Basic Statistical Concepts and Methods for Research Studies

Rick White

Department of Statistics, UBC

Pathology and Lab Medicine

Faculty of Medicine

October 25, 2011

Discussion Point

What does the word
"Statistics"
mean to you?

The discipline of Statistics

- Wikipedia: Statistics is the study of the collection, organization, and interpretation of data.
- American Heritage Science Dictionary: The branch of mathematics that deals with the collection, organization, analysis, and interpretation of numerical data.
- Merriam-Webster Dictionary: A branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data.
- Dictionary.com: the science that deals with the collection, classification, analysis, and interpretation of numerical facts or data, ...

Planning your Study

- Define the questions of interest.
- Determine the appropriate populations (people/animals/etc) that will allow the questions to be answered.
- Create a plan to sample the populations. Randomization may be required.
- Determine what information is needed from the sample to answer the questions.
- Create an appropriate analysis plan.

Statisticians can help

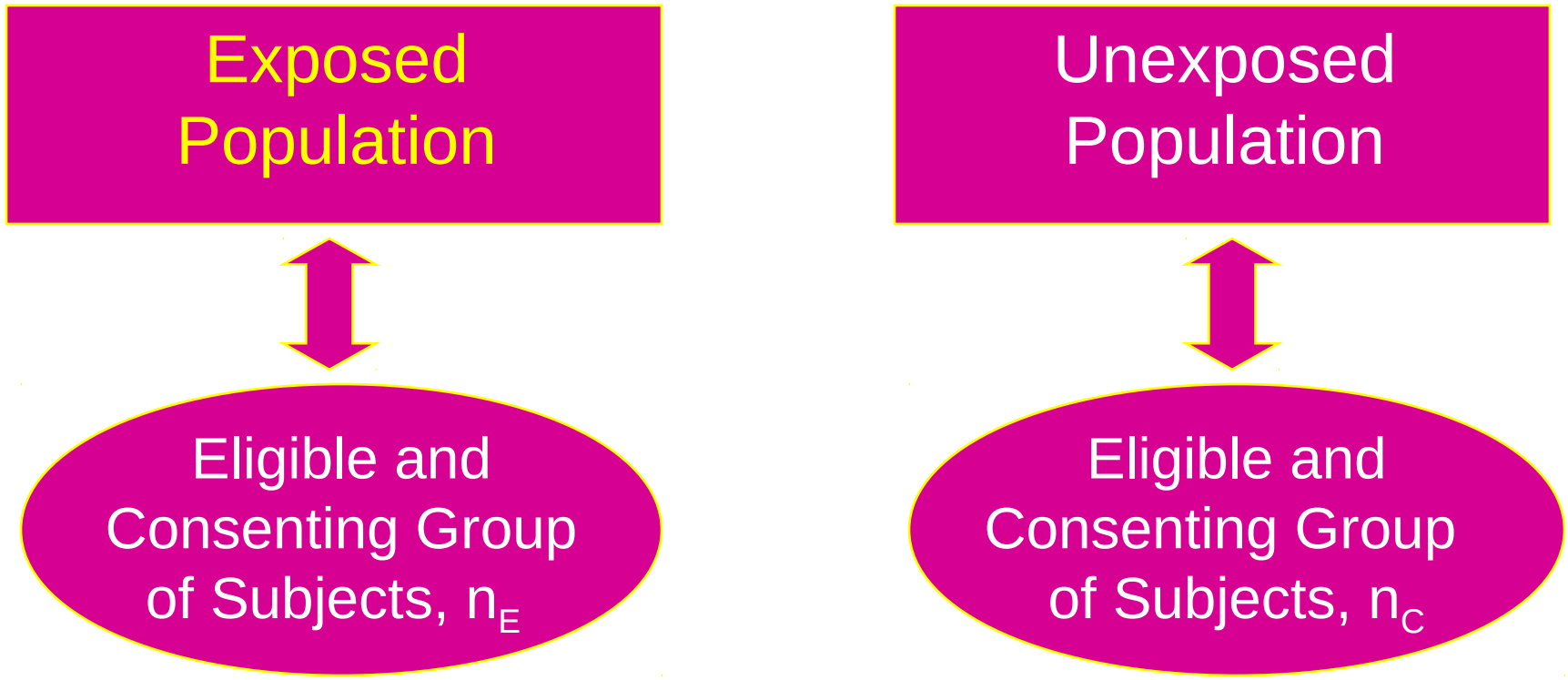
- Focus and clarify the objectives
- Design an appropriate sampling plan
- Provide a randomization scheme
- Design an appropriate analysis plan

➔ Talk to a statistician before you collect data !!!

Two Types of Research Studies

- Observational Studies
  - In epidemiology and statistics, an observational study draws inferences about the possible effect of a treatment on subjects, where the assignment of subjects into a treated group versus a control group is outside the control of the investigator.
- Controlled Experiments
  - There are many forms of controlled or designed experiments. A relatively simple one randomly assigns research subjects or specimens into two groups: an experimental group and a control group.

A Two-Group Observational Study



Observational Studies: Some Issues

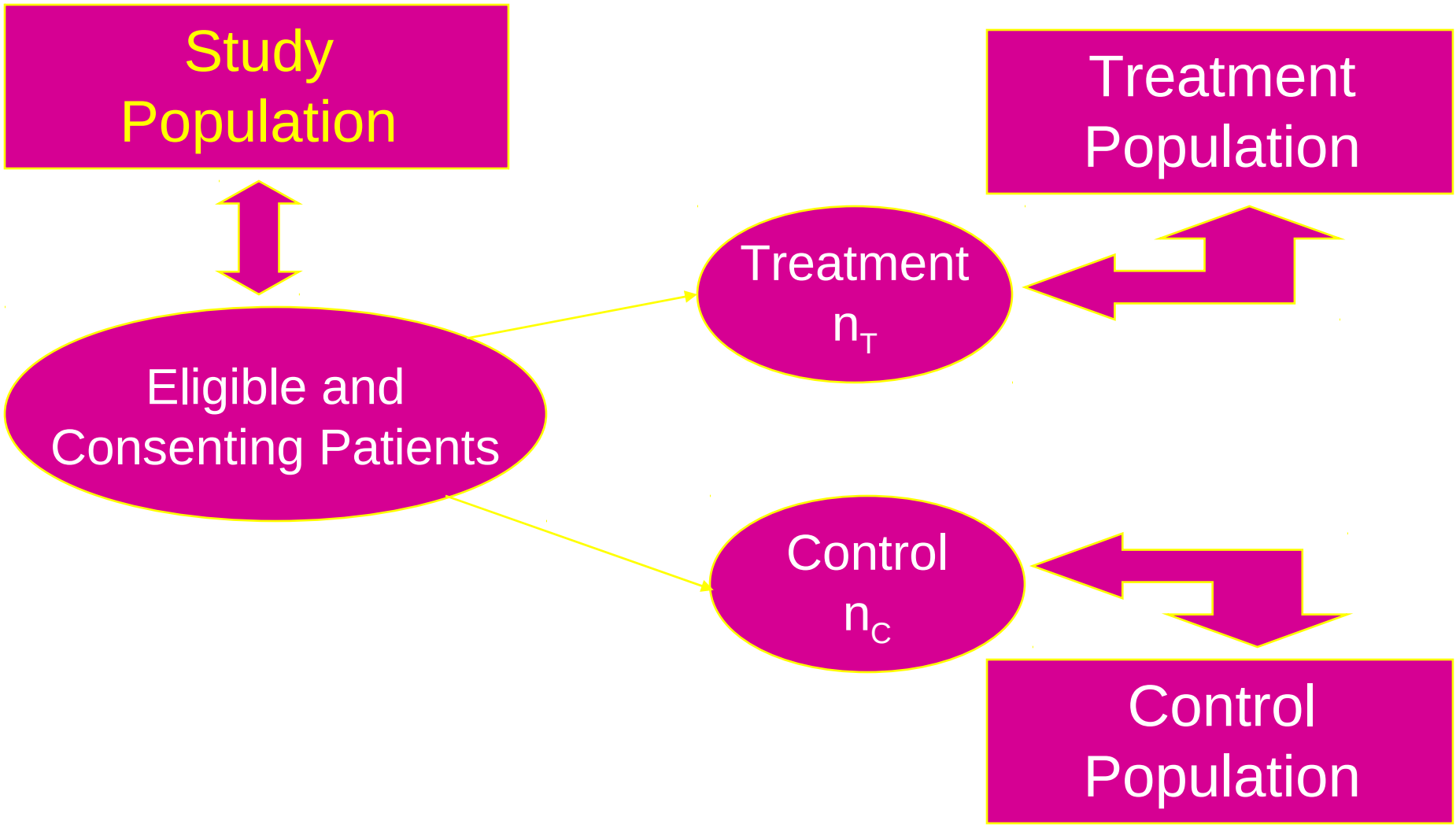
- Describing differences is straightforward
  - Examination of associations with exposure
- Confounding is a major issue
  - Differences could be due to a factor other than exposure that differs across the groups
  - Analyses must adjust for such confounders
- Analyses can suggest, but not establish causality
  - Bradford Hill's criteria for causality
- If based on samples, are samples representative?

Confounding Example

- Kidney Stones
- The table shows the success rates and numbers of treatments for treatments involving both small and large kidney stones, where Treatment A includes all open procedures and Treatment B is percutaneous nephrolithotomy:

	Treatment A	Treatment B
Small Stones	**93% 81/87**	87% 234/270
Large Stones	**73% 192/263**	69% 55/80
Both	78% 273/350	**83% 289/350**

A Two-Armed Clinical Trial



Clinical Trials: Some Issues

- Ethics of experimentation.
  - Cannot force someone to smoke
- Choice of control: nothing, placebo or active.
  - The placebo effect
  - ethically must treat if one already exists
- Blinding of subjects and evaluators.
  - subconscious effects
- Target population versus study population.

Defining the Question

- Make sure questions are clear and focused.
- All questions should be based on the same populations or perhaps a subset.
- Each question should define a single hypothesis to test or quantity to measure.
- “Does Betaseron decrease the relapse rate in relapsing-remitting MS patients?”
- “What is the change in relapse rate between relapsing-remitting MS patients treated with Betaseron and those who are not.”

Multiple Questions

- The more questions the greater the overall type 1 error rate (familywise error rate).
- Worse case scenario: Independent events
  - P(A & B) = P(A)*P(B)
- Suppose a researcher is trying to find a molecule that would be good biomarker for a given disease. If they look at 5, 10 or 20 markers with $\alpha = 0.05$ then the probability of at least 1 false positive = 0.23, 0.40 or 0.64 respectively.
- Using a Bonferroni correction ($\alpha^* = \alpha/n$) with 5, 10 or 20 biomarkers then $\alpha^* = 0.01$, 0.005 or 0.0025 and the probability of at least 1 false positive = 0.049 in all cases.

Multiple Question: cont

- Hypotheses in same study are usually not independent
- Bonferroni is conservative and leads to a larger sample size to maintain the same power
  - if n=64 is needed to detect a difference of 1/2 a standard unit with power = 0.80 for a single question in a two-sample t-test, then with 5, 10, 20 questions the sample size becomes 96, 109, 122 respectively
- Less conservative methods are available such as False Discovery Rate
- Recommendations: limit the number of questions or select a few questions as primary

The Sampling Plan

- Quite often the sample is based on the available population for experiments and observational studies.
  - clinical trials or case-control studies.
- Are sampled units independent?
  - Most analyses make this assumption.
- Does the sample represent the population?
  - This is who is sampled not how many.
  - This is statistical accuracy.
- Does the sample give enough precision or power?
  - This is how many are sampled.
  - Precision for estimation, power for testing.

Randomization

- Should assign a subject or specimen to any treatment or control group with probability based on relative sample size within those groups.
- Controls bias in the study.
- Does not guarantee that the groups are "matched" or equivalent, only that any differences are due to chance.
- Randomization should be used to balance the experiment as much as possible if several factors are considered
- Can be combined with some observational components such as gender or age.
- Randomization should be used to remove order effects within a subject if using repeated measures under different conditions

Experimental Unit

- The level of randomization or sampling.
- Sampling unit does not always equal the experimental unit
- The level of independent observations.
- Pseudoreplication are observations within an experimental unit: repeated measures
  - results in over exaggeration of statistical significance
  - some pseudoreplicates are hard to detect.
- Discussion Point: 100 mice are assigned to 20 cages. Each cage is fed a diet which contains either treatment or placebo. What is the experimental unit and why?

Sample Size Considerations

- Most statistical analysis depend on a reasonable sample size to be valid
- As sample size increases
  - statistical power of a hypothesis test increases
  - precision of estimates increases by the square root law
  - distribution of statistics become more predictable
    - Central limit theorem -> normal distribution
  - statistical accuracy does not change
- Demonstration **http://onlinestatbook.com/stat_sim/sampling_dist/index.html**

Sample Size/Power example

two sample t-test: detecting a 1/4 standard unit difference

N	Power
20	11.7%
40	19.6%
60	27.4%
80	34.9%
100	42.0%

two sample t-test: detecting a 1/2 standard unit difference

N	Power
20	33.8%
40	59.8%
60	77.5%
80	88.1%
100	94.0%

The Data

- Data must answer your questions
- Blind collectors to avoid bias
- Accuracy of measurements
- Accurate data entry
- Dealing with missing data
- Data issues should be resolved before analysis begins
- Data should be finalized and locked before analysis.
- Discussion Point: Why should the data be finalized before analysis begins?

Types of variables

- Numeric variables
  - Continuous variables (interval or ratio)
  - Discrete variables (binary or counts)
- Categorical variables (always discrete)
  - ordinal (can be ordered or ranked)
  - nominal (categories without order)
- Binary data can be treated as numeric, ordinal or nominal
- Response variables (dependent variables)
- Explanatory variables (independent variables)

Common Analyses

- Continuous response variable:
  - t-test, paired t-test, rank tests
  - regression, ANOVA, linear models, survival analysis
- Discrete numeric response variable:
  - Count: Poisson regression (possibly over-dispersed)
  - Binary: logistic regression
  - Contingency tables: Chi Squared or Fishers Exact tests
- Categorical responses are more complicated
  - Ordinal: Proportional odds models
  - Nominal: multinomial models
- Pseudoreplication complicates the analysis
  - Mixed effects models, repeated measures analysis

Model Assumptions

- Every model is based on a set of assumptions
- Most important assumption is data are i.i.d.
  - independent and identically distributed
  - required for most common statistical models and test
- Many models assume normally distributed errors
  - t-test, regression, ANOVA
  - with large sample size not a critical assumption
- Other distributions for logistic or Poisson regression models
- non-parametric tests make no assumptions about distribution
- effect of violating assumptions heavily depends on the situation
- some violations are more critical than others
- Assumptions should be checked if possible

Hypothesis test

- Define null and alternative hypothesis
  - Null Hypothesis: the default condition
    - Absence of evidence is not evidence of absence
  - Alternative Hypothesis: what we want to show
- Determine threshold needed to reject null
  - usually based on α (false positive probability)
- Collect data needed to test hypothesis
  - sample size determined by power analysis
- Analyze data to determine result
  - Null is true, do not reject null: correct decision
  - Null is true, reject null: Type 1 error
  - Alt is true, do not reject null: Type 2 error
  - Alt is true, reject null: correct decision (power)

The End

Questions?