BNI Biostatistics BootCamp Lecture 4:

# Selection of Appropriate Statistical Methods in Protocol Development

**August 31, 2023**


**Wonsuk Yoo, PhD**

Dept. of Translational Neuroscience

wonsuk.yoo@barrowneuro.org

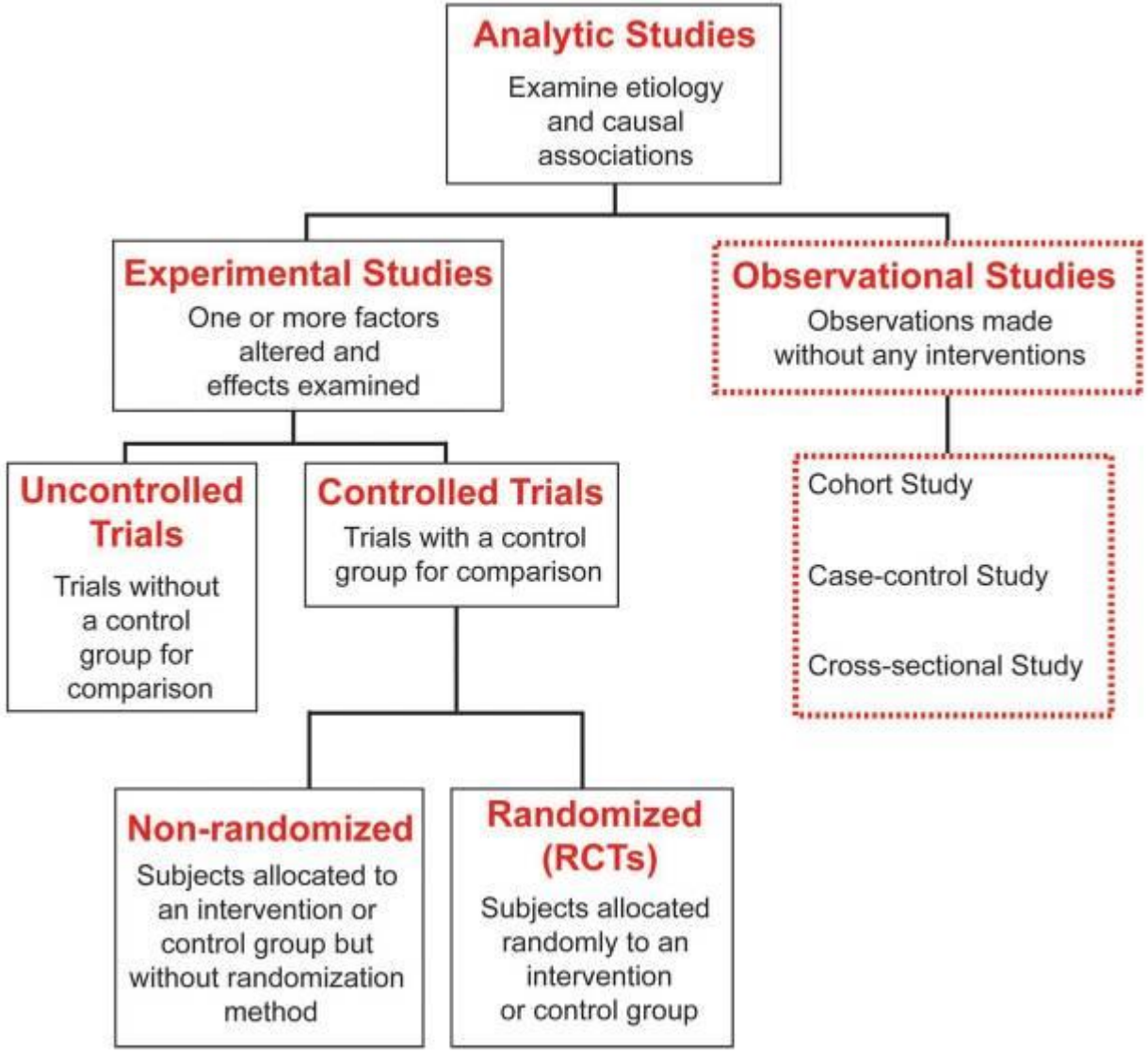

Ivy Brain Tumor Center | Barrow Neurological Institute

# Quick review of last three Lectures:

- Lecture 1: Statistical uncertainty and hypothesis testing
  - ➡ Conditions of good study design related to statistical uncertainty
- Lecture 2: Statistical considerations in protocol development – study design
  - ➡ Design components to control type 1 error rate
- Lecture 3: Understanding of the statistical power and sample size in protocol development.
  - ➡ Design components to maximize the study power

# Analytic study designs:

Source:
Song JW, Chung KC. Observational studies: cohort and case-control studies. Plast Reconstr Surg. 2010 Dec;126(6):2234-2242. doi: 10.1097/PRS.0b013e3181f44abc. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2998589/



# Key Elements of Statistical Methods:

- The type of data analysis: descriptive or inferential purposes?

  ➡ Descriptive statistics or inferential statistics. Go to slide 5.

- The research type: Randomized trials or Non-randomized trials?

  ➡ go to next slide 6.

- The type of outcome variables: Do the primary endpoints be continuous or categorical or even survival data?

  ➡ go to next slide 7.

- The size of study data: more than 30 or less than 30 (small)?

  ➡ go to next slide 8.

# Descriptive Statistics:

- There are two data analysis purposes: descriptive and analytic.
- Descriptive statistics is about description of how to characterize demographic, baseline and study data. How to summarize?
  - ➡ Numerical data (normal data): Mean ± SD or Mean(SD)
  - ➡ Numerical data (non-normal data): Median(IQR) or Median(25%-75%)
  - ➡ Categorical data: frequencies (proportions/percentages) – N(%)
- Normality test: Why is it important to check outcome data normality?
  - ➡ Graphical approach: Q-Q plot
  - ➡ Analytical method: Shapiro-Wilk test… problem or issue?

# Studies in Neuroscience Research

| Randomized Experimental Studies: | Non-randomized Observational Studies: |
| --- | --- |
| ● Random allocation of patients to treatment or control group ➡ Averaging the other factors between two groups<br>● Affords researchers ability to draw unbiased conclusions about treatment effect. | ● Allocation of patients to treatment or control group not random<br>● Results might be biased due to pre-existing covariate characteristics.<br>● Usually work with EMR, chart review, most public-accessed database |
| Q. What is quasi-randomized trials? And when do we use it? | |

# The Types of Outcomes and Predictors:

| | | OUTCOME | | |
|---|---|---|---|---|
| | | Categorical | Continuous | Survival |
| **INPUT** | Categorical (N=2) | Chi-square test | Student's t-test<br>Wilcoxon rank sum test | Log-rank test<br>Cox regression |
| | Categorical (N>2) | Chi-square test | ANOVA<br>Kruskal-Wallis test | Log-rank test<br>Cox regression |
| | Continuous | Logistic regression | Correlation analysis<br>Linear regression | Cox regression |

# Parametric .vs. Non-parametric Methods

- What happen for small data (less than 30 samples0 studies?
- What is the core concept of non-parametric approach?
- Nonparametric tests do not assume a particular form of the distribution as long as it is continuous. These tests compare the medians of the groups instead of the means and they use the rank of the observations instead of the actual observed values in the calculation of the test statistic.

| Parametric Methods | Non-parametric Methods |
|---|---|
| Paired t-test | Wilcoxon signed rank test |
| Two sample independent t-test | Wilcoxon rank sum test |
| ANOVA | Kruskal-Wallis test |

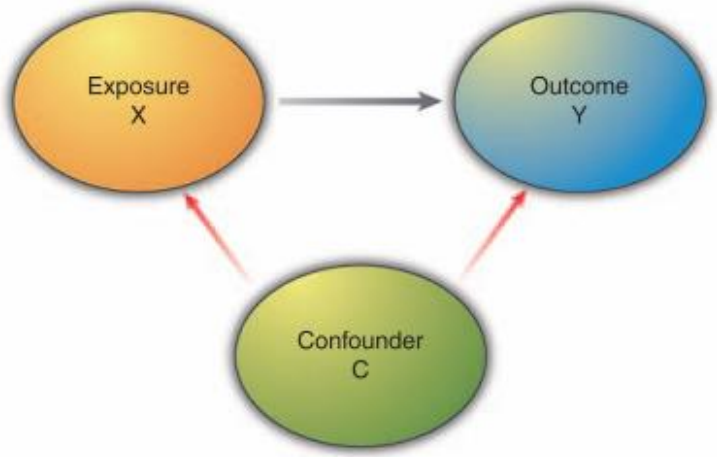
# Observational Studies are …

- In observational data exist pre-determined existing conditions.
- That is, observational studies are not able for investigator to randomly allocate patients into treatment or control group, which could result in…

➡ Potential of selection bias when we get a sample

➡ There exist large variations in pre-treatment characteristics between treatment and control groups. We call these confounding factors…

➡ True effect of treatment may be over-estimated.



# Result of Confounding Effects?

- What is a confounding variable?
  - ➡ A third variable that influences both outcome and predictor variables.
- Why does a confounding variable be important?
  - ➡ You may find a cause-and-effect relationship that does not actually exist. Why? because the effect you measure is caused by the confounding variable and not by your independent variable.
- What is the result of confounding variables? BIAS…
- What should we do with non-randomized studies using observational data?
  - ➡ We need to reduce the impact of confounding variables. W

# How to Reduce Confounding Effects?

- Current techniques to control selection bias include epidemiologic and statistical methods: Matching and Regression methods.
- **Matching method**: An epidemiologic method to match control patients and experimental patients with similar pre-treatment characteristics.
- **Regression method**: A statistical method to adjusts for pre-treatment characteristics (covariates) by incorporating them into the model.
- What are limitations with current methods?
  - ➡ These methods cannot control for large number of confounders/covariates.
  - ➡ The propensity scores method

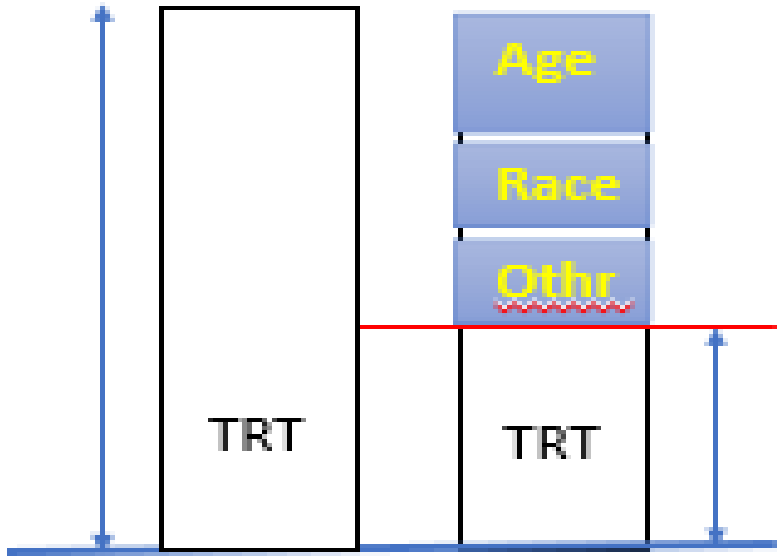
# Regression modeling strategy

- Three types of outcome values:  continuous, categorical, and survival
- Multiple regression model for the **continuous** outcome.
- Logistic regression model for the **dichotomous** outcome
- Cox regression model for the **survival** outcome
- The multivariable model looks like

$$Y = f(t, c_i|data) = t + c_1 + c_2 + \cdots + c_k$$

# Regression approach to reduce confounding effects in observational studies?

- Pre-existing conditions exits in observational database ➡ Bias...
- How can we identify the confounders information?
- How does the confounding effects be reduced with regression mode?

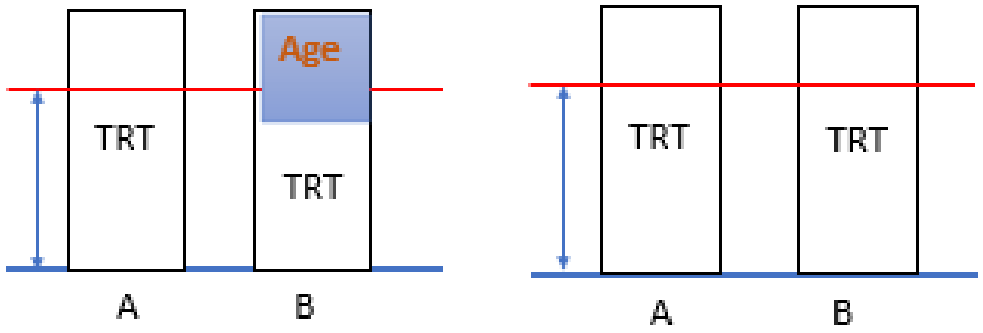


# Should we consider a multivariable model in randomized trials?

- A comparative study without randomization:

  ➡ might lead to a bias

- Why do we perform a randomized trial?




- In randomized trials, do we need to develop a multivariable model?

  ➡ if necessary … if so, when necessary?

**Table 1. Demographic Characteristics of the Patients at Baseline.**

| Characteristic | Radiotherapy (N=286) | Radiotherapy plus Temozolomide (N=287) |
|---|---|---|
| Age — yr | | |
| Median | 57 | 56 |
| Range | 23–71 | 19–70 |
| Age — no. (%)* | | |
| <50 yr | 81 (28) | 90 (31) |
| ≥50 yr | 205 (72) | 197 (69) |
| Sex — no. (%) | | |
| Male | 175 (61) | 185 (64) |
| Female | 111 (39) | 102 (36) |
| WHO performance status — no. (%)*† | | |
| 0 | 110 (38) | 113 (39) |
| 1 | 141 (49) | 136 (47) |
| 2 | 35 (12) | 38 (13) |
| Extent of surgery — no. (%)* | | |
| Biopsy | 45 (16) | 48 (17) |
| Debulking | 241 (84) | 239 (83) |
| Complete resection | 113 (40) | 113 (39) |
| Partial resection | 128 (45) | 126 (44) |

- **Stupp** et al. Radiotherapy plus Concomitant and Adjuvant **Temozolomide** for **Glioblastoma**. NEJM 2005;352: 987-996.
- Look at demographic and baseline characteristics for both groups. What do you learn from the table?
- The primary endpoint of “overall survival” was compared between two groups: **p-value < 0.0001**. What is your conclusion?
- Let’s assume the younger age (<50) of treatment was 50% rather than current 31%. Can we ensure that the result of this study is correct? If not, why?

# ANOVA (I):

Suppose 12 recent college graduates are assigned three groups with 4 subjects: a control (I), a drug treatment (II), a combination of drug and exercise (II). Age, pulse rates, systolic blood pressure (SBP), and LDL cholesterol levels were measured after 8 weeks. Explain whether there are differences among three groups with respect to above four factors.

DATA:

| Age | | | PR | | | SBP | | | LDL | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| I | II | III | I | II | III | I | II | III | I | II | III |
| 22 | 22 | 22 | 64 | 59 | 72 | 74 | 81 | 90 | 85 | 72 | 78 |
| 22 | 22 | 22 | 63 | 58 | 71 | 76 | 80 | 89 | 101 | 130 | 91 |
| 22 | 22 | 22 | 63 | 60 | 72 | 75 | 80 | 90 | 68 | 91 | 141 |
| 22 | 22 | 22 | 64 | 59 | 72 | 75 | 81 | 90 | 121 | 99 | 121 |

- Hypothesis:
  $H_0$: The means are all equal
  $H_a$: At least one mean is different equal
- Within group variation
- Between group variation
- ANOVA F-test statistic = $\frac{Bet.Var}{With\ Var}$
- D/M: Reject $H_0$ for large F-values or for small p-value
- For PR, p=0.002. What is your conclusion?

# ANOVA (II): ad-hoc multiple comparison

```
Pairwise comparisons using t tests with pooled SD
data:  qstir$days and factor(qstir$treat)

  1       2       3
2 0.00906 -       -
3 5.6e-07 0.00059 -
4 4.3e-14 1.5e-11 1.9e-06

P value adjustment method: Holm


  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = days ~ treat, data = qstir)

$`factor(treat)`
         diff        lwr      upr     p adj
2-1  7.40000  0.1825386 14.61746 0.0427729
3-1 18.85000 11.1947262 26.50527 0.0000008
4-1 36.65556 29.2403189 44.07079 0.0000000
3-2 11.45000  3.7947262 19.10527 0.0016066
4-2 29.25556 21.8403189 36.67079 0.0000000
4-3 17.80556  9.9635403 25.64757 0.0000037
```

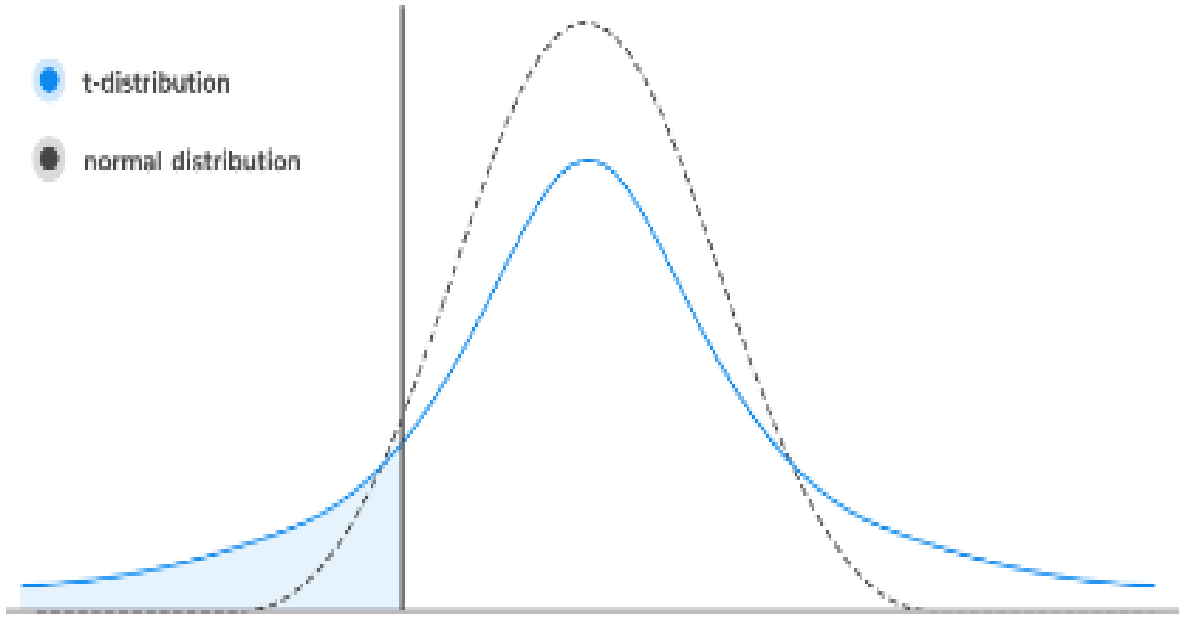
- We need this process only when the overall ANOVA test is statistically significant.
- Approach: Bonferroni, Fisher, Dunnett, LSD, **Tukey**, **Holm**, etc
- How many pairwise groups?
  - ➡ For 4 groups, 6 pairwise groups.
- Holm's conclusion:  All six pairwise groups are significantly different.
- Tukey's conclusion: All six pairwise groups are significantly different.
- Tukey method is most conservative.

# Thought 1:

The distribution of systolic blood pressure (SBP) in a large group is approximately normally distributed. The mean is 120 and approximately 95% of the SBP are between 90 and 150. Then what is standard deviation (SD) of the distribution of SBPs equal to?

# Thought 2: Normal .vs. t-distribution

The critical value at 97.5 percentile in normal distribution is $Z_{0.975}$=**1.96** while the critical value at 97.5 percentile in t-distribution is $t_{0.975}$ = **2.23**. Even though the t-distribution seems to be similar to the normal for more than 30 samples, why do these two numbers be different?



# Thought 3:

Our goal in statistical inference is minimize the two types of errors. But it's impossible to control two errors simultaneously because the two errors ($\alpha$ and $\beta$) have the reverse relationship. How can we construct a clinical trial with good study designs?


Q&A

# Topic 1: Normal rule of thumb

- For normal data (bell-shape), apply the data show a 68-95-99% Rule.