}{n_i p_{i|i-1}} \right)$$

$100(1-\alpha)\%$ CI for π_j

$$\hat{\pi}_j \pm z_{1-\alpha/2} \times \hat{\pi}_j \sqrt{\sum_{i=1}^{j} \frac{(1-p_{i|i-1})}{n_i p_{i|i-1}}}$$

27

Ex: KM Estimate of Survival Probability

- We see that the standard error of our estimator will depend heavily on
  - the distribution of the time to event, and
  - the censoring distribution
- We could take an approach based on, say, an exponential distribution as we did for HR based inference
  - This would likely not differ much from an approach using the parametric estimator
  - It would likely not be too reliable in general
- Similar issues would arise when trying to use the restricted mean as the primary endpoint

28

28

Options to Increase Precision

- Increase sample size
- Decrease V
  - With HR: adjust for important prognostic variables
  - Using anecdotal simulations shown in Lecture 17:
    - A prognostic variable having HR_W= 1.82 per 1 SD_W of its distn
      - Unadjusted HR_X = 0.638 corresponds to adjusted HR_X= 0.607
      - Adjustment increase power from 61% to 70%
      - Equivalent to 23% increase in number of events
    - A prognostic variable having HR_W= 2.46 per 1 SD_W of its distn
      - Unadjusted HR_X = 0.668 corresponds to adjusted HR_X= 0.607
      - Adjustment increase power from 52% to 70%
      - Equivalent to 55% increase in number of events
- (Decrease confidence level)

29

29

Subgroups

- Testing for effects in K subgroups
  - Does the treatment work in each subgroup?
  - Bonferroni correction: Test at α / K
    - No subgroups: N = 100
    - Two subgroups: N = 230
- Testing for interactions across subgroups
  - Does the treatment work differently in subgroups?
    - Two subgroups: N = 400

30

30