Chapter 11: Chance

Fletcher Seminaire

3rd of May 2023

Presented by: Ida Sigvardsson

Two approaches to chance

- **Hypothesis testing:**

  Tests if one method/treatment is statistically significant better than another **(unlikely only by chance).**

Statistical significance? ← P-value

- **Estimation:**

  Uses statistical methods to estimate the range of values that is likely to include the true value.

Hypothesis testing

Four ways in which the statistical conclusions might relate to reality:

- True difference present & statistical test significant (Correct)
- True difference absent & statistical test not significant (Correct)
- True difference present & statistical test not significant (Type 2 error)
- True difference absent & statistical test significant (Type 1 error)

		True difference	
		Present	Absent
Conclusion of test	Significant	CORRECT	TYPE 1 ERROR
	Not significant	TYPE 2 ERROR	CORRECT

False-negative

False-positive

Statistical tests in clinical research

Test the Statistical Significance of a Difference

Test	When used
Chi square	Between 2 or more proportions
Fishers exact test	Between two proportions
Mann-Whitney U	Between two medians
Student t	Between two means
F test	Between two or more means

To Describe the Extent of Association

Test	When used
Regression coefficient	Between an independent variable and a dependent variable
Pearson's r	Between two variables

To model the Effects of Multiple Variables

Test	When used
Logistic regression	With a dichotomous outcome
Cox proportional hazard	With a time-to-event outcome
Propensity score	Select matching controls

Statistical power

- The probability that a study will find a statistically significant difference when a difference really exists *(not by chance).*

- A study is powerful when it has a high probability of detecting differences when treatments really do have different effects.

How many study participants are enough to compare the effect of two treatments?

Estimating sample size requirements

Effect size

Type 1 error

Type 2 error

Characteristics of the data

Interrelationships

Requirements

- Then

VS

- Now

Large sample sizes were not needed to discover powerful treatments, we came a long way on clinical experience.

Many effective treatments are well established. Larger sample sizes needed to discover chronic diseases with multiple interacting causes & when treatment effects are small.

Statistical precision & effect size

- Point estimate
- Confidence intervals

“If the study is unbiased, there is a 95% chance that the interval includes the true effect size”.

“The true value is most likely to be close to the point estimate, less likely to be near the outer limits of the interval, and could (5 times out of 100) fall outside these limits altogether”.

Statistical precision & effect size

- Point estimate
- Confidence intervals

Statistically significance does not always equal clinical relevance (and vice versa).

Statistical power

- Statistical power can be estimated by means of readily available formulas, tables, nomograms, computer programs, or web sites.

- **Statistical precision increases with the statistical power of the study**

- **After a study is completed,** there is no longer a need to test statistical power (estimate effect size, outcome event rates, and variability among participants) because it is already known.

  Attention should instead be on the point estimates and confidence intervals.

Detecting rare events

A lot of people needs to be observed in order to have a good chance of detecting rare events. (Probability of detecting an event as a function of the number of people under observation:

To have a good chance of detecting a 1/x event, one must observe 3x people (for example if the underlying rate is 1/10,000, one must observe 30,000 people).

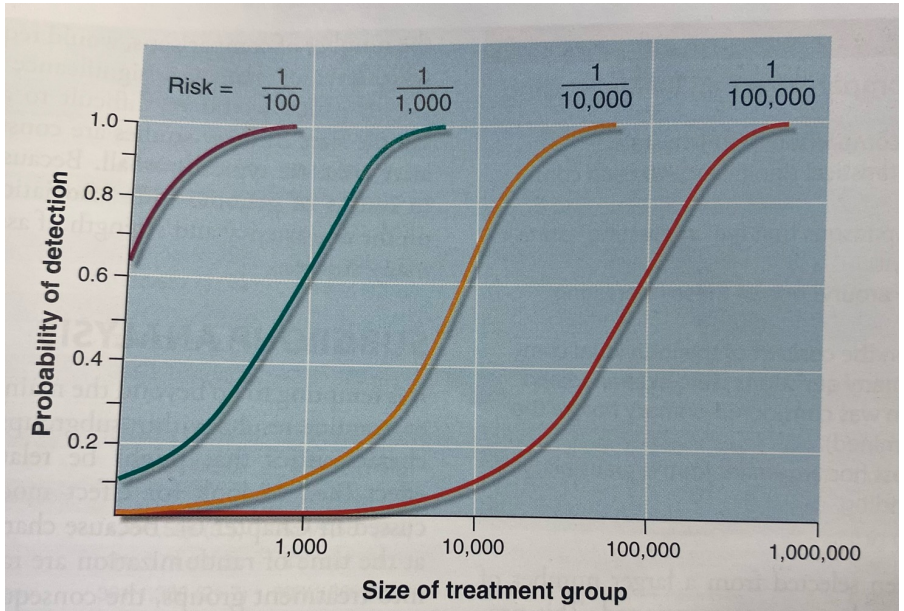

Figure 11.6. The probability of detecting one event according to the rate of the event and the number of people observed. (Reprinted from Guess HA, Rudnick SA. Use of cost-effectiveness analysis in planning cancer chemoprophylaxis trials. *Control Clin Trials* 1983;4:89–100, with permission from Elsevier.)

Multiple comparisons

- A common way to manage multiple comparisons:
  divide the usual P value of 0.05 by the number of comparisons.

- Multiple outcomes

$$\frac{\text{Usual P-value}}{\text{Number of comparisons}} = \text{New P-value to consider}$$

Different kinds of trials

- **Superiority trials:**

  Conducted to establish if one treatment is better than another.

- **Noninferiority trials:**

  The purpose is to show that an alternative treatment is no worse than an established treatment (at least for clinically important difference).

Multivariable methods

To examine the effect of several variables together:

Multivariable modeling
(a mathematical expression)

Advantage: No other way to adjust for or include many variables

Disadvantage: Validity based on assumpsions that may not be met, easily affected by random variation

General process:

1) Identify all variables possibly related to outcome (confounders/effect modifiers)
2) If few outcome events: select the variables most likely related to outcome (if statistical testing: $P<0.10$)
3) Check if assumptions of the data structure are met (e.g., if risk does not vary over time)
4) All variables entered in the statistical model (e.g., Cox proportional hazards model)