What to do today (March 2, 2023)?

Part I. Introduction
Part II. Epidemiologic Concepts and Designs
Part III. Clinical Trials

Part IV. Modern Biostatistical Approaches

Part IV.1 Incomplete Data Analysis

Part IV.2 Some Other Important Topics (Chp 8 - 18, Koepsell and Weiss, 2003)

- *IV.2.1 Measures of Risks*
- *IV.2.2 Measurement Error Revisit*
- *IV.2.3 Confounding and Its Control*
- **IV.2.4 Causation vs Association**

Part IV.3 Selected Widely-Used Algorithms

- *IV.3.1 Bootstrap and Related*
- *IV.3.2 EM Algorithm and Related*

Part IV.4 Lifetime Data Analysis

Part IV.2.4A Causation vs Association: Historical Development

Causality and Self-Experimentation

- In 1984, Barry Marshall, an Australian physician, swallowed *Helicobacter pylori* to prove it (not stress) causes stomach ulcers.

  H. Pylori: a Gram-negative, microaerophilic bacterium found in the stomach. (It was identified in 1982 by Marshall and Warren.)

  - He did get sick (although no ulcers developed).
  - Marshall and his colleague Robin Warren received the 2005 Nobel prize in medicine.
- Suppose that Marshall had developed ulcers - would this have proven a causal link between *H. pylori* and the development of stomach ulcers?

Part IV.2.4A Causation vs Association: Historical Development

Recall **Koch's Postulates** (Porta, 2008)

- Robert Koch (1843-1910) along with Louis Pasteur (1822-1895) was responsible for the development of the germ theory of disease which eventually replaced miasmatic disease theories.
- A specific disease agent (e.g. Mycobacterium tuberculosis) causes a specific disease (e.g. tuberculosis, TB) if:
  - The agent is present in every case of the disease by isolation in pure culture.
  - The agent must not be found in cases of other disease.
  - Once isolated, the agent must be capable of reproducing the disease in experimental animals.
  - The agent must be recovered from the experimental disease produced.

Part IV.2.4A Causation vs Association: Historical Development

H. pylori in peptic ulcer: have Koch's postulates been satisfied? (Marshall, 1995)

- In the first experiment an isolate of H. pylori was taken from an elderly man with nonulcer dyspepsia.
- By day 5 the volunteer was vomiting with symptoms largely resolved by day 14.
- Biopsy on day 8 showed heavy colonization.
- A second volunteer ended up being sick with gastritis for three years!

$\Longrightarrow$ Marshall concludes that "only one of the many steps required for the development of peptic ulceration has so far been fulfilled, i.e. the ability of H. pylori to produce histological gastritis in a susceptible host."

Part IV.2.4A Causation vs Association: Historical Development

Limitations of Koch's Postulates (Evans, 1978)

- Applies only to bacterial infections.
- Infection is, in almost all cases, a necessary but not a sufficient component for development of disease.
  For example, not all people exposed to the tubercle bacillus develop tuberculosis.
- The probability that an infected person develops tuberculosis increases (Dye, 2003) for
  - people co-infected with HIV,
  - people who are malnourished,
  - people with diabetes, and
  - smokers.

Part IV.2.4B Causation vs Association: Necessary and Sufficient Causes (Porta, 2008)

- **necessary cause:** a causal factor whose presence is required for the occurrence of the effect
  - If A is necessary for B, then B cannot occur without A being present
  - e.g. female for breast cancer

- **sufficient cause:** A set of conditions, factors or events needed to produce a given outcome
  - If A is sufficient for B, then if A is present, B will occur.
  - There may however be other factors which can also result in B occuring.

Part IV.2.4C Causation vs Association: Challenge of Establishing Causality

We observe associations but must infer causation.

For example, oral contraceptive use is associated with a higher risk of developing cervical cancer (Franco, 2003).

- Do oral contraceptives cause cervical cancer?
  Most women who use oral contraceptives never develop the disease.
- Is oral contraceptive use necessary for the development of cervical cancer?
- Is oral contraceptive use sufficient for the development of cervical cancer?

Part IV.2.4C Causation vs Association: Challenge of Establishing Causality

Randomized Controlled Trials and Causality

- Evidence for a causal relationship is generally considered greatest if obtained from a randomized controlled trial (Fletcher and Fletcher, 2005 page 201).
  - Greatest weight is given to double-blind trials with no missing obser- vations where all subjects followed their assigned interventions.
  - Disagreements are likely in practice given that most randomized trials do not provide unequivocal results.

Part IV.2.4C Causation vs Association: Challenge of Establishing Causality

- Causal relationships can never be proven but rather are a matter of judgment so disagreements are quite common (Aschengrau and Seage, 2003 page 377).

  For example, mammography screening to prevent breast cancer mortality has been quite controversial in spite of trials suggesting benefit (see e.g., Gotzsche and Olsen (2000), Duffy and Tabar (2000)).

Part IV.2.4D Causation vs Association: Assessing Causation

Bradford Hill's guidelines for assessing causation (Hill, 1965)
"Our observations reveal an association between two variables, perfectly clear-cut and beyond what we would care to attribute to the play of chance. What aspects of that association should we especially consider before deciding that the most likely interpretation of it is causation?"

- 1. Temporality
- 2. Strength of the association
- 3. Biological gradient
- 4. Consistency
- 5. Specificity
- 6. Plausibility

Part IV.2.4D Causation vs Association: Assessing Causation

1. **Temporality**
   - Temporality is a necessary but not a sufficient condition for causality.
   - Common logical error:
     If A precedes B, then A caused B $\Longrightarrow$ e.g. the roosters crow causes the sun to rise.

Example: Logical Error? Regression to the Mean?

- Patients screened to identify hypertensives.
- Hypertensives given a new, experimental medication are found to have lower blood pressure, on average, in the week after starting treatment.
- Did the medication cause blood pressure to go down?

Part IV.2.4D Causation vs Association: Assessing Causation

2. **Strength of the Associtation** (Greenland and Rothman, 2005)
   - Strong associations are less likely to be a result of selection bias, measurement error, omitted confounders or chance.
     - A weak association may of course still be causal.
       Cigarette smoking is believed to be a cause of heart disease; the association is weaker than for lung cancer.
   - A strong association may be due to confounding.
     - Down's syndrome prevalence at birth increases both with maternal age and birth order.
       The more modest increase with birth order likely reflects the greater age of mothers who have had several children.
       Chromosomal errors resulting in children born with Down's syndrome is a consequence of maternal age.

Part IV.2.4D Causation vs Association: Assessing Causation

3. Biological Gradient

- An association is more likely to be causal if the strength of association increases (or decreases) with the amount of exposure, i.e. dose-response effect.
  - For example, lung cancer incidence increases with number of cigarettes smoked per day.
- The absence of a biological gradient does not preclude a causal relationship.
  - Some exposures can cause nonlinear effects.
  - Moderate consumption of alcohol is protective against heart disease while never drinkers and heavy drinkers all have higher rates of heart disease (Ellison, 2007).

Part IV.2.4D Causation vs Association: Assessing Causation

4. Consistency

- An association is more likely causal if found in different studies, when using different study designs, or when carried out in different populations.
- Lack of consistency does not preclude causality.
  - There is strong evidence that human papilloma virus (HPV) is a necessary cause of cervical cancer (Franco, 2003).
  - There is also evidence that oral contraceptive use increases the risk of women infected with HPV to develop cervical cancer.
  - Consequently the absence of an association between oral contra- ceptive use and cervical cancer would be expected in a population of uninfected women.

Part IV.2.4D Causation vs Association: Assessing Causation

5. Specificity

- Developed originally in reference to infectious disease.
  - An exposure should cause a unique effect.
  - Rarely holds for non-infectious diseases, e.g. smoking causes heart disease, lung cancer etc.
- Specificity can still be useful in some studies.
  - Does screening for colorectal cancer using sigmoidoscopy reduce mortality?
  - Sigmoidoscopes do not reach the upper colon.
  - Screening sigmoidoscopy reduces mortality from tumors in the rectum and lower colon but not the upper colon (Koepsell and Weiss 2003, pp. 187-188).

Specificity (disambiguation) as a causal criterion has a different meaning than specificity of a diagnostic test.

Part IV.2.4D Causation vs Association: Assessing Causation

6. Plausibility, 7. Coherence, 8. Analogy [*often treated as interchangeable* (Goodman and Samet, 2006)]

- associations are more likely causal when supported by existing theory
  - Associations may be causal even when not so supported. Resistance to claims that H. pylori caused peptic ulcer was that the stomach was believed to be sterile.
- Analogy:
  - Comparison to a known causal relationship can bolster arguments that an observed association is causal.
    Thalidomide caused birth defects implies drugs with similar chemical properties might also cause birth defects.

Part IV.2.4D Causation vs Association: Assessing Causation

9. Experiment

- Animal experiments are useful for understanding how risk factors cause disease (Editorial, 2004), e.g. kittens were a model for examining the effect of oxygen on retinopathy of prematurity (Koepsell and Weiss, 2003 page 12).
- Experimental studies of people are neither ethical nor practical to evaluate the effect of most exposures.
- Randomized vaccine trials provided additional evidence that human pappiloma virus (HPV) causes lesions which can lead to cervical can- cer (Koutsky and Harper, 2006).
- Trials may have multiple interpretations, e.g. a cluster randomized trial of swamp draining could mislead as to the cause of malaria.

Part IV.2.4E Causation vs Association: Misperceptions of the Nature of Causation

- direct and indirect causes of disease: the direct causes are more important?
  e.g. cigarette smoking: "direct" due to our ignorance of the downstream consequences of the actions of that agent
- to be a cause of disease, an exposure must be present in every case?
  e.g. excessive alcohol consumption: a cause of some motor vehicle injury
- to be a cause of disease, an exposure must be capable of producing that disease on its own?
  e.g. infection with the tubercle bacillus: a cause of tuberculosis (TB)

Part IV.2.4E Causation vs Association: Misperceptions of the Nature of Causation

- A statistically significant association implies causality?

An association must be statistically significant to provide convincing evidence of causality; statistical significance is not, however, sufficient to establish causality.

Selection bias, measurement error, confounding and chance may result in spuriously statistically significant associations.

Part IV.2.4E Causation vs Association: Misperceptions of the Nature of Causation

- A statistically insignificant association implies the absence of a causal relationship?

  An association may not reach nominal levels of statistical significance (0.05)
  - if the sample size is too small,
  - due to a combination of selection bias, measurement bias, or data analytic bias (e.g. omitted confounders),
  - because the true association is too small to be detected (and of no practical interest),
  - if there truly is no causal relationship.

Part IV.2.4F Causation vs Association: Models of Causes in Epidemiology

- **Line of Causation:**
  disinguishing genetic and environmental factors
- **Triangle of causation:**
  agent, host, environment
- **Diseases are multifactorial:** web of causality

Part IV.2.4F Causation vs Association: Causal Inference (Tsiatis, 2006)

Statistical inference generally focuses on the associational relationships between variables in a population.

- Data are iid realizations of $Z \sim p_Z(\cdot; \phi)$, with ϕ the parameter(s) describing important features of the relationships

Causal relationships, "does 'A' cause 'B'?", can be of the primary interest: e.g. Does a treatment intervention or exposure at one point in time have a causal effect on subsequent response?

- How to formulate it from a statistical perspective? van der Laan and Robins (2003)

Part IV.2.4F Causation vs Association: Causal Inference (Tsiatis, 2006)

Point Exposure Studies

- Variables Involved:
  - r.v. A: the possible treatments/exposures can be given/experienced by an individual
    e.g. A=1 or 0 for a statin drug or not
  - r.v. Y: the response variable
    e.g. change in blood pressure after three months
  - r.v. X: other covariates, such as baseline covariates
- Study Goal: to establish a causal link between treatment and response

  Does treatment with the statin drug reduce blood pressure after three months as compared with no treatment?

Part IV.2.4F Causation vs Association: Causal Inference (Tsiatis, 2006)

- Data Collected: a study of indept $i = 1, \ldots, n$ subjects $Z_i = (Y_i, A_i, X_i)$

- Associational Analysis: $\mu_1 = E(Y|A = 1), \mu_0 = E(Y|A = 0)$ and $\Delta = \mu_1 - \mu_0$, the difference in mean responses for individuals receiving treatment 1 and treatment 0.
Δ is estimated by $\hat{\Delta} = \hat{\mu}_1 - \hat{\mu}_0$,

$$\hat{\mu}_1 = \frac{\sum A_i Y_i}{n_1}, \quad \hat{\mu}_0 = \frac{\sum (1 - A_i) Y_i}{n_0}$$

Part IV.2.4F Causation vs Association: Causal Inference (Tsiatis, 2006)

Discussing .....

- Has the causal question of interest been answered? How does $\hat{\Delta}$ work? What does Δ tell?

- What if the subjects receiving the statin are different from those who not?
  e.g. younger, wealther, or smoking less
  That is, if there're confounders? $\Longrightarrow$ the relationship between relationship and response is confounded (distorted).

$\therefore$ $\hat{\Delta}$ does not present the causal effect adequately.

Part IV.2.4F Causation vs Association: Causal Inference (Tsiatis, 2006)

Let's take a close look ... ...

- Assume a potential outcome $Y^*(a)$ for treatment $A = a$
- The causal (subject-specific) treatment effect is $Y^*(1) - Y^*(0)$
- $\delta = E[Y^*(1) - Y^*(0)]$ is the average causal treatment effect

However, it is impossible to measure both $Y^*(1)$ and $Y^*(0)$ from one individual.

Part IV.2.4F Causation vs Association: Causal Inference (Tsiatis, 2006)

If the observed response $Y = AY^*(1) + (1 - A)Y^*(0)$ [! still an assumption – Stable Unit Treatment Value Assumption (SUTVA) by Rubin (1978)],

$\Longrightarrow$ the desired (full) data are "coarsened" into (Y, A, X) from $(Y^*(1), Y^*(0), X)$

"A" serves as "R", the missing indicator

Part IV.2.4F Causation vs Association: Causal Inference (Tsiatis, 2006)

Randomization and Causality

It's been accepted that a randomized intervention study will result in an unbiased estimate of the average treatment effect with causal interpretations.

Formalizing this notion ... ...

Part IV.2.4F Causation vs Association: Causal Inference (Tsiatis, 2006)

Together with SUTVA, assume $A \perp \{Y^*(1), Y^*(0)\}$ [possible when the treatment is randomly assigned]

- the associational effect $\Delta = E(Y|A=1) - E(Y|A=0)$ is then the same as the average causal effect $\delta = E[Y^*(1) - Y^*(0)]$:

$$E(Y|A=a) = E[AY^*(1) + (1-A)Y^*(0)|A=a] = E[Y^*(a)]$$

- $\hat{\Delta} = \hat{\mu}_1 - \hat{\mu}_0$ is an unbiased estimator for Δ and for δ, too.

Part IV.2.4F Causation vs Association: Causal Inference (Tsiatis, 2006)

Observational Studies

The assumption of randomization, $A \perp \{Y^*(1), Y^*(0)\}$, may not be reasonable as subjects are not assigned to treatment but rather by choice. e.g. randomization of an exposure is infeasible/unethical, smoking?

If, together with SUTVA, we have $A \perp \{Y^*(1), Y^*(0)\} \big| X$: the treatment choice and the potential outcome are indept within the same stratum, $E[Y^*(a)]$:

$$E_X[E\{Y^*(a)|X\}] = E_X[E\{Y^*(a)|A = a, X\}] = E_X[E\{Y|A = a, X\}]$$

$\Longrightarrow \delta = E_X[E\{Y|A = 1, X\}] - E_X[E\{Y|A = 0, X\}]$, invoving only the distn of (Y, A, X).

Part IV.2.4F Causation vs Association: Causal Inference (Tsiatis, 2006)

Estimating δ:

- Regression Modeling: specifying $E(Y|A,X) = \mu(A,X;\theta)$
- Estimating θ: estimating euqation

$$\sum_{i=1}^{n} \frac{\partial \mu(A_i, X_i; \theta)}{\partial \theta} V^{-1}(A_i, X_i)[Y_i - \mu(A_i, X_i; \theta)] = 0$$

- Estimating δ: $\hat{\delta} = \frac{1}{n}\sum_{i=1}^{n}[\mu(1, X_i; \hat{\theta}) - \mu(0, X_i; \hat{\theta})]$

How to cast it as a coarsened data analysis?

What to study next?

Part I. Introduction
Part II. Epidemiologic Concepts and Designs
Part III. Clinical Trials

Part IV. Analytic Epidemiology

Part IV.1 Incomplete Data Analysis (*supplementary*)

Part IV.2 Some Other Important Topics (Chp 8 - 18, Koepsell and Weiss, 2003)

Part IV.3 Selected Widely-Used Algorithms

IV.3.1 Bootstrap and Related
IV.3.2 EM Algorithm and Related

Part IV.4 Lifetime Data Analysis