STA305/1004 - Class 5

January 23, 2017

Today's Class

- REMINDER: Assignment #1 due Jan. 27 on Crowdmark via portal by 22:00
- In-class problem on last week's material
- Introduction to Phase III Clinical Trials
- Introduction to power

What are clinical trials?

Clinical trials are prospective intervention studies with human subjects to investigate experimental drugs, new treatments, medical devices, or clinical procedures (Yin, 2012).

Phases of clinical trials

Developing a new drug for cancer.

- **Preclinical studies:** In vitro (e.g. slides, test tubes) and in vivo (living organism such as rodents) studies on wide range of doses of experimental agents. This stage of study provides preliminary toxicity and efficacy data including pharmacokinetics (PK) and pharmacodynamics (PD) information.
- **Phase I:** Usually first study in humans to investigate the toxicity and side effects of the new agent. Identify MTD.
- **Phase II:** Assess if drug has sufficient efficacy. The drug is usually administered around the MTD. If drug does not show efficacy or is too toxic then further testing is discontinued.

Phases of clinical trials

- **Phase III:** If drug passes phase II testing then it is compared to the current standard of care or placebo. These are long-term, large scale randomized studies that may involve hundreds or thousands of patients.
- If the drug is proven to be effective (e.g. two positive phase III trials required for FDA approval) the company will file an application with regulatory agencies to sell the drug. If approved then the drug will be available to the general population in the country where it was approved.
- **Phase IV:** After approval a study might follow a large number of patients over a longer period of time to monitor side effects and drug interactions. For example, findings from these studies might add a warning label to the drug.

Phases of clinical trials

- The four phases are usually conducted sequentially and separately.
- Each trial requires an independent study design and a study protocol.
- Every aspect of trial design, monitoring, and data analysis call upon statistical methods.
- In randomized clinical trials a treatment group is often referred to as an **arm**.

Phases of clinical trials

- Experimental design plays a very important role in the design of clinical trials.
- Two arm clinical trials use all of theory of randomization that we learned about last week. Randomization is used to design phase III clinical trials since causation can usually be assessed using a randomized design.

How can causation be assessed using a randomized design?

- Suppose that patients are randomized in a two arm clinical trial where one of the arms is the standard treatment and the other arm is an experimental treatment
- A statistically significant difference in the outcome between the two arms is observed showing the experimental treatment is more efficacious.
- The interpretation is that the experimental treatment *caused* patients to have a better outcome since the only difference between the two arms is the treatment. Randomization is supposed to ensure that the groups will be similar with respect to all the factors measured in the study and all the factors that are not measured.

How can causation be assessed using a randomized design?

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 JANUARY 19, 2017 VOL. 376 NO. 3

Ocrelizumab versus Placebo in Primary Progressive Multiple Sclerosis

X. Montalban, S.L. Hauser, L. Kappos, D.L. Arnold, A. Bar-Or, G. Comi, J. de Seze, G. Giovannoni, H.-P. Hartung, B. Hemmer, F. Lublin, K.W. Rammohan, K. Selmaj, A. Traboulsee, A. Sauter, D. Masterman, P. Fontoura, S. Belachew, H. Garren, N. Mairon, P. Chin, and J.S. Wolinsky, for the ORATORIO Clinical Investigators*

How can causation be assessed using a randomized design?

Table 1. Baseline Demographic and Disease Characteristics (Intention-to-Treat Population).*

Characteristic	Ocrelizumab (N=488)	Placebo (N=244)
Age — yr		
Mean	44.7±7.9	44.4±8.3
Median (range)	46.0 (20–56)	46.0 (18–56)
Female sex — no. (%)	237 (48.6)	124 (50.8)
Time since onset of MS symptoms — yr†		
Mean	6.7±4.0	6.1±3.6
Median (range)	6.0 (1.1–32.9)	5.5 (0.9–23.8)
Time since diagnosis of PPMS — yr‡		
Mean	2.9±3.2	2.8±3.3
Median (range)	1.6 (0.1–16.8)	1.3 (0.1–23.8)
No previous use of disease-modifying therapy — no. (%)§	433 (88.7)	214 (87.7)
Score on EDSS¶		
Mean	4.7±1.2	4.7±1.2
Median (range)	4.5 (2.5–7.0)	4.5 (2.5–6.5)
Gadolinium-enhancing lesions on T_1-weighted images — no./total no. (%)		
Yes‖	133/484 (27.5)	60/243 (24.7)
No	351/484 (72.5)	183/243 (75.3)
No. of lesions on T_2-weighted images**		
Mean	48.7±38.2	48.2±39.3
Median (range)	42.0 (0–249.0)	43.0 (0–208.0)
Total volume of lesions on T_2-weighted images — cm^3**		
Mean	12.7±15.1	10.9±13.0
Median (range)	7.3 (0–90.3)	6.2 (0–81.1)
Normalized brain volume — cm^3††		
Mean	1462.9±84.0	1469.9±88.7
Median (range)	1462.2 (1214.3–1711.1)	1464.5 (1216.3–1701.7)

* Plus–minus values are means ±SD. Patients were stratified according to geographic region (United States vs. rest of the world) and age (≤45 vs. >45 years). MS denotes multiple sclerosis, and PPMS primary progressive MS.
† Data were not available for 14 patients in the ocrelizumab group and 7 patients in the placebo group.
‡ Data were not available for 2 patients in the ocrelizumab group and 1 patient in the placebo group.
§ Shown are data for patients with no use of disease-modifying therapy in the 2 years before trial entry.
¶ Scores on the Expanded Disability Status Scale (EDSS) range from 0 to 10, with higher scores indicating greater disability. Data were not available for 1 patient in the ocrelizumab group.
‖ A breakdown of the categorical numbers of gadolinium-enhancing lesions on T_1-weighted images is provided in Table S7 in the Supplementary Appendix.

How many patients should be enrolled in a Phase III clinical trial?

- In a phase III trial sample size is the most critical component of the study design. The sample size has implications for how many subjects will be exposed to a drug that has no proven efficacy.
- The investigator needs to specify type I, II error rates, and the effect sizes.
- Standard practice is to compute the smallest sample size required to detect a clinically important/significant treatment difference with sufficient.

How many patients should be enrolled in a Phase III clinical trial?

- If the sample size is too small then the trial might fail to discover a truly effective drug because the statistical test cannot reach the significance level (5%) due to a lack of power.
- If the sample size is overestimated then resources wasted and drug development delayed since patient enrollment is often the main factor in time to complete a trial.

Statistical hypotheses

Suppose that subjects are randomized to treatments A or B with equal probability. Let μ_A be the mean response in the group receiving drug A and μ_B be the mean response in the group receiving drug B. The null hypothesis is that there is no difference between A and B, the alternative claims there is a clinically meaningful difference between them.

$$H_0 : \mu_A = \mu_B \text{ versus } H_0 : \mu_A \neq \mu_B$$

Statistical hypotheses

The type I error rate is defined as:

$$\begin{aligned}\alpha &= P\left(\text{type I error}\right) \\ &= P\left(\text{Reject } H_0 \text{ when } H_0 \text{ is true}\right).\end{aligned}$$

Statistical hypotheses

The type II error rate is defined as:

$$\begin{aligned}\beta &= P\left(\text{type II error}\right) \\ &= P\left(\text{Accept } H_0 \text{ when } H_1 \text{ is true}\right).\end{aligned}$$

Statistical hypotheses

Power is define as:

$$\begin{aligned}\text{power} &= 1 - \beta \\ &= 1 - P\left(\text{Accept } H_0 \text{ when } H_1 \text{ is true}\right) \\ &= P\left(\text{Reject } H_0 \text{ when } H_1 \text{ is true}\right).\end{aligned}$$

Power

The probability that a fixed level α test will reject H_0 when a particular alternative value of the parameter is true is called power of the test to detect that alternative.

Power

Can a 6-month exercise program increase the total body bone mineral content (TBBMC) of young women? Based on results of a previous study $\sigma = 2$ for the percent change in TBBMC over the 6-month period. A change in TBBMC of 1% would be considered important. Is 25 subjects a large enough sample size for this project?