

BIOSTATISTICS METHODS

FOR TRANSLATIONAL & CLINICAL RESEARCH



DESIGNING CLINICAL RESEARCH

Part A: Concepts & Designs

TRIAL BY JURY: RATIONALE

“Trial by Jury” is the way our society deals with uncertainties. Its goal is to minimize errors/mistakes (not to eliminate them; mistakes are still being made every day!)

HOW DOES SOCIETY DEAL WITH UNCERTAINTIES ?

- We form Assumption/Hypothesis: “Every person is innocent until proven guilty” (written in our Constitution),
- We gather **data**: Evidence against Hypothesis- not against the suspect, then
- We decide whether Hypothesis should be rejected (If it is, the verdict is “Guilty”)

ELEMENTS OF A SUCCESSFUL TRIAL

- A probable **CAUSE** (a crime and a suspect)
- A thorough **INVESTIGATION** (by police)
- An efficient **PRESENTATION** (by D.A.'s office/attorneys- including the organization and summarization of evidence)
- A fair & impartial **ASSESSMENT** by Jury, after that a decision is made.

HOW DOES SCIENCE DEAL WITH UNCERTAINTIES ?

- **We form Assumption/Hypothesis: From experience & observations (The process leads to the so-called research questions)**
- **We gather data: Experiments & Trials, Surveys, Medical Records Abstractions.**
- **We make decision based on what we find in the data.**

ELEMENTS OF GOOD RESEARCH

- A good Research Question with well-defined objectives & endpoints,
- A thorough Investigation, lots of data
- An efficient Presentation: data organization & summarization, and
- A proper Statistical Inference (the process & methods of drawing conclusions)

AREAS OF BIOSTATISTICS

Research/Biostatistics is a three-step process:

- ❖ **Sampling/design**: Find a way or ways to collect data (**going from population to sample**).
- ❖ **Descriptive statistics**: Learn to organize, summarize and present data which can shed light on the research question (**investigating sample**).
- ❖ **Inferential statistics**: Generalize what we learn from the sample or samples to the target population and answer the research question (**going from sample to population**).

THE IMPORTANT PHASE

Just as in the case of “Trial by Jury”, the most **important stage** of the “Research Process” is the **Design**: How & How Much data are collected! Also, It dictates how data should be analyzed. **May be it’s less on the question of “how” to collect your data but the decision on “when to do what”!**

WHAT IS A RESEARCH DESIGN

Very briefly, designed experiments are conducted to demonstrate a cause-and-effect relation between one or more explanatory factors (or predictors) and a response variable. The demonstration of a cause-and-effect relationship is accomplished by altering the level or levels of the explanatory factors and observing the effect of the changes (i.e. designed values of predictors X's) on the response variable Y.

A Simple Example:

An experiment on the effect of Vitamin C on the prevention of colds could be simply conducted as follows. A number of n children (the sample size) are randomized; half were each give a 1,000-mg tablet of Vitamin C daily during the test period and form the “experimental group”. The remaining half , who made up the “control group” received “placebo” – an identical tablet containing no Vitamin C – also on a daily basis. At the end, the “Number of colds per child” could be chosen as the outcome/response variable, and the means of the two groups are compared.

Assignment of the treatments (factor levels: Vitamin C or Placebo) to the experimental units (children) was performed using a process called “randomization”. The purpose of randomization was to “balance” the characteristics of the children in each of the treatment groups, so that the difference in the response variable, the number of cold episodes per child, can be rightly attributed to the effect of the predictor – the difference between Vitamin C and Placebo.

STUDY DESIGNS

- **How data will be collected? It's the complex issue of when to do what:**
 - ❖ **Cross-section Design,**
 - ❖ **Case-Control Design (retrospective),**
 - ❖ **Cohort Design (prospective); Clinical Trials (one-arm, two-arm; open-label or randomized; double or triple blind)**

The cross-sectional designs are very popular in social/behavioral studies, e.g. teen surveys. As for health research data, since **diseases are rare**, there are very few of these; fundamental designs are case-control and cohort.

Retrospective Studies gather past data from selected cases (with disease) and controls (without disease) to determine differences, if any, in exposure to a suspected risk factor. Advantages: Economical & Quick. Major Limitations: Accuracy of exposure histories & Appropriateness of controls.

Prospective studies enroll two groups of subjects, say Treatment and Placebo in Clinical Trial; subjects are followed over time to obtain result (say, new SBP or occurrence of an event).

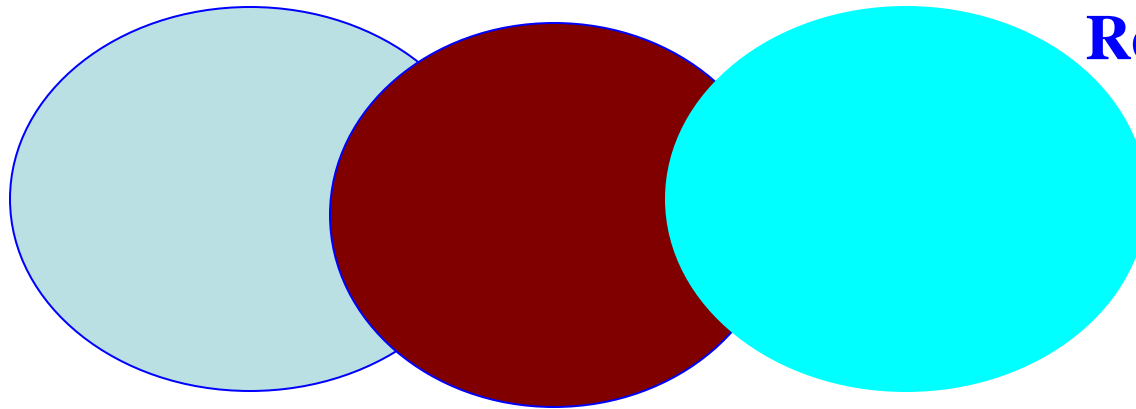
Randomization is very important but there are one-group prospective studies and “one-arm trials**” in translational and clinical research.**

SOME TERMINOLOGIES

- **Research Designs**: Methods for data collection
- **Clinical Studies**: Class of all scientific approaches to evaluate Disease Prevention, Diagnostics, and Treatments.
- **Clinical Trials**: Subset of clinical studies that evaluates Investigational Drugs; they are in prospective/longitudinal form (the basic nature of trials is prospective).

Clinical Research

**Population
Research**



**Laboratory
Research**

Translational Research is the component of basic science that interacts with clinical research or with population research.

It often starts with some common/ typical, health-oriented curiosity (e.g. **What is the average fish consumption in American diet?**) leading to a descriptive/observational research study. For example, a “**survey**” which requires minimal statistical supports – mostly descriptive.

The it would be followed by an analytic/observational study for some more scientific/statistical curiosity (e.g. Is there an association between fish intake and risk of myocardial infarction?). More statistical supports here: Correlation & Regression

Then concluded with a randomized, controlled clinical trial to establish the case for interventions (e.g. Does treatments with fish oil capsules reduce total cardiovascular mortality?)

Sometimes, an one-arm, early-phase trial or trials could precede this if treatment or intervention is more experimental – say, with possible side effects.

THE ANATOMY & PHYSIOLOGY OF CLINICAL RESEARCH

We form or evaluate a research or research project from/on two different angles or parts: the **anatomy** and the **physiology** of research; just like the hardware and software to run a computer operation.

THE ANATOMY PART

- ❖ From the anatomy of the research, one can describe/see **what it's made of**; this includes the tangible elements of the study plan: **research question, design, subjects, measurements, sample size calculation, etc...**
- ❖ The **goal** is to create these elements in a form that will make the project **feasible, efficient, and cost-effective.**

THE PHYSIOLOGY PART

- ❖ From the physiology of the research, one can describe/see how it works; first about what happened in the study sample and then about how study findings generalized to people outside the study.
- ❖ The goal is to minimize the errors that threaten conclusions based on these inferences.

Separating the two aspects is artificial because the anatomy does not make much sense without some understanding of its physiology; but the separation would clarify our thinking and understanding as learners of clinical research. Just like medical students learn about human anatomy and physiology before being allowed to make hospital rounds.

THE PROTOCOL

- The structure of a Research Project, or its anatomy, is described in its protocol; the written part of the study.
- The Protocol have a vital scientific function to help the investigator organize his/her research in a logical, focused, and efficient way.

RESEARCH PROTOCOL

- Before a research project is open to enroll patient, investigators have to write a “protocol” – a detailed written proposal.
- That proposal, “the protocol”, have to be reviewed and approved by two separate regulatory groups: an Institutional Review Board (IRB) to scrutinize its safety aspects, such as the patients’ written “consent form” and a research-oriented committee of peers to finalize the peer-review, peer-approval process (eg. Cancer Protocol Review Committee in cancer centers)

PROTOCOL REVIEW

❖ The protocol review by peers are focused on two different aspects:

- (1) Scientific merit of the protocol: **scientific relevance, validity of the hypothesis, adequate study design, biostatistics input, adequate patient population, and feasibility of timely completion;**
- (2) Priority of the proposed study in regards to competing protocols, priority of its scientific merit and its **impact on existing studies.**

COMPONENTS OF THE ANATOMY

- ❖ **Research Question:** What is the objective of the study, the uncertainty the investigator wants to resolve?
- ❖ **Background and Significance:** Why these questions important?
- ❖ **Design:** How is the study structured?
- ❖ **Subjects:** Who are the subjects and how they will be selected and recruited.
- ❖ **Variables:** What measurements will be made: predictors, confounders, and outcomes.
- ❖ **Statistical Considerations:** How large is the study and how will data be analyzed (“**Design**” is an important statistical component but listed in the Design Section).

RESEARCH QUESTION

- The research question is the objective of the study, the uncertainty the investigator wants to resolve
- A good research question should pass the “so what?” test; setting the answer should contribute usefully to our state of knowledge

Research questions often begin with a “general concern” but that must be narrowed down to concrete, “researchable issues”. For example:

Concern: Should people eat more fish

Specific issues:

(1) Does eating fish lower the risk of cardiovascular disease?

(2) Is there a risk of mercury toxicity from increasing fish intake in older adults?

The acronym **FINER** denotes five essential characteristics of a good research question:

Feasible,

Interesting,

Novel,

Ethical, and

Relevant.

BACKGROUND & SIGNIFICANCE

- This section of a protocol sets the proposed study **in context** and gives its rationale
 - ❖ **What is known** about the topic at hand?
(Citing previous research that is relevant – including the investigator's own work)
 - ❖ What are problems with the prior/cited research & **what uncertainties remain.**

MORE SPECIFICS

- Two basic items are:
 - ❖ Why is the research question important?
 - ❖ What kind of answers will study provide?
- Often work on the significance section would help/lead investigator to modifications in the issues of the research question.

STUDY SUBJECTS

- Decisions on two major issues: **Who** are included and **How** to recruit them?
 - ❖ The first is to specify **Inclusion and Exclusion criteria** that define the kinds of patients best suited to the research questions
 - ❖ **Where/When** to recruit enough people in order to answer the research question – including feasibility of recruitment.

INCLUSION/EXCLUSION

- **Inclusion criteria:** Patient characteristics required for entry, describing the population of patients that the drug is intended to serve. There are also **exclusion criteria** as well.
- For eligibility, consideration should be given to patients who are likely to benefit from treatment **and** to the generalization of the results:
 - (i) Effectiveness of the treatment may be masked by the inclusion of patients with little chance of responding;
 - (ii) On the other hand, with narrow criteria, generalization may be compromised.

VARIABLES

A set of decisions concerning the choice or choices of which variables to measure.

- ❖ **Predictor Variables**
- ❖ **Outcome Variables** (primary, secondary)
- ❖ **Confounders or Confounding Variables**
(and how to control them)

CONFOUNDERS

- A **confounder** is not under investigation, but may be **related to the primary “outcome” and/or predictor variables**; an effect modifier is a special case – An **effect modifier alters effect of some predictor variables**.
- We can control for confounders through the design (stratification, stratified randomization) or through the analysis (use of regression).

BIOSTATISTICS REVIEW

- (1) The endpoint or endpoints of the study
- (2) Patient characteristics that may affect response (leading to possible stratification)
- (3) The design of the study
- (4) Accrual goal and statistical power.
- (5) A Plan for Data Management
- (6) Method or Methods for data analysis
- (7) Criteria for stopping the trial

Trials & Validity

SURVEYS & SAMPLING

The format of surveys is such that one can assume there is an identifiable, existent parent population of subjects. We act as if the sample is obtained from the parent population according to a carefully defined technical procedure called random sampling.

VALIDITY OF BIOMEDICAL RESEARCH

- **Not only diseases are rare, there is a basic issue here: WE usually do not have “random samples” in real-life biomedical studies.**
- **The laboratory investigator uses animals in his projects but the animals are not randomly selected from any large population of animals. The clinician, who is attempting to describe the results he has obtained with a particular therapy, cannot say that his patients is a random sample from a parent population of patients.**

LOGICAL APPROACH

- Because they are not population-based (there is not an identifiable, existent parent population of subjects for sample selection), biomedical studies - both cohort and case-control- are “comparative”. **That is the validity of the conclusions is based on a comparison.**
- In a cohort study, say a clinical trial, we compare the results from the “treatment group” versus the results from the “placebo group”.
- In a case-control study, we compare the “cases” versus the “controls” with respect to an exposure under investigation (“exposure” could be binary or continuous).

Efforts are needed to assure that we have “**comparable groups**”; for example, in prospective trials, **randomization** is a necessary component/procedure. Sometimes, special form of randomization: **stratified randomization.**

Inference & Validity

Validity is an important concept/component; it involves the assessment against accepted absolute standards which are often not available; or in a milder form, **to see if the evaluation appears to cover its intended target or targets.**

INFERENCES & VALIDITIES

- Two major levels of inferences are involved in interpreting a study, a clinical trial
- ❖ The first level concerns Internal validity; the degree to which the investigator draws the correct conclusions about what actually happened in the study.
- ❖ The second level concerns External Validity (also referred to as generalizability or inference); the degree to which these conclusions could be appropriately applied to people and events outside the study.

External Validity

Internal Validity

Truth in
The Universe

Truth in
The Study

Findings in
The Study

Research Question

Study Plan

Study Data



A Simple Example:

An experiment on the effect of Vitamin C on the prevention of colds could be simply conducted as follows. A number of n children (the sample size) are **randomized**; half were each give a 1,000-mg tablet of Vitamin C daily during the test period and form the “experimental group”. The remaining half , who made up the “control group” received “placebo” – an identical tablet containing no Vitamin C – also on a daily basis. At the end, the “Number of colds per child” could be chosen as the outcome/response variable, and the means of the two groups are compared.

Assignment of the treatments (factor levels: Vitamin C or Placebo) to the experimental units (children) was performed using a process called “**randomization**”. The purpose of randomization was to “balance” the characteristics of the children in each of the treatment groups, so that the difference in the response variable, the number of cold episodes per child, can be rightly attributed to the effect of the predictor – the difference between Vitamin C and Placebo. **Randomization helps to assure Internal Validity.**

What about External Validity?

For instance, when looked to establish a relationship but found **no statistically significant correlation**. On this basis it is concluded that there is no relationship between the two factors. How could this conclusion be wrong -- that is, **what are the "threats to validity"**? A “conclusion” is a generalization from findings of the study to truth in the universe; it involves external validity.

REJECTION

- Back to a “Statistical Test”.
- The Null Hypothesis H_0 is a “Theory”. The Data are “Reality”. When they do not agree, then **we have to trust the reality**; That’s when H_0 is rejected.
- How do we tell if Theory and Reality do not agree? When the data show overwhelmingly that it is almost impossible to have the data that we already collected if H_0 is true (that it is possible but with a very small probability).

ABOUT REJECTIONS

**Hypothesis Testing
is similar to Trial
by Jury**

A very important concept: when a null hypothesis is not rejected it does not necessarily lead to its acceptance, because a “not guilty” verdict is just an indication of “lack of evidence” and “innocence” is just one of its possibilities. That is, when a difference is not statistically significant, **there are still two possibilities:**

- (i) The null hypothesis is true,
- (ii) There is not enough evidence from sample data to support its rejection (i.e. sample size may be too small).

STATISTICAL ISSUES

- Statistics is a way of thinking, thinking about ways to gather and analyze data.
- The gathering part (i.e. data collection) comes before the analyzing part; the first thing a statistician or a learner of statistics does when faced with a biomedical project is **data collection (followed by data management and data analysis)**.
- Studies may be inconclusive because they were poorly planned, not enough data were collected to accomplish the goals and support the hypotheses.
- **To assure external validity, we have to assure of adequate sample size**

Of course, it is always an issue of possible **trade-offs**: On the one side are the issues of **internal and external validities** (say, you need a study with large enough sample size); on the other **feasibility** (dictated by your ability to recruit patients). Therefore, once the study plan has been formulated, it's still a final decision: whether or not to go for it.

Different Forms of Designs

The simplest form of designed experiments is the “completely randomized design” where treatments are randomly assigned to the experimental units – regardless of their characteristics. This design is most useful when the experimental units are relatively homogeneous with respect to known confounders.

A confounder is a factor which may be related to the treatment or the outcome even the factor itself may not be under investigation. A study may involve one or several confounders. In the above clinical trial example, the primary outcome is SBP reduction and the baseline SBP is a potential confounder. Patients' age may be another one. In theory, values of confounders may have been balanced out between study groups because patients were randomized. But it is not guaranteed; especially if the sample size is not very large.

If confounder or confounders are known, heterogeneous experimental units are divided into homogeneous “block”; and randomizations of treatments are carried out within each block. The result would be a “**randomized complete block design**”.

A Simple Example:

An experiment on the effect of Vitamin C on the prevention of colds could be simply conducted as follows. A number of n children (the sample size) are randomized; half were each give a 1,000-mg tablet of Vitamin C daily during the test period and form the “experimental group”. The remaining half , who made up the “control group” received “placebo” – an identical tablet containing no Vitamin C – also on a daily basis. At the end, the “Number of colds per child” could be chosen as the outcome/response variable, and the means of the two groups are compared.

Some other factors might affect the numbers of colds contracted by a child: age, gender, etc... Let say we focus on gender.

THE CHOICES

- We could perform complete randomization – disregard the gender of the child, and put Gender into the analysis as a covariate; or
- We could randomize boys and girls separately; at the end the proportions of boys in the two groups are similar and there would be no need for adjustment.
- The first approach is a **complete randomized design**; the second is a **randomized complete block design**.
- **Similarly, we could block using “age groups”.**

The term “treatment” may also mean different things; a treatment could be a factor or it could be multifactor. For example, let consider two different aspects of a drug regiment: Dose (Low, High) and Administration mode (say, one tablet a day or two tablets every other day). We could combine these two aspects to form 4 combinations; then treating them as 4 treatments and apply a complete randomize design. We call it a (balanced) **Factorial Design**; the analysis is similar to that of a randomized complete block design.

THE ROLE OF STUDY DESIGN

In a “standard” experimental design, a linear model for a continuous response/outcome is:

$$Y = \begin{bmatrix} \text{Overall} \\ \text{Constant} \end{bmatrix} + \begin{bmatrix} \text{Treatment} \\ \text{Effect} \end{bmatrix} + \begin{bmatrix} \text{Experimental} \\ \text{Error} \end{bmatrix}$$

The last component, ‘experimental error’, includes not only error specific to the experimental process but also includes “**subject effect**” (age, gender, etc...). Sometimes these subject effects are large making it difficult to assess “**treatment effect**”.

Blocking (to turn a completely randomized design into a randomized complete block design) would help. But it would only help to “reduce” subject effects, not to “eliminate” them: subjects in the same block are only similar, not identical – unless we have “blocks of size one”. And that the basic idea of “Cross-over Designs**”, a very popular form in biomedical research.**

In the most simple cross-over design, subjects are randomly divided into two groups (often of equal size); subjects in both groups/series take both treatments (experimental treatment and placebo/control) but in different “orders”.

Group 1: Period #1 (Treatment) – washout – Period #2 (Placebo)

Group 2: Period #1 (Placebo) – washout – Period #2 (Treatment)

Of course, “order effects” and “carry-over effects” are possible. And the cross-over designs are not always suitable. They are commonly used when treatment effects are not permanent; for example some treatments of rheumatism.

**DESIGNING CLINICAL RESEARCH:
NUTS & BOLTS OF THE BASICS**

How data will be collected? (It's the complex issue of when to do what):

- ❖ Cross-section Design,
- ❖ Case-Control Design (retrospective),
- ❖ Cohort Design (prospective); Clinical Trials (one-arm, two-arm; open-label or randomized; double or triple blind)

CROSS-SECTIONAL DESIGNS

In a cross sectional study, investigators draw a sample, randomly, from the population, then make all measurements for all variables on a single occasion- or within a very short period of time – without a follow up. They study distributions the variables within that sample; sometimes designating predictor and outcome variables based on “biological plausibility”, then correlating one to the other.

Example:

In the National Health and Nutrition Examination Survey (NHANES), a sample designed to represent the U.S. population is interviewed and examined. These surveys have been carried out periodically and all data are available for public use. They make up a major source of information about the health and the habits of the U.S. population; one could obtain estimates such as prevalence of smoking or a disease.

In addition to studying distributions and obtaining parameter estimates, cross-sectional studies can also be used for examining associations. For examples, a cross-sectional finding in NHANES III is an association between childhood obesity and hours watching television. The choice of which variables to label as predictors and which as outcomes depends on the cause-and-effect hypotheses of the investigator rather than on the study design.

Serial Surveys:

These form a special case. A series of cross-sectional studies of a single population observed at several points in time – the case of those NHANES – is sometimes used to draw (informal) inferences about changing patterns of population characteristics over time.

Strengths:

A major strength of cross-sectional studies is that there is no waiting time for the outcome to occur, and no loss to follow-up. This makes them fast and inexpensive. And the obvious strength of their sizes. A cross-sectional study, because of its low cost, could be included as the first step in a cohort study or an experiment.

Weaknesses:

The major weakness of cross-sectional studies is the difficulty of establishing causal relationships from “observational” data collected in a cross-sectional time frame.

Cross-sectional studies are also “impractical” for the study of rare diseases if the sample was collected from the general population; we might need 10,000 subjects or more to find just one case of a rare disease. What would happen to statistical power?

CASE-CONTROL STUDIES

Case-control studies “began” as epidemiologic studies to try to identify risk factors for diseases. Therefore, the term “cases” refer to those with the disease under investigation. However, the term has become more generic; case-control design can also be used to look at other outcomes, such “disability” among those who already have a disease. In that case, “controls” are those with the disease but not disability. The “case” in a case-control study maybe a patient who has had a good but rare outcome; say, recovery from a usually fatal disease.

Some investigators and scientists refer to case-control studies as “confirmatory observational studies” and to cross-sectional studies as “exploratory observational studies”. Both are observational; without interventions. Case-control studies are “**retrospective**”; obtaining “past” data from cases and from controls. **The research focus is the disease** (‘status’ under investigation).

STEPS IN THE DESIGN

- **Select the sample of cases**
- **Select the sample of controls**
- **Measure predictor variable & potential confounders and effect modifiers**

Case-control studies cannot yield estimates of the incidence or prevalence of a disease because the proportion of the study subjects who have the disease is determined by how many cases and how many controls the investigator chooses to sample, rather than by their proportions in the population.

What case-control studies do provide is descriptive information on the characteristics of the cases and, more important, an estimate of the strength of the association between each predictor variable and the presence or absence of the disease. These estimates are in the form of the odds ratio which approximates the Relative Risk if the disease is relatively rare.

Strength: Efficiency for Rare Outcomes

A major strength of case-control studies is their rapid, high yield of information from relatively few subjects. For rare diseases or outcomes, cross-sectional design is impractical; it requires a study size that no investigator could afford. Cohort design, which is normally larger, is also impractical because it requires a follow-up time longer than most investigators could afford.

Strength: Generating Hypotheses

The retrospective approach of case-control studies, and their ability to investigate a large numbers of possible predictors make them useful for **generating hypotheses** about, say, the causes of new outbreak of disease so that a more thorough investigation or investigations could follow.

Weaknesses:

Case-control studies have great strengths but they also have major limitations; among them:

- (1) Accuracy of exposure histories.**
- (2) Appropriateness of controls.**
- (3) Unlike cohort designs, we can only study one disease/outcome at a time.**

Sampling Bias:

The data collection in a case-control study begins with the **cases**. But how do we know if these cases are representative of all patients who developed the disease; those who are undiagnosed, or died, are not included. And some included might be misdiagnosed (easier problem here). The more difficult decisions faced by investigators of a case-control study, however, relates to the more open-ended task of selecting **controls**. The ones included might be inappropriate and confound study conclusion.

Hospital- or Clinic-based Controls

One strategy to compensate for the possible selection bias caused by obtaining cases from a hospital or clinic is **to select controls from the same facility**. However, the risk factor of interest might be related to causes for which those control seek care; if so, prevalence of the risk factor in the control group would be falsely high, biasing the study results toward “the null”. That’s why, some studies use **two control groups**.

Matching & Multiple-matching:

Matching, and multiple matching, is a relatively simple method of ensuring that cases and controls are comparable with respect to major factors that are related to the disease but are not interest to the investigators. Examples are gender and Age group. **Matching does have its adverse consequences, especially when “modifiable factors”, such as income and cholesterol levels, are matched.**

Population-based Cases:

Population-base case-control studies are now possible for many diseases, like cancers, because of a rapid increase in the creation of maintenance of “**disease registries**”. Cases obtained from disease registries are generally more representative of the population of patients. When information on the cases and controls can come from the same sources, the design has the potential for eliminating sampling bias. Later, we will cover such a form, the “**nested case-control design**”.

Two or more Control Groups:

Selection of controls can be very tricky, especially might not be representative of patients; for example, hospital-based cases. In those studies, it is advisable to **use more than one control groups**; for example a hospital-based and a population-based control groups. The former might be biased toward the Null. On the other hand, be prepare to deal with “**multiple decision problem**”.

Measurement Bias: The Need for Blinding

Besides selection bias, case-control studies might be biased due to measurement error caused by their retrospective approach: the “recall bias” especially when it occurs to a different extent in cases and in controls. A necessary solution is the need for blinding: **both observers (interviewers, for example) and study subjects could be both blinded to the case-control status of each subject and to the risk factor being studied (not an easy task at all!).**

Case-crossover Studies:

A variant of the case-control design, useful for the short-term effects of varying or intermittent exposures, is the “case-crossover design”. As with the “regular” case-control studies, case-crossover studies are retrospective studies that begin with a group of cases. However, in regular case-control studies, the exposures of the cases are compared with exposures of a group of controls. Each case in case-crossover studies serves as his/her own control. Exposures of the cases at or right before outcome time are compared with exposures of those same cases at one/more other points in time.

TASKS FOR ENHANCING CAUSAL INFERENCE

There is a sub-field of statistics/biostatistics called **“Causal Inference”**; but that’s part of **“Data Analysis”**. There are a few things we can do in the **“Design Stage”** to enhance results of causal inference.

Suppose that a study reveals an “association” between coffee drinking and myocardial infarction (MI). There are 5 possibilities: (1) Coffee drinking and MI are not related; what revealed was a chance finding (random error); (2) Coffee drinking and MI are not related; what revealed was caused by some bias, systematic error; (3) MI is a cause of coffee drinking, a so-called “Effect-Cause” phenomenon; (4) Coffee drinking is associated with a third extrinsic factor, called confounder, and the confounder is a cause for MI; and (5) Coffee drinking is a cause for MI; this is the real Cause-Effect phenomenon (bingo!) – the ideal possibility .

Before reaching the ideal conclusion – something the investigators wanted to prove, the other four rival possibilities have to be considered and ruled out. How? What can or should we do in the design stage?

Strategies for addressing random errors are available in both design and analysis stages. In data analysis, you focus on “statistical significance” (p-value). The design strategies include: (1) Increasing the “precision” of measurements, and more important, (2) Increasing sample size. So, sample size estimation is needed not just for budget justification!

Ruling out spurious associations due to bias is trickier, more difficult. Here are 3 basic questions to ponder: (1) Do the samples really represent the target populations; (2) Do the measurements of the predictor variables really represent the predictors of interest (the issue of randomization included here); and (3) Do the measurements of the outcome variables really represent the outcomes of interest (the use of surrogate markers included here).

Strategies for coping with confounders require that investigators be aware of, be able to measure, and use them. The most common way to “use” them is matching – especially for factors which are not easy to quantify for use in data analysis (e.g. geographical factor). However, be cautious, you might overdo it! Overmatching can reduce statistical power and making it more difficult to generalize the findings.

About the only way to rule out “Effect-Cause” possibility from an observational study, the possibility that “the cart has come before the horse”, is **to follow up with a cohort, longitudinal study** – as outlined in an earlier section of this lecture, “Natural History of Research”.

DESIGNING A COHORT STUDY

“Cohort” was the Roman term for a group of soldiers that march together. In clinical research, a cohort is a group of subjects followed over time. In itself, the term **“cohort”** does not yet mean **“prospective”**. In the design terminology, we have **“prospective cohorts”** but we also have **retrospective cohorts** which may appear under **“nested case-control”** or **“case-cohort”** options.

PROSPECTIVE COHORTS

In a prospective cohort study, the investigator:

- ❖ **Selects a sample from a target population;**
- ❖ **Measures (baseline) values of predictor variables;**
- ❖ **Measures the outcomes during follow-up**

In the most simple case, one binary predictor: presence or absence of a risk factor and for the outcome, whether a disease occurs. This type of design is prospective & longitudinal.

An Example:

- (1) In 1976, investigators obtained lists of registered nurses aged 25 to 42 in the most populous states and mailed them an invitation to participate in the study; those who agreed became the cohort;**
- (2) They mailed a questionnaire about weight, exercise, and other potential risk factors; they obtained 121,700 completed questionnaires, that's the size of the cohort;**
- (3) They send periodic questionnaires about the occurrence of a variety of disease outcomes, heart diseases and cancers included.**

Some Results:

The investigators succeeded in following 95% of the nurses and 1,517 cases of breast cancer were confirmed during the next 12 years. They found that, for example, women who gained more weights have a higher risk of breast cancer after menopause; those who gained more than 20 kg since age 18 had a twofold increased risk of developing breast cancer.

Strengths:

- (1) Suitable for assessing “**disease incidence**” (new cases); helpful in investigating potential “**causes**” because cohort members were free of the disease under investigation to start with;
- (2) Measurements of predictors are not influenced by knowledge of the outcome;
- (3) Prospective approach allows investigators to measure variables more completely and more accurately, to update the status of risk factors – especially important for “**time-dependent**” covariates; the large size of the cohort and long period of follow-up provide substantial “**statistical Power**”.

Weaknesses:

- (1) Cohort studies, even prospective cohort studies, are basically “**observational**”; causal inference could be challenged and interpretation often muddied by potential influences of confounders and effect modifiers;
- (2) **Time and cost consuming.** It could be more feasible if outcomes are more common and immediate; for example, a prospective study of risk factors for progression (or relapse) after treatment of patients with breast cancer.

RETROSPECTIVE COHORTS

A retrospective cohort differs from a prospective cohort in that the assembly of the cohort, baseline measurements, and follow-up all happened in the past. It was assembled for other purposes; however, important data about risk factors are still possible to be obtained for the new purpose – for example, from banked blood samples.

Design of Retrospective Cohorts:

The Investigator:

- (1) Identifies a cohort that has been assembled;**
- (2) Collects data on predictors (“measured” in the past);**
- (3) Collects data on the Outcome (measured in/at the present).**

Example #1:

To study thoracic aortic aneurysm, investigators:

- (1) Search the database of Olmsted County, Minnesota – which is considered a cohort because of thorough medical records of its residents – and found 133 cases of aneurysm;**
- (2) They reviewed patients' records to collect data on age, size of aneurysm, and other factors of cardiovascular diseases at the time of diagnosis;**
- (3) For the outcomes, they collected data from the medical records of these 133 patients to determine whether the aneurysm ruptured or was surgically repaired**

Example #2:

The Singapore Cohort was drawn from residents in government-built housing estates (roughly 86% of the population resided in such facilities); enrollment period was 1993 - 1998. Men and women between the ages of 45 and 74 years (35,298 were women), representing 85% of eligible subjects, were enrolled. At the time of recruitment, each cohort subject was interviewed in-person using a structured **questionnaire** that focused on current diet. **Blood samples** were requested and a total 28,346 blood samples were archived and banked. To date, only <0.05% of subjects are lost to follow-up.

To date, there are 304 incident breast cancer cases with a stored blood sample, and a study is proposed to investigate the roles of some genetic factors and diets as possible risk factors/protectors for breast cancer.

Strengths and Weaknesses:

Retrospective cohorts have many of the same strengths as prospective cohorts and they have the advantage of being less costly and less time consuming.

The main disadvantages are the limited control investigators could have over the nature and the quality of data; existing data on predictors could be incomplete (too late now!) and not ideal for answering the research question.

NESTED CASE-CONTROL

A nested case-control design is a case-control study “nested” within a cohort study.

Investigators begin with a suitable cohort having enough cases (to assure adequate statistical power) to answer the research question. Then, they select a random sample of the subjects who have not developed the outcome/disease under investigation (the controls); they could increase the power by selecting two or three controls matched to a case

Example:

Back to the Singapore cohort assembled in 1993-1998. By the end of 2011, there are 304 incident breast cancer cases with a stored blood sample, and a study is proposed to investigate the roles of some genetic factors and diets as possible risk factors/protectors for breast cancer. The proposed design was a 2-to-1 matched case-control study of roughly 900 women with the following Specific Aims (there are more aims):

- (1) Investigating the T-reg, T-cell and NK cell levels as a risk factor for breast cancer occurrence;**
- (2) Correlating T-reg, T-cell and NK cell levels with diet factors (from baseline interview) focusing on soybean products and green tea.**

NESTED CASE-COHORT

The nested case-cohort option is almost the same design as the nested case-control except that the “controls” are a random sample of all the members of the cohort “regardless of outcomes”. This means there might be some cases among those sampled for the comparison group; these cases appear in both groups. This approach has the advantage that the “controls” (even some of them are cases) represent the cohort in general, and therefore provide a basis for estimating incidence and prevalence in the population from which it was drawn.

Strengths and Weaknesses:

Nested case-control and nested case-cohort designs are especially useful for costly measurements on serum, electronic images (MRI and mammograms), and hospital charts, etc... that have been archived at the assembly time of the cohort and preserved for later analysis.

When data are available, or can be obtained easily, for the entire cohort, nothing is gained by studying only a sample; the whole cohort should be used.

MULTIPLE-COHORT STUDIES

Multiple-cohort studies begin with two or more groups of subjects; typically, one group with no exposure to a potential risk factor and one or more other groups with different levels of exposure. This is different from case-control design because in a case-control study the two groups are chosen based on the presence or absence of the outcome.

Multiple-cohort design is particularly useful and popular for studying rare exposures such as occupational and environmental hazards.

ISSUES WITH COHORTS

The hallmark of a cohort study is the identification of a group of subjects at the beginning of a period of follow-up:

- ❖ Subjects should be appropriate to the research question;
- ❖ Subjects should be available for follow-up;
- ❖ Subjects should resemble to the population to which the results will be generalized;
- ❖ Number of subjects should provide adequate (statistical) power.

The quality of the study (and future studies) will depend on the precision and accuracy of the measurements of predictor (s) and outcome variable.

The ability to draw inferences about cause and effects will also depend on the degree to which investigators have identified and measured all potential confounders and effect modifiers.

Predictors may change during the follow-up; whether and how frequently measurements should be repeated depends on how they are likely to change and, of course, depends on the cost and the importance to the research question of observing these changes.

Outcomes should be observed/assessed using standardized criteria and, ideally, blindly without knowing the values of the predictor variable.

Follow-up of the entire cohort is important; investigators should take a number of steps to achieve this goals; for example:

- ❖ Exclude those likely to be lost, or collect adequate information (physician or friends or relatives) that can be used if they move or die;**
- ❖ Prepare for periodic contacts (by mail, by phone, etc...);**
- ❖ Show respect and appreciation!**

DESIGNING A RANDOMIZED TRIAL

In clinical trials, investigators apply an “intervention” and observe the effect on outcomes. The major advantage is the ability to demonstrate causality; in particular: (1) random assigning subjects to intervention helps to reduce or eliminate the influence of confounders, and (2) blinding its administration helps to reduce or eliminate the effect of biases from ascertainment of the outcome.

Of course, not every research question is amenable to the clinical trial design. For example: (1) By ethical reasons, we cannot assign subjects to smoking in a trial in order to learn about its harmful effects, or (2) It is not feasible to study whether drug treatment of high LDL-cholesterol in children will prevent heart attacks many decades later.

In addition, clinical trials are generally expensive, time consuming, address narrow clinical questions, and sometimes expose participants to potential harm. For these reasons, clinical trials are best reserved for relatively “mature” research questions, and when observational studies strongly suggest that an intervention might be effective and safe. Even then, ones should learn to conduct these major studies in a responsible way to ensure success and to protect participants.

Intervention and Control:

The choice and dose of intervention is a difficult decision that balances effectiveness and safety; other considerations include relevance to clinical practice, simplicity, suitable for blinding, and feasibility of enrolling subjects. These are often results of a long process of “early phase clinical trials”. The best control group is a **placebo control** that allows participants, investigators, and study staff to be blinded. The strategy of using place compensates for any psychological effects so that any outcome difference between study groups can be ascribed to a biological effect.

Measurements:

Clinically relevant outcome measures such as resolution of the disease/condition, pain, quality of life, occurrence/relapse of cancer, and death are the most meaningful outcomes of trials. Sometimes, investigators have choice but to rely on intermediary such as bone or breast density, HIV viral load. These intermediary markers are valid surrogate markers for clinical outcomes to the degree that treatment-induced changes in the marker consistently predict changes in the clinical outcomes.

All clinical trials should include measures of potential adverse effects of the intervention – even plan, called “Stopping Rule” to stop the trial when it goes wrong, as measured by these adverse effects.

Selecting Participants:

In a clinical trial, inclusion and exclusion criteria, together, govern the selection process. The criteria for selecting participants should identify those who are likely to benefit and not be harmed by the treatment, easy to recruit, and likely to adhere to treatment and follow-up protocols. On the other hand, criteria also maximize our ability to generalize the findings from the trial to target population. For example, choosing only participants at high risk of an uncommon outcome can decrease sample size and cost, but may make recruitment more difficult and decrease our ability to general the findings.

Baseline Data:

Even though, in theory, randomization is supposed to eliminate the problem of confounding by factors that are present at the outset, and a lot of measurements adds expense and complexity, baseline data are important in many trials. If outcomes include change in a variable, the outcome variable must be measured at the beginning of the study in the same way that it will be measured at the end. And ones can check to see if randomization works well, or ones should back it up in the analysis – say, using “regression”.

Randomization and Blinding:

Randomization, which eliminates bias due to confounding variables, should be “tamperproof”. Thoroughly consider special randomization techniques: Blocked randomization, Stratified randomization, etc... Blinding the intervention is as important as randomization and serves to control bias through outcome ascertainment and adjudication. Consider “double blind”, “triple blind” features.

COMPARISON OF DIAGNOSTIC TESTS

If we want to compare two diagnostic tests, say tests for diabetes by measuring sugar levels from urine and from blood, or more often, **a new versus a more established one for certain disease**, we could proceed like conducting a clinical trial and decision is made through a “statistical test”. In addition to the comparison; the more difficult problem is how “express” the **“level of difference”** if the two screening tests do not have the same performance.

Designs:

- If each study subject is tested by all tests, we refer to as “**paired design**”, even more than two tests could be considered.
- If each study subject is tested by one test, we will refer to the design as “**unpaired**”.

USE UNPAIRED DESIGNS

- If tests are **invasive**, cause the patient discomfort, time consuming, or have significant risk associated with them,
- If ethical considerations require unpaired designs be made **to minimize the burden on the subjects.**
- If the performance of one test might interfere with the implementation and/or the result of another; for example, two surgical procedures.

UNPAIRED DESIGNS

- Follow the same design principles of multi-arm randomized clinical trials.
- Those include **well-defined inclusion-exclusion criteria**, clear apriori definition of disease and test result - including measurement scale, preparation of study protocol, and randomization to ensure that study arms are balanced with regards to factors affecting test performance and/or result; analysis plan must be in place.
- **Blinding** – if feasible- may be needed to ensure integrity of disease and test assessments.

COMPARISON OF TESTS WITH BINARY ENDPOINT

We can perform two separate Chi-square tests, one for cases and one for controls; for an overall level of α , each test is performed at $\alpha / 2$

MEASURING DIFFERENCES

- If the difference between two diagnostic tests are found to be significant; the level of difference should be summarized and presented.
- The two commonly used parameters are the ratio of two sensitivities (RS^+) and the ratio of two specificities (RS^-); these are ratio of independent proportions.

PAIRED DESIGNS

- If feasible, paired designs are more desirable.
- Most important, only valid if tests do not interfere with each other; be cautious because interference can be subtle.
- Also paying attention to cooperation of the subjects; “order” should/may be randomized.

ADVANTAGES

- **More efficient because impact of between-subject variability is minimized.**
- **Possibilities of confounding are eliminated,**
- **One can examine characteristics of subjects where tests yield different results; this can lead to insight about test performance and, sometimes, strategies for improving tests.**
- **One can assess value of combinations of tests compared to single tests.**

COMPARISON OF TESTS

We can perform two separate McNemar's Chi-square tests, one for the set of cases and one for the set of controls.

MEASURING DIFFERENCES

- If the difference between two diagnostic tests are found to be significant; the level of difference should be summarized and presented.
- The two commonly used parameters are still the ratio of two sensitivities (RS^+) and the ratio of two specificities (RS^-). However, these are no longer ratio of independent proportions, the method becomes a little more complicated.

EXERCISES

#1. Suppose we want to compare the use of medical care by black and white teenagers. The aim is to compare the proportions of kids without physical check-ups within the last two years. Some recent survey shows that these rates for blacks and whites are 17% and 7% respectively. How large should a total sample be so that it would be able to detect such a 10% difference with a power of 90% using a statistical test at the two-sided level of significance of .01?

#2. A study will be conducted to investigate a claim that oat bran will reduce serum cholesterol in men with high cholesterol levels. Subjects will be randomized to diets that include either oat bran or cornflakes cereals. After two weeks, LDL cholesterol level (in mmol/L) will be measured and the two groups will be compared via a two-sample t test. A pilot study with cornflakes yields

$$n = 14; z = 4.44, s = 0.97$$

How large should a total sample size be if we decide to preset $\alpha = 0.01$ and that it is important to detect an LDL cholesterol level reduction of 1.0 mmol/L with a power of 95%?