Session 2: Probability & Study Design

Li (Sherlly) Xie

Session 2 Flow

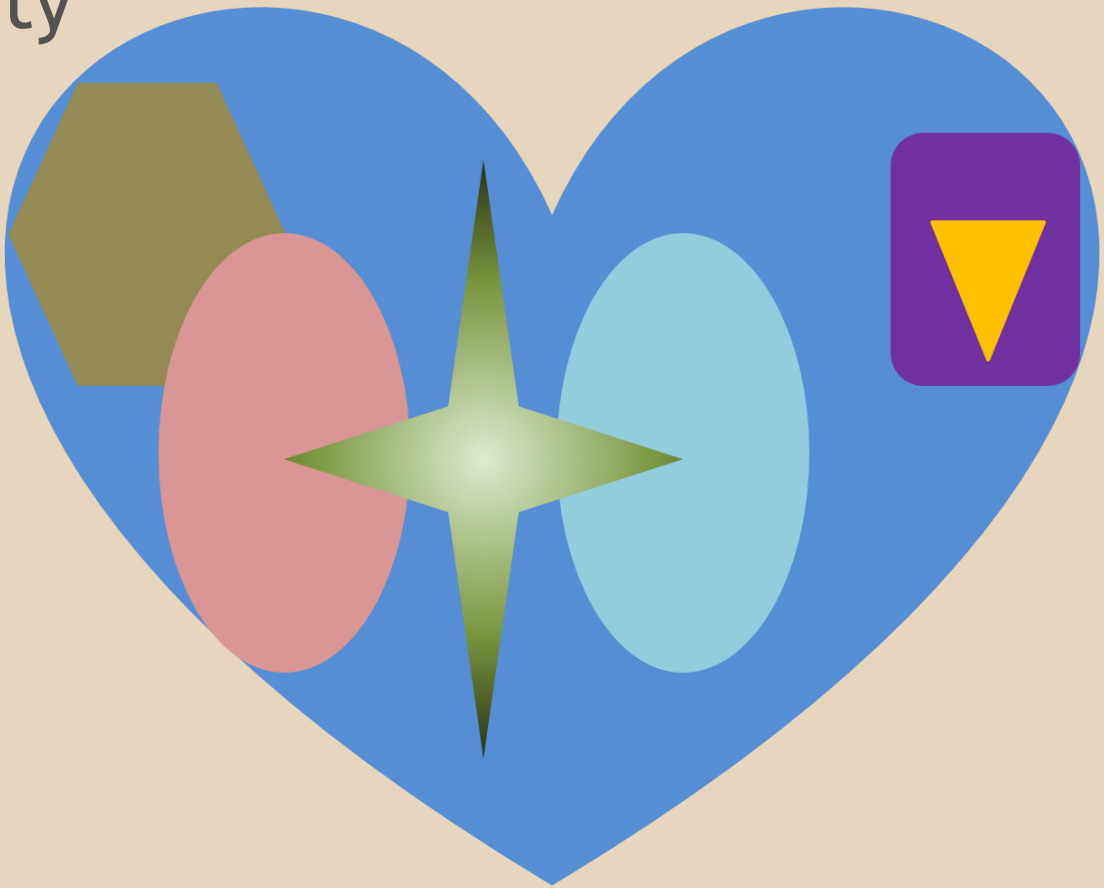
Probability Principles

Break

Study Design

Probability Concepts

Law of total probability
Conditional probability
Intersection
Union
Mutual exclusivity
Subset
Exchangeability
Complimentary sets



Complement & Union, Illustrated

blue: children age 0-18

lighter blue: having family history of asthma

4-horn star: having clinically diagnosed asthma

pink: overweight (at or above 95th age & gender adjusted pop percentile)

brown: high-energy low-nutrient diet

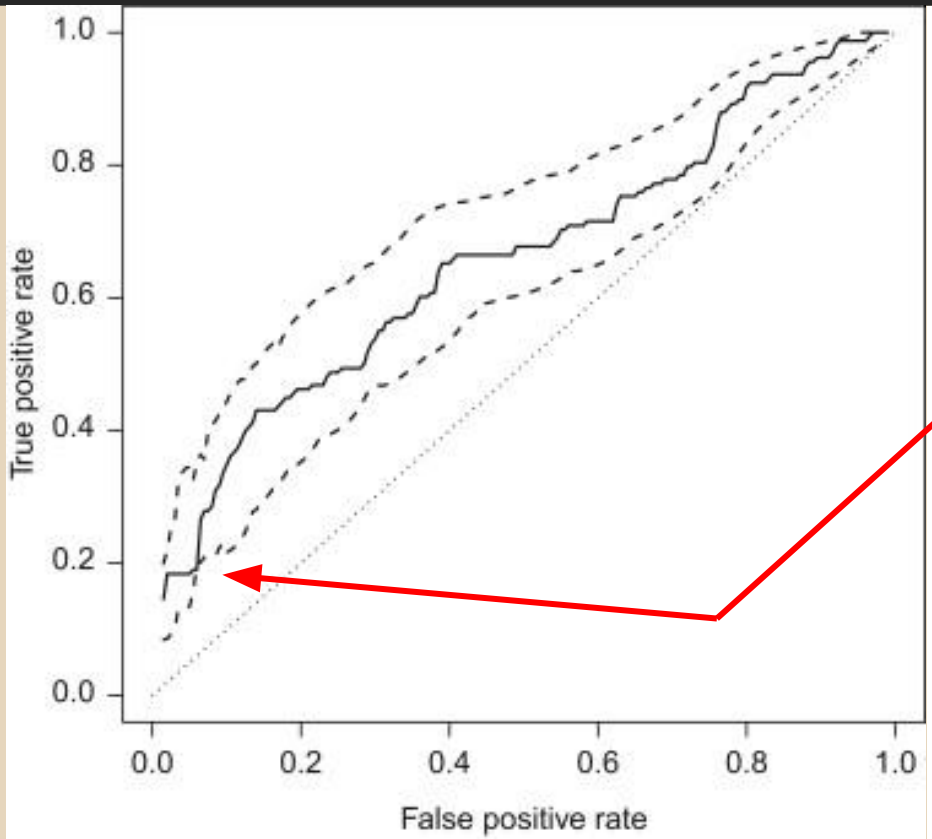
purple: underweight (below 30th perc)

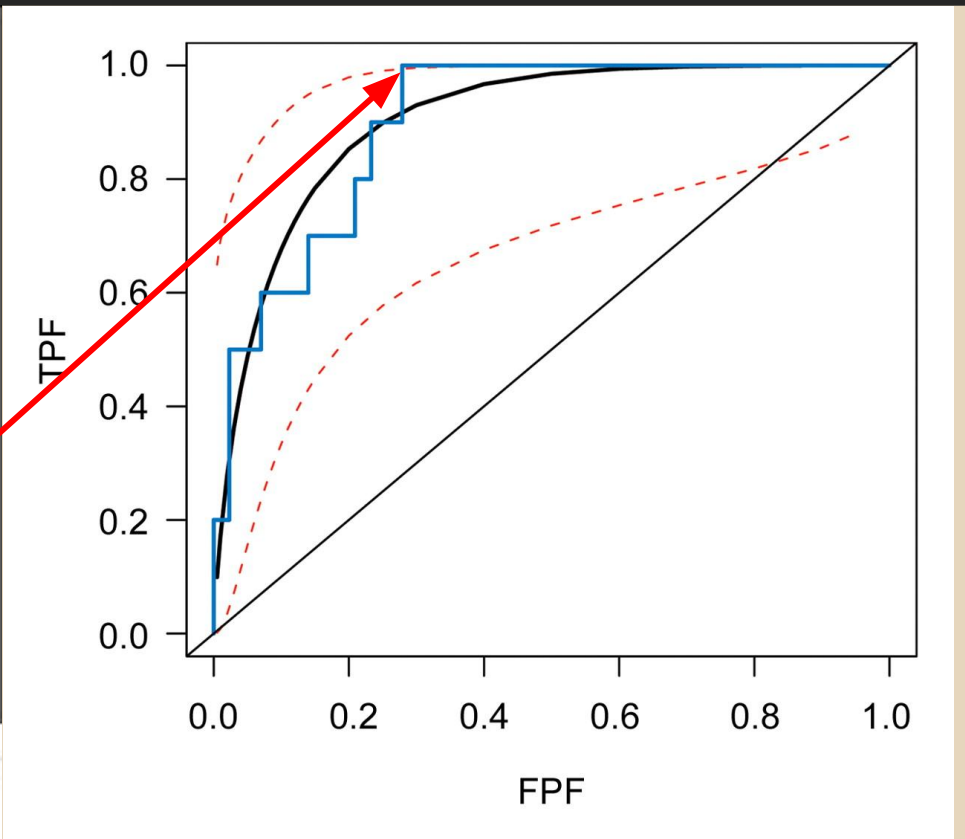
orange: severely underweight (below 15th perc)

Now answer these questions

1. Characterize the entire population
2. Is it possible to tell where the two ovals SHOULD belong or their intersections with the horns of the star based solely on their positions?
3. Estimate the % overlap between the severely underweight and underweight individuals and the % overlap between the severely underweight and the whole population of individuals. Are the 2 percentages the same? If you are to sample a severely underweight individual, from which population--the entire population or the underweight population--are you more likely to encounter one?
4. What else do you observe? What are some other questions about this population you could ask?

What about the EDGE?





Respiratory Research (IF 3.36)

Radiology (IF 5.73)

Statistics-friendly research aims/hypotheses

Use words like "explore" and "describe" with "the association between X and Y"Numbers (appearances) answer the DICHOTOMOUS QUESTION: to be or not to be

A pair of compatible hypotheses pertains to dichotomy AND ONLY dichotomy

But we ask incompatible questions all the time, because we think numbers are statistics

Example

Typical pairs of hypotheses:

Null (H0): drug effective at delaying cancer recurrence

Alternative (Ha): drug ineffective at delaying cancer recurrence

COMPLIMENTARY but incompatible:

there are a million different ways to be effective and ineffective

Questions to address during study design

- How do you define and measure your outcome (dependent variable) and other variables in your data (independent variables)?
- How do you account for potential confounders that could cripple your inference?
- How do you sample?
- What EXACTLY do you wish to know & how flexible can you be with your aim?
- This is what a statistician cannot tell you--use your medical expertise & pubmed.

Variations of... the same aim

Original aim: which factor is ASSOCIATED with the lowest # of occurrences of disease X (ie the "protective" factor)?

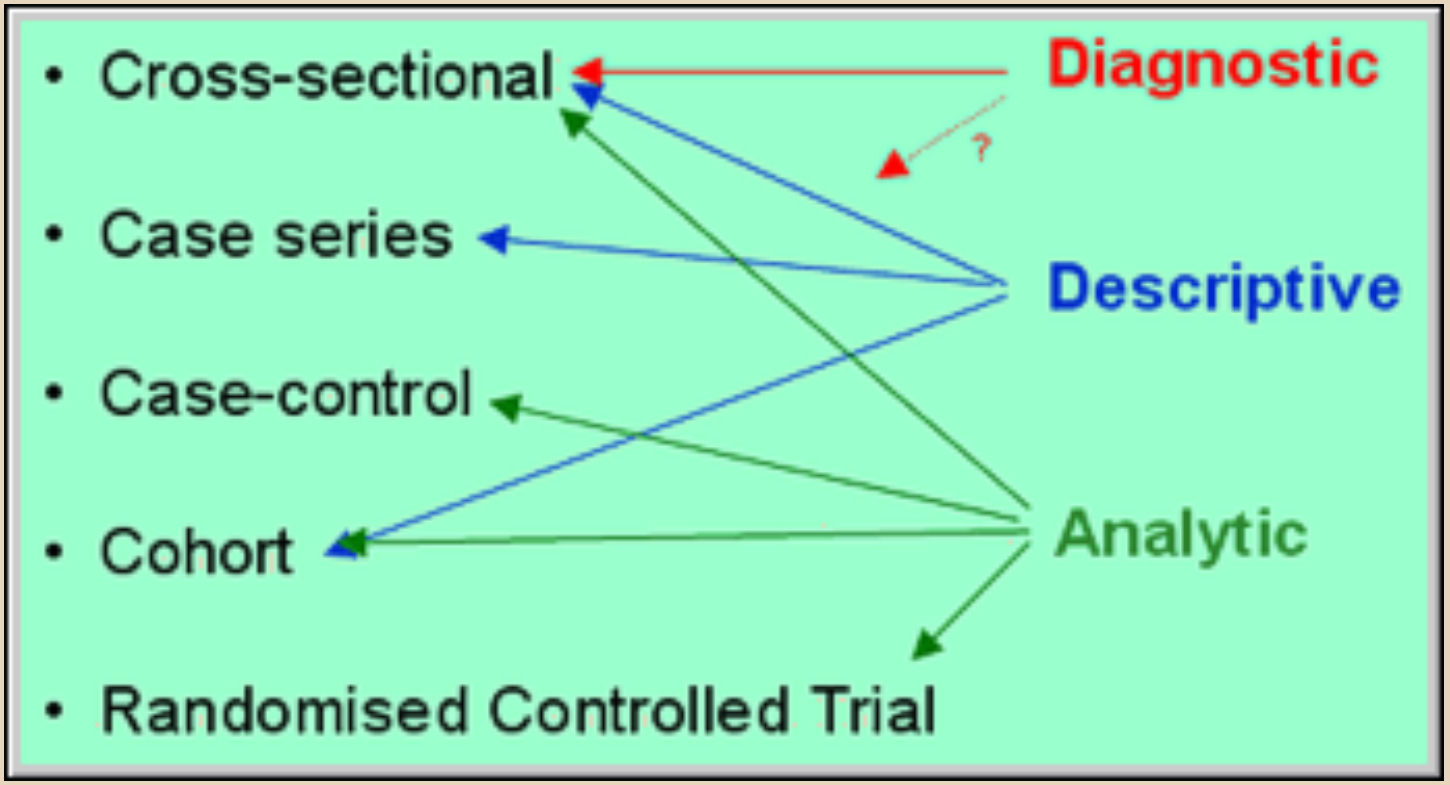
Variation 1: ... occurrences of stomach pain?

Variation 2: ... lowest concentration of pathogen _____?

Variation 3: ... least # of ER visits by cases within 3 days after exposure?

There are many ways to express the idea of "protective"

Study Designs



emergencymed.co.za/TRRM/module3

Strengths

Case control

Good for studying rare conditions, fast, could examine multiple risk factors simultaneously, useful as initial studies to establish associations

Cohort

Each cohort is relatively homogeneous (subjects can be matched to limit confounding, results could be standardized)

Randomized controlled trial

Randomization reduces the probability of population bias, easier to blink/mask than observational studies, clearly defined study population, data obtained under this design are in better agreement with the assumptions hypothesis tests and statistical models make about the observations.

Weaknesses

Case control studies:

Recall bias; control may be hard to find, no causal inference, cannot be used to evaluate diagnostic test

Cohort:

Cohort identification may be difficult due to confounding variables, no causal inference, expensive (longitudinal), possible larger loss to follow-up

Randomized controlled trial:

Expensive, volunteer bias (representativeness of sample and generalizability of results), hard to reveal causation, loss to follow-up

Brief Comment on Systematic Review and Meta-Analysis

Systematic review: Exhaustive review of the current literature and other sources.

Meta-Analysis: A subset of systematic reviews; A method of systematically pooling qualitative and quantitative results across studies to develop a single conclusion.

Advantage: great statistical power (large overall sample size), more reliable than single studies, meta-analytic results generalizable to a greater population.

Disadvantage:

Heterogeneity of study populations;

Definition of the same outcome may vary

Outcome may be measured in different ways (instruments, protocols, etc)

Quality of pooled evidence depends on the quality of each included study

Possible Designs

RCT: always prospective

Observational studies:

- prospective and retrospective cohort
- retrospective case-control
- cross-sectional

Recommended References:

1. Research Design Comparison/Contrast
2. National University of Health Sciences Overview of Study Designs in Clinical Research
3. University of Minnesota Libraries UNDERSTANDING RESEARCH STUDY DESIGNS

This week's task

Data organization. Due Tuesday 2/19/2013 NOON via e-mail (LI.XIE@NEMOURS.ORG)

If you missed hw1, please do both this week. hw1 can be obtained from Mrs Rhonda Carter

See you next Wednesday!