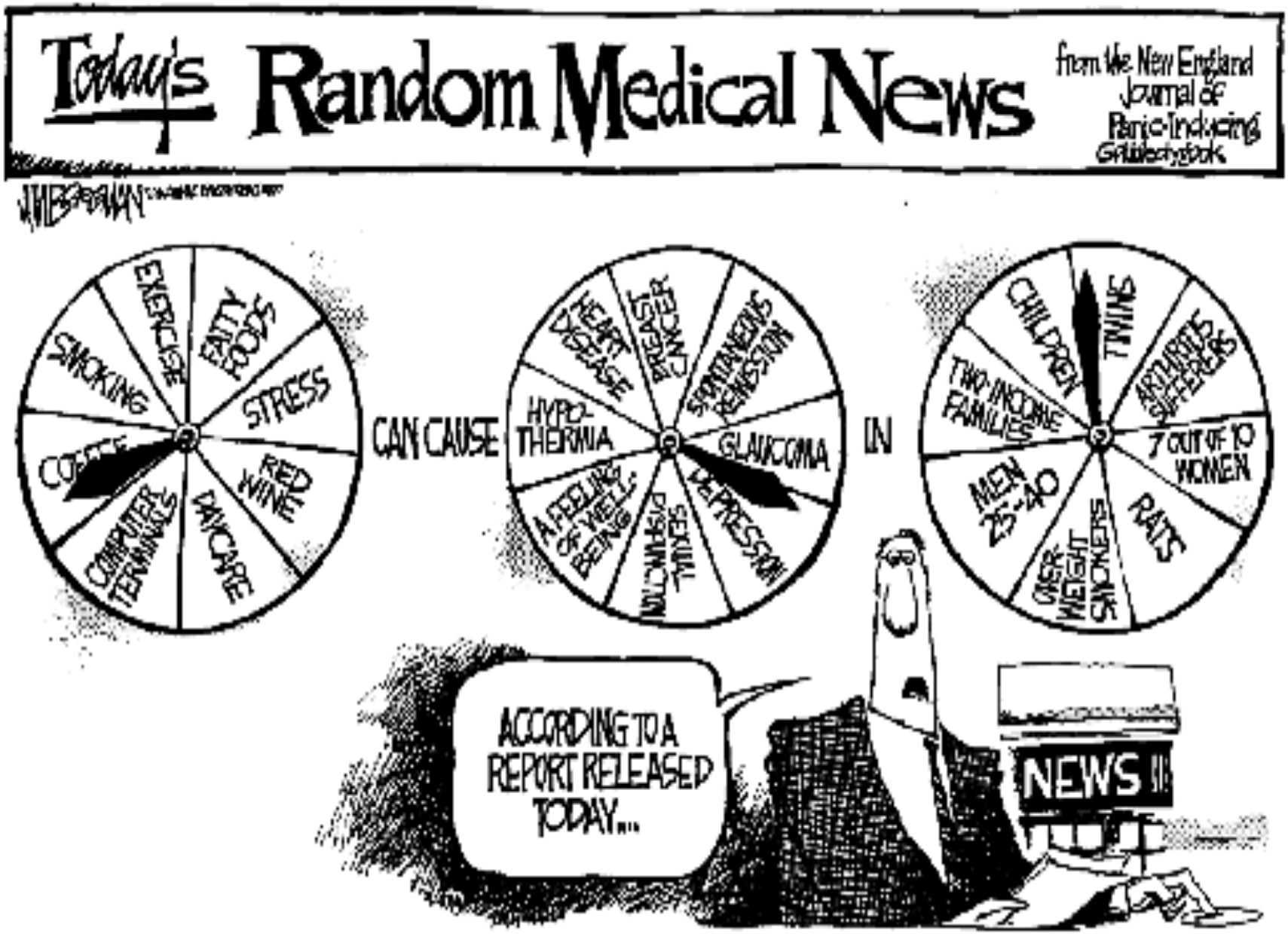
Short Course: Causal Inference

Introduction

Menggang Yu

Summer 2014


Today's Random Medical News
from the New England Journal of Panic-Inducing Gratuitous Info
EXERCISE
FATTY FOODS
STRESS
RED WINE
DAYCARE
COMPUTER TERMINALS
COFFEE
SMOKING
CAN CAUSE
HEART DISEASE
BREAST CANCER
SPONTANEOUS REMISSION
GLAUCOMA
DEPRESSION
SEXUAL DYSFUNCTION
A FEELING OF WELL-BEING
HYPO-THERMIA
IN
CHILDREN
TWINS
ARTHRITIS SUFFERERS
7 OUT OF 10 WOMEN
RATS
OVER-WEIGHT SMOKERS
MEN 25-40
TWO-INCOME FAMILIES
ACCORDING TO A REPORT RELEASED TODAY...
NEWS

RA Fisher

Le Fanu, James; The Rise and Fall of Modern Medicine

Twelve milestones of modern medicine

1. 1941: Penicillin
2. 1949: Cortisone
3. 1950: Streptomycin, Smoking and Sir Austin Bradford Hill
4. 1952: Chlorpromazine and the Revolution in Psychiatry
5. 1952: The Copenhagen Polio Epidemic and the Birth of Intensive Care
6. .....

Austin Bradford Hill

The Bradford Hill criteria for causation

A group of minimal conditions necessary to provide adequate evidence of a causal relationship between an incidence and a consequence:

- Strength: A small association does not mean that there is not a causal effect, though the larger the association, the more likely that it is causal.
- Consistency: Consistent findings observed by different persons in different places with different samples strengthens the likelihood of an effect.
- Specificity: Causation is likely if a very specific population at a specific site and disease with no other likely explanation. The more specific an association between a factor and an effect is, the bigger the probability of a causal relationship.

The Bradford Hill criteria for causation

- Temporality: The effect has to occur after the cause (and if there is an expected delay between the cause and expected effect, then the effect must occur after that delay).
- Biological gradient: Greater exposure should generally lead to greater incidence of the effect. However, in some cases, the mere presence of the factor can trigger the effect. In other cases, an inverse proportion is observed: greater exposure leads to lower incidence.
- Plausibility: A plausible mechanism between cause and effect is helpful (but Hill noted that knowledge of the mechanism is limited by current knowledge).

The Bradford Hill criteria for causation

- Coherence: Coherence between epidemiological and laboratory findings increases the likelihood of an effect. However, Hill noted that "... lack of such [laboratory] evidence cannot nullify the epidemiological effect on associations".
- Experiment: "Occasionally it is possible to appeal to experimental evidence".
- Analogy: The effect of similar factors may be considered.

A causal question

At age 45, Ms. Smith is diagnosed with stage II stomach cancer.

- Her oncologist Rachael discusses with her two possible treatments: (i) surgery alone, or (ii) surgery + chemo. They decide on (ii).
- Ten years later, Ms. Smith is alive and the tumor has not recurred.
- Her surgeon, Steve, and Rachael debate.
  - Rachael says: "The chemo prevented the recurrence - without it, the tumor would have recurred."
  - Steve says: "You can't know that. It's a fantasy - you're making it up. We'll never know."

Two notions of causation

- Causes of an effect/outcome
  What are causes of lung cancer?

  Historian/detective paradigm

- Effects of a cause
  Does smoking cause lung cancer?

  Experimental paradigm

For this course, we concentrate on effects of a cause/intervention.

Association versus causality

- Most scientific inquiry/data analyses have one of two goals:
  - Association/prediction, i.e., determine predictors or variables associated with the outcome of interest.
  - Causality, i.e., understand factors that cause or have an effect on the outcome of interest.
- We are often told that association is not causation;
  - when can we claim an association is causation?
- The area of causal inference provides the answer.

Whether an association is causal

Typical approach for determining whether an association is causal:

1. Observe treatments/exposures, outcomes, and other variables in population
2. Use standard statistical methods to derive inferences about associations between observable variables
3. Use series of *ad hoc* rules to determine whether associations have causal interpretation

Useful to have more precise definitions of causal effects!

An intuitive definition of cause

- Jim took the pill on Sept 1, 2001
  - Five days later, he was dead
- Had Jim not taken the pill on Sept 1, 2001 (all others things being equal)
  - Five days later, he would have been alive
- Did the pill cause Jim's death?

Estimands: asking the right question

A fellow decides to wallpaper a bedroom in his house, but he isn't sure how many rolls of wallpaper he'll need. He knows that his next-door neighbor had recently done the same job, and the two rooms are identical in size.

"So, how many rolls of wallpaper did you buy for your bedroom?" the man asks his neighbor.

"Ten," says the neighbor.

Estimands: asking the right question

The man buys 10 rolls of paper and does the job. When it's finished, it looks wonderful, but he has four rolls of wallpaper left over.

"Hey, neighbor," he says. "I bought 10 rolls of wallpaper for the bedroom, but I've got four left over."

"That's funny," says the neighbor. "X XXX XXX"

Definitions for causal inference

- **Unit**: Person, place, or thing upon which a treatment will operate
- **Treatment**: An intervention or exposure investigators wish to assess the effect of treatment compared to no treatment
- Treatment or intervention should be manipulable
  Does anti-retroviral therapy affect CD4 count in HIV+ individuals?
  Does caloric intake affect CD4 count?
  Does obesity affect CD4 count?
  Does gender affect CD4 count?

Definitions for causal inference

- **Potential outcomes**: Values of a unit's measurement of interest when (i) treatment received and (ii) treatment not received

  Potential outcomes often referred to as counterfactuals

- **Individual causal effect**: For each unit, comparison between potential outcome under treatment and potential outcome under control

Potential outcomes

- Let a denote whether treatment received
  - $a = 1$ for treatment
  - $a = 0$ for no treatment (or "control")

- Let $Y_i(a)$ denote the potential outcome for unit i given treatment a
  - notation $Y_i(a), y_{ia}, Y_{ia}$ etc also common
  - Often subscript i is dropped: $y(a), Y(a)$, etc
  - Sometimes a_1 denotes $a = 1$, e.g., $Y(1) = Y(a_1)$

- Each individual has a set of potential outcomes $\{Y_i(0), Y_i(1)\}$

Individual causal effect

- Individual causal effect (ICE) is some comparison between individual potential outcomes, e.g.

$$Y_i(0) - Y_i(1) \text{ or } 1 - \frac{Y_i(1)}{Y_i(0)}$$

- Potential outcomes model of causation; Neyman-Rubin causal model
- Fundamental problem of causal inference:
  Cannot observe both potential outcomes for same person

Actual or observed data

- $Y = 1$ if patient died, 0 otherwise
- $A = 1$ if patient treated, 0 otherwise

ID	A	Y
Ian	1	1
Jim	0	0

Ideal data

ID	A	Y	Y(0)	Y(1)
Ian	1	1	1	1
Jim	0	0	0	0

Available data on Ian and Jim

ID	A	Y	Y(0)	Y(1)
Ian	1	1	?	1
Jim	0	0	0	?

- If untreated, $A = 0$, $Y = Y(0)$ and $Y(1)$ is missing
- If treated, $A = 1$, $Y = Y(1)$ and $Y(0)$ is missing
- Counterfactual and actual outcomes are linked via $Y(A) = Y$ (More details later!)

Your table vs God's table

	Yours		God's		
ID	A	Y	A	Y(0)	Y(1)
1	0	0	0	0	0
2	0	1	0	1	1
3	0	0	0	0	0
4	0	0	0	0	0
5	1	1	1	1	1
6	1	1	1	1	1
7	1	0	1	0	0
8	1	0	1	0	0

- Yours:
  $Pr(Y = 1|A = 1) - Pr(Y = 1|A = 0) = 2/4 - 1/4 = 1/4$
- God's: $Pr(Y(1) = 1) - Pr(Y(0) = 1) = E(Y(1) - Y(0)) = 0$
- Association $\neq$ Causation: Confounding for the effect of A on Y

Average causal effect

- Association (identifiable): crude risk difference

$$\delta = E(Y|A=1) - E(Y|A=0) = E(Y(1)|A=1) - E(Y(0)|A=0)$$

- Causation (not identifiable): causal risk difference = average causal effect (ACE)

$$\delta^* = E(Y(1) - Y(0)) = E(Y(1)) - E(Y(0))$$

proportion diseased if all treated - proportion diseased if all untreated

- ACE, instead of ICE, will be the main focus of this short course with the exception of Fisher's randomization procedure.

Reinforcing the notations

- $Y(1)$ and $Y(0)$ are potential outcomes
- Y is the observed outcome
- There may be no link among the three.
- Need assumptions to link them.
- One key step in causal inference is to delineate these assumptions.

Observed quantities

- **Data**: $Z_i = (Y_i, A_i, X_i), i = 1, \ldots, n$, where for i^{th} individual,
  Y_i : response
  A_i : the treatment received
  X_i : the covariates that have been collected on the individual prior to the intervention.

- **Interest**: establish a causal relationship between the intervention and the outcome

Statistical paradigm

- $Y|A = 0 \sim N(\mu_0, \sigma^2)$
  $Y|A = 1 \sim N(\mu_1, \sigma^2)$
- Population: patients with hypertension who have diastolic blood pressure greater than 140.
- A sample from this population are studied, some who receive statin drugs ($A = 1$) and others who don't ($A = 0$); Y: the change in a patient's blood pressure after 6 months
- $\delta = \mu_1 - \mu_0$: the observed difference in mean blood pressure for patients receiving treatment versus those that don't

Association

- Associational relationship!
- Individuals who receive treatment may be inherently different from those who do not.
- δ: reflect such inherent differences as well as any effect of treatment.
- **Confounding variables:** factors that may be related both to outcome as well as who may get treatment.

Main task

Under certain assumptions, causation is identifiable.

Main focus: estimate $E(Y(1))$ and $E(Y(0))$, which must be done using the observed data.

Causal parameters or estimands

- In causal inference, careful attention is given to the estimands ("science", "nature")
- For a population of N units, average causal effect (ACE)

$$\frac{1}{N}\sum_{i=1}^{N}\{Y_i(1) - Y_i(0)\}$$

- Median causal effect

$$\text{Median } \{Y_i(1) - Y_i(0) : i = 1, ..., N\}$$

Causal parameters or estimands

- Difference in median potential outcomes

$$\text{Median } \{Y_i(1) : i = 1, ..., N\} - \text{Median } \{Y_i(0) : i = 1, ..., N\}$$

- Critical requirement is comparison is on the same set of units

$$\text{Median } \{Y_i(1) : i \in S\} - \text{Median } \{Y_i(0) : i \in T\}$$

not a causal effect unless $S = T$

Other causal estimands

- Conditional average causal effects also common
- Average treatment effect in the treated

$$E\{Y(1) - Y(0)|A = 1\} \equiv ATET$$

and avg treatment effect in the controls

$$E\{Y(1) - Y(0)|A = 0\} \equiv ATEC$$

Stable Unit Treatment Value Assumption (SUTVA)

- That the notation $\{Y_i(0), Y_i(1)\}$ sufficiently describes all possible potential outcomes requires SUTVA

- 1. Consistency: Value $Y_i(a)$ for unit i when exposed to treatment a will be the same no matter what mechanism is used to assign treatment a to unit i; no different versions of treatment

$$Y_i = Y_i(1)A_i + Y_i(0)(1 - A_i)$$

Not to be confused with the statistical concept of an estimator being consistent (i.e., converges in probability)

Stable Unit Treatment Value Assumption (SUTVA)

Individual i receives treatment 1 in a hypothetical situation where

- everyone in the population is given treatment 1
- a sample of individuals are randomized to receive treatment 1 or 0
- some individuals receive treatment 1 or 0 according to their physician

Observed response would be the same for patient i.

SUTVA

- 2. No interference between units: Outcome of one individual assumed to be unaffected by treatment assignment of others

  Violated in infectious diseases (vaccines)?

- Without SUTVA things are potentially much more complicated

- Need to expand notation to reflect the dependence of the outcomes on treatment of others or different possible treatments

Nonparametric bounds

- Identification (statistical) of parameter: the parameter can be calculated (correctly) from the joint distribution of observable quantities
- Here: we would like to be able to learn something about the potential outcomes $Y(a)$:
  - Distributions and expectations of $Y(a)$
  - Comparisons (e.g., differences or ratios of expectations; causal risk differences or ratios)

Nonparametric bounds

- Problem: potential outcomes $Y(1)$, $Y(0)$, not observed for every subject in population
- What can be stated about these sometimes unobserved quantities?
  - No additional assumptions: bounds – introduce briefly

Nonparametric bounds

- Bounds: consider binary Y
- What is largest value for $Pr(Y(1) = 1)$ that is consistent with data? What is smallest?
- What about $Pr(Y(0) = 1)$?
- What is maximum value for causal risk difference $Pr(Y(1) = 1) - Pr(Y(0) = 1)$?
  (Consider inference either for finite population (no sampling) or sample of infinite size)

Nonparametric bounds

- Can expand probability:

$$\begin{aligned} Pr(Y(0)=1) &= Pr(A=0, Y(0)=1) + Pr(A=1, Y(0)=1) \\ &= Pr(A=0)Pr(Y(0)=1|A=0) \\ &\quad + Pr(A=1)Pr(Y(0)=1|A=1) \\ &= Pr(A=0)Pr(Y=1|A=0) \\ &\quad + Pr(A=1)Pr(Y(0)=1|A=1) \end{aligned}$$

- What quantities above are observable, estimable from data?
- Where is the SUTVA used?

Nonparametric bounds

- All quantities known or estimable except $Pr(Y(0) = 1|A = 1)$
- What about bounds on $Pr(Y(0) = 1|A = 1)$?

  Limits for $Pr(Y(0) = 1|A = 1)$: $0 \leq Pr(Y(0) = 1|A = 1) \leq 1$

Nonparametric bounds

- Limits for entire population: i.e., on $Pr(Y(0) = 1)$:

$$Pr(A = 0)Pr(Y = 1|A = 0) \leq Pr(Y(0) = 1) \leq Pr(A = 0)Pr(Y = 1|A = 0) + Pr(A = 1)$$

- For binary outcome, can learn something about potential outcomes without making any further assumptions about unobserved quantities

- For other potential outcome $Y(1)$

$$Pr(A = 1)Pr(Y = 1|A = 1) \leq Pr(Y(1) = 1) \leq Pr(A = 0) + Pr(A = 1)Pr(Y = 1|A = 1)$$

Nonparametric bounds

- How can one construct bounds on risk difference?
- For lower bound, take largest value for $Pr(Y(0) = 1)$ consistent with data, smallest value for $Pr(Y(1) = 1)$

$$\begin{aligned} Pr(Y(1) = 1) - Pr(Y(0) = 1) & \\ \geq \quad & Pr(A = 1)Pr(Y = 1|A = 1) \\ & -Pr(A = 0)Pr(Y = 1|A = 0) - Pr(A = 1) \end{aligned}$$

- Similarly, for upper bound:

$$\begin{aligned} Pr(Y(1) = 1) - Pr(Y(0) = 1) & \\ \leq \quad & Pr(A = 1)Pr(Y = 1|A = 1) \\ & +Pr(A = 0) - Pr(A = 0)Pr(Y = 1|A = 0) \end{aligned}$$

Nonparametric bounds

- Difference between upper and lower bounds on Pr(Y(1)=1) - Pr(Y(0)=1) is 1
- So half of theoretical range (-1 to 1; range of 2) excluded by observable variables with no other assumptions
- No matter how large study, how much data obtained, this approach cannot get narrower range for causal risk difference
- Can never exclude null hypothesis of no treatment effect

Nonparametric bounds

- How would one derive inference from sample about population?
- One could calculate sampling variances of bounds
  Will generally lead to even wider intervals
- Nonparametric bounds are generally fairly broad; More precise inference desired
- Suggest methods (very general) for obtaining inferences

Nonparametric bounds

- To get more precise and informative inference, two approaches:
  1. Consider assumptions that allow tightening of nonparametric bounds approach, e.g. Manski (1990)
  2. Specify assumptions sufficient to identify causal parameters/effects; consider how to weaken/make assumptions more reasonable.

  Approach 2 will often justify familiar approaches; more widely used in practice in epidemiology and biostatistics

Nonparametric bounds

- Most of course will follow approach 2, will consider the following different settings:
  1. Randomized trials without noncompliance
  2. Simple observational studies

Short Course: Causal Inference

Randomized Experiments

Menggang Yu

Summer 2014

Simple examples

- Example 1 - Additive

i	$Y_i(1)$	$Y_i(0)$	$Y_i(1) - Y_i(0)$
1	10	7	3
2	3	0	3
3	5	2	3
4	12	9	3
Mean	7.5	4.5	3

- Example 2 - Non-additive

i	$Y_i(1)$	$Y_i(0)$	$Y_i(1) - Y_i(0)$
1	10	10	0
2	3	0	3
3	5	5	0
4	12	3	9
Mean	7.5	4.5	3

Additivity

- Causal effect is additive if

$$Y_i(1) - Y_i(0) = \delta \text{ for all } i$$

  where δ is some constant

- Example 1 additive, Example 2 non-additive

- Additivity assumption plays important role in some of the results of Neyman and Fisher below

Assignment mechanism

- In order to learn about the causal effects of interest, we need to know or posit an assignment mechanism
- **Assignment mechanism:** The process of deciding which units receive treatment and which receive control
- Define $Y^{obs} \equiv Y(A) = (1 - A)Y(0) + AY(1)$ (consistency)
- Assigned treatment unmasks the potential outcome $Y(A) = Y^{obs}$ but masks the other potential outcome $Y(1 - A) = Y^{mis}$

Completely randomized experiment

- Exactly k of N units receive treatment, with each assignment equally likely
- Let $A \equiv (A_1, ..., A_N)$ and $a \equiv (a_1, ..., a_N)$
- For completely randomized experiment

$$Pr[A = a] =$$

- Unconfounded or strongly ignorable assignment mechanism, also called full exchangeability (Hernan and Robins 2013)

The randomization assumption

- Data from a randomized intervention study can be used to derive an unbiased estimation of **causal** treatment effect

- Suppose you randomize the population of patients to either treatment with probability $p > 0$ or to control with probability $1 - p > 0$. Then, with $\amalg$ denoting independence, it holds that

$$Y(a) \amalg A, a = 0, 1$$

or

$$(Y(0), Y(1)) \amalg A$$

Randomization leads to causality

- In this case

$$\begin{aligned} E(Y|A=1) &= E(Y(1)A + Y(0)(1-A)|A=1) \\ &= E(Y(1)|A=1) \\ &= E(Y(1)). \end{aligned}$$

- Thus, the probability distribution of the counterfactuals $Y(a), a = 0, 1$, can be written in terms of the distribution of the observed data (Y, A) and hence it is identified.

- Hence $\delta^* = \delta = E(Y|A=1) - E(Y|A=0)$.

The null hypothesis of no treatment effect

- Note that randomization does not imply $Y \amalg A$ since $Y = AY(1) + (1 - A)Y(0)$ is determined by treatment and therefore is a posttreatment variable.

- $Y \amalg A$: represents the null hypothesis of no treatment effect.

$$Y \quad = \quad Y(1)A + Y(0)(1 - A) = A\{Y(1) - Y(0)\} + Y(0)$$

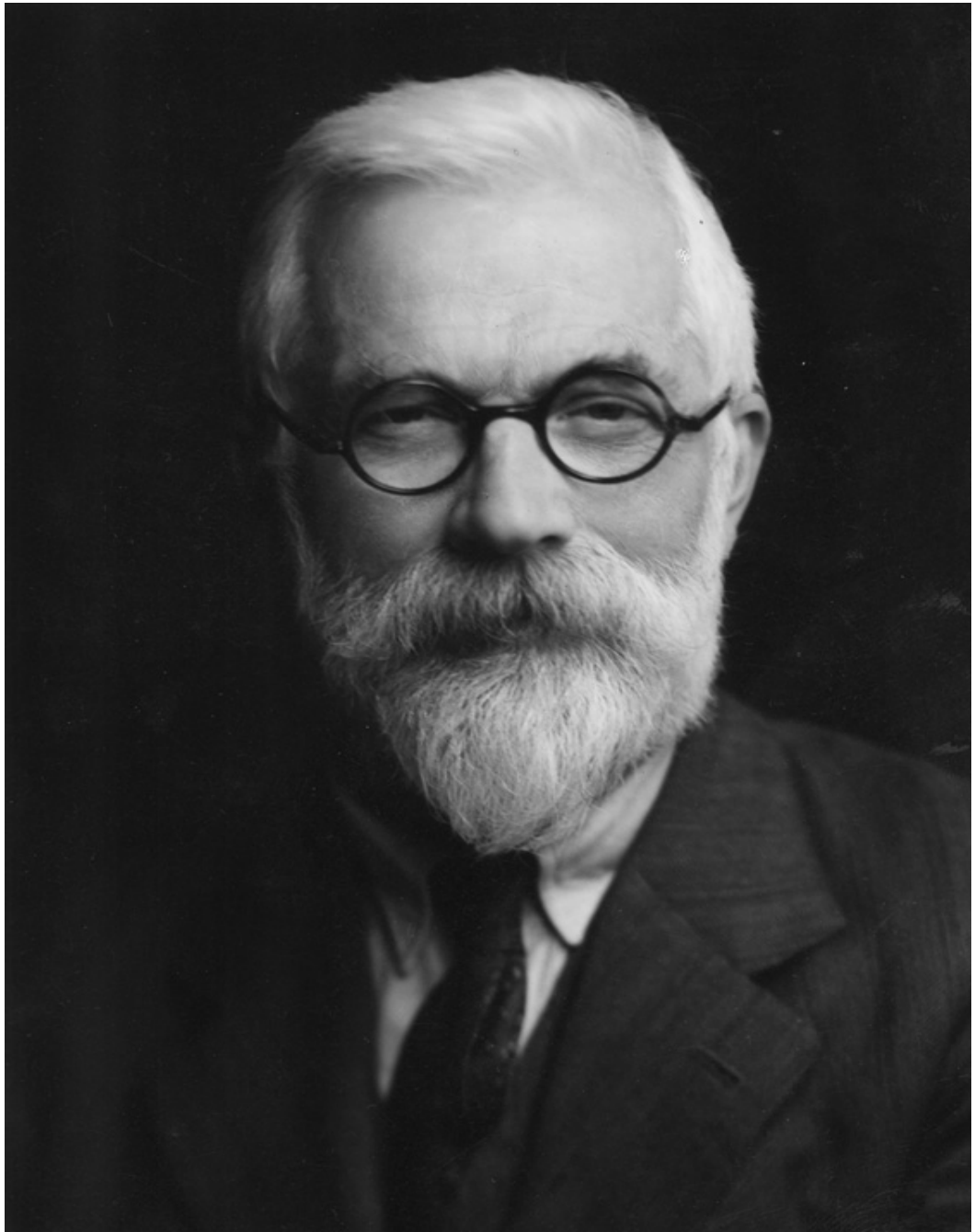
Therefore $Y \amalg A$ iff $Y(1) = Y(0)$.

Causal Inference in randomized experiments

- Randomization-based inference
  - Neyman 1923 lays out potential outcome framework
  - Fisher 1925 realizes importance of randomization
- Model-based inference
  - Frequentist
  - Bayesian

Neyman

RA Fisher



Causal Inference in randomized experiments

- Neyman approach – based on ACE:

$$H_0 : ACE = \sum_{i=1}^{N} \{Y_i(1) - Y_i(0)\} = 0$$

Not a "sharp" null hypothesis.

- Fisher approach – based on ICE:

$$H_0 : Y_i(0) = Y_i(1), \forall i = 1, \ldots, N$$

No treatment effect on any unit

$\Rightarrow$ A "sharp" null hypothesis – specifies all potential outcomes

Neyman

- Suppose we have a completely randomized experiment
- Then an unbiased estimator of ACE

$$\widehat{ACE} \equiv \frac{\sum_i Y_i^{obs} A_i}{\sum_i A_i} - \frac{\sum_i Y_i^{obs}(1 - A_i)}{\sum_i (1 - A_i)}$$

- Expectation is taken over all possible randomizations
- Potential outcomes $\{Y_i(0), Y_i(1) : i = 1, ..., N\}$ regarded as fixed
- Only treatment assignments A_i are considered random
- Randomization-based repeated-sampling inference

Neyman

- Draw simple random sample without replacement of size k from population of size N

- Then the sample average

$$\frac{\sum_i Y_i^{obs} A_i}{\sum_i A_i} = \frac{1}{k} \sum_{i:A_i=1} Y_i(1)$$

  is an unbiased estimator of the population average

$$\frac{1}{N} \sum_{i=1}^{N} Y_i(1) = \bar{Y}(1)$$

- Prove $E(\widehat{ACE}) = ACE$.

Example 1 revisited

- Example 1

i	$Y_i(1)$	$Y_i(0)$	$Y_i(1) - Y_i(0)$
1	10	7	3
2	3	0	3
3	5	2	3
4	12	9	3
Mean	7.5	4.5	3

- EXERCISE: Completely randomized experiment where $k = 2$

a	$Pr[A = a]$	$\widehat{ACE}$
0011	1/6	(5+12)/2 -(7+0)/2=5
1010	1/6	
1001	1/6	
0110	1/6	
0101	1/6	
1100	1/6	
Mean		

Neyman

- Estimator of variance

$$\widehat{Var}(\widehat{ACE}) \equiv \frac{\hat{\sigma}_1^2}{k} + \frac{\hat{\sigma}_0^2}{N-k}$$

where $\hat{\sigma}_j^2$ is the sample variance of $\{Y_i^{obs} : A_i = j\}$ for $j = 0, 1$,

- Positively biased unless effect is additive, i.e.,

$$E\{\widehat{Var}(\widehat{ACE})\} = Var(\widehat{ACE}) + \frac{1}{N}\sigma_{1-0}^2$$

where $\sigma_{1-0}^2 \equiv \sum_i [\{Y_i(1) - Y_i(0)\} - \{\bar{Y}(1) - \bar{Y}(0)\}]^2/(N-1)$

Example 1 revisited

- Example 1: additive causal effect

i	$Y_i(1)$	$Y_i(0)$	$Y_i(1) - Y_i(0)$
1	10	7	3
2	3	0	3
3	5	2	3
4	12	9	3

- Completely randomized experiment where $k = 2$

a	$Pr[A = a]$	$\widehat{ACE}$	$\widehat{Var}(\widehat{ACE})$
0011	1/6	(5+12)/2 -(7+0)/2=5	24.5
1010	1/6	(10+5)/2-(0+9)/2=3	26.5
1001	1/6	(10+12)/2-(0+2)/2=10	2.0
0110	1/6	(3+5)/2 -(7+9)/2=-4	2.0
0101	1/6	(3+12)/2-(7+2)/2=3	26.5
1100	1/6	(10+3)/2-(2+9)/2=1	24.5
Mean		3	17.67

Example 2 revisited

- Example 2: non-additive causal effect

i	$Y_i(1)$	$Y_i(0)$	$Y_i(1) - Y_i(0)$
1	10	10	0
2	3	0	3
3	5	5	0
4	12	3	9

- Completely randomized experiment where $k = 2$

a	$Pr[A = a]$	$\widehat{ACE}$	$\widehat{Var}(\widehat{ACE})$
0011	1/6	3.5	37.25
1010	1/6	6.0	8.5
1001	1/6	8.5	7.25
0110	1/6	-2.5	13.25
0101	1/6	0.0	26.5
1100	1/6	2.5	13.25
Mean		3	17.67
Variance*		13.17	

Large sample approximation

- Assuming additivity, as $k, (N-k) \to \infty$

$$\frac{\widehat{ACE} - ACE}{\widehat{Var}(\widehat{ACE})^{1/2}} \to N(0,1)$$

- For large samples, $(1-\alpha)\%$ confidence interval

$$\widehat{ACE} \pm z_{1-\alpha/2} \widehat{Var}(\widehat{ACE})^{1/2}$$

- Hypothesis testing: $H_0 : ACE = 0$

$$Z_{obs} = \frac{\widehat{ACE}}{\widehat{Var}(\widehat{ACE})^{1/2}} \sim N(0,1)$$

Neyman summary

- Credited with founding potential outcomes model of causality
- Derived unbiased estimator of ACE and
  (positively biased) estimator of variance of ACE
- Leads to valid large sample confidence intervals
- Assumes only SUTVA and completely randomized experiment

Extensions of Neyman's potential outcomes

- A may take on multiple levels, say 0, 1, 2, ... such that each individual has potential outcomes $\{Y_i(0), Y_i(1), Y_i(2)\}$
- Similarly, may have treatment at multiple time points, eg: $Y_i(a_1, a_2)$ potential outcome when treatment a_j received at time j
- A continuous, $\{Y_i(a)\}$ infinite dimensional
- $Y_i(a)$ random instead of deterministic function of a
- Interference: Outcome for individual i may depend on treatment assignment of others $Y_i(a)$

Fisher: motivating example

- Suppose want to understand vaccine's ability to prevent infection/disease from some pathogen $a = 1$ vaccine, $a = 0$ control; $Y = 1$ infection, $Y = 0$ uninfected

- Four types of individuals

$Y_i(0)$	$Y_i(1)$	ICE	
0	0	0	Immune
1	0	1	Protected
0	1	-1	Harmed
1	1	0	Doomed

- Fundamental problem of causal inference: cannot say to which stratum individuals belong, do not know ICE or ACE

Observed table - Fisher

- Conduct completely randomized experiment; observe 2×2 table

	$Y = 1$	$Y = 0$	
Vaccine	$\sum_i A_i Y_i^{obs}$	$\sum_i A_i(1 - Y_i^{obs})$	$\sum_i A_i$
Placebo	$\sum_i (1 - A_i) Y_i^{obs}$	$\sum_i (1 - A_i)(1 - Y_i^{obs})$	$\sum_i (1 - A_i)$
	$\sum_i Y_i^{obs}$	$\sum_i (1 - Y_i^{obs})$	n

- Vesikari et al. (1990) randomized $n = 200$ infants aged 2-5 months to rotavirus vaccine or placebo

	Diarrhea $Y = 1$	NoDiar $Y = 0$	
Vaccine	37	63	100
Placebo	52	48	100
	89	111	200

Causal table - Fisher

For completely randomized experiment, under sharp null the observed 2×2 table

	$Y = 1$	$Y = 0$	
Vaccine	$\sum_i A_i Y_i^{obs}$	$\sum_i A_i(1 - Y_i^{obs})$	$\sum_i A_i$
Placebo	$\sum_i (1 - A_i) Y_i^{obs}$	$\sum_i (1 - A_i)(1 - Y_i^{obs})$	$\sum_i (1 - A_i)$
	$\sum_i Y_i^{obs}$	$\sum_i (1 - Y_i^{obs})$	N

equivalent to

	$Y = 1$	$Y = 0$	
Vaccine	$\sum_i A_i Y_i(0)$	$\sum_i A_i(1 - Y_i(0))$	n
Placebo	$\sum_i (1 - A_i) Y_i(0)$	$\sum_i (1 - A_i)(1 - Y_i(0))$	$N - n$
	$\sum_i Y_i(0)$	$\sum_i (1 - Y_i(0))$	N

Fisher's randomization or exact test

- Assuming H_0 and completely randomized experiment, row and column totals are fixed
- $\sum_i A_i Y_i(0)$ is the total of a simple random sample (without replacement) of size k from a population of size N with $\sum_i Y_i(0)$ ones and $\sum_i (1 - Y_i(0))$ zeros
- $\sum_i A_i Y_i(0)$ has a hypergeometric distribution

Hypergeometric distribution

From Wiki: the hypergeometric distribution is a discrete probability distribution that describes the probability of k successes in n draws without replacement from a finite population of size N containing exactly K successes.

	success	no success	
drawn	k	$n-k$	n
not drawn	$K-k$	$N-K-n+k$	$N-n$
	K	$N-K$	N

$$P(X=k)=\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

Fisher's randomization or exact test

- Veiskari data: one-sided p-value $\Pr[\sum_i A_i Y_i(0) \leq 37] = 0.023$
- In R

```
phyper(37,89,111,100)

phyper(q, m, n, k, lower.tail = TRUE, log.p = FALSE)

q   vector of quantiles representing the number of white balls
     drawn from an urn which contains both black and white balls.

m  the number of white balls in the urn.

n   the number of black balls in the urn.

k   the number of balls drawn from the urn.

p   probability, it must be between 0 and 1.
```

Randomization-based inference

- Before looking at the observed data:
  - Specify a sharp null hypothesis
  - Specify a test statistic for estimating the treatment effect and evaluating the null hypothesis.

- Using the Y_{obs} data
  - Calculate the value of the test statistic, and specify values that are more extreme (i.e., unusual).
  - Fill in the missing potential outcomes using the sharp null hypothesis and the observed Y values.

- Obtaining a p-value
  - For each possible assignment, calculate the value of the test statistic of interest that would have been observed under that assignment .
  - Determine how extreme the value observed is.

Permutation test

- Test statistic $D = \widehat{ACE}$
- Under $H_0 : Y_i(1) - Y_i(0) = 0$, randomization-based distribution of D is computed by considering all possible treatment assignments k
- Under H_0, can compute D for each of these randomization assignments
- For completely randomized experiment, each possible randomization assignment is equally likely

Permutation test: example

- A new drug is being tested in humans for the first time to assess effect on CD4+ T cells in patients with HIV
- 7 individuals are randomized to 2 groups: control ($N - k = 3$) or drug ($k = 4$)
- Outcome $Y_i(a)$: percent increase in CD4+ count from baseline for $a = 0$ control, $a = 1$ drug
- Null hypothesis is the drug has no effect

$$H_0 : Y_i(1) - Y_i(0) = 0 \text{ for } i = 1, \ldots, N$$

Permutation test: example

- Observed data: control (65, 69, 73); drug (70, 88, 89, 92)
- There are 35 possible randomizations
  i.e. there are 35 possible ways to assign 4 individuals to drug
- Under H_0, can compute D for each of these randomization assignments

Example: all possible group assignments

$A=0$	$A=1$	d	$A=0$	$A=1$	d
65 69 73	70 88 89 92	15.75	65 69 70	73 88 89 92	17.50
65 69 89	70 88 73 92	6.42	65 69 88	70 73 89 92	7.00
...	...	...	...	...	...

Example: CDF of D

d	$P(D \leq d)$	d	$P(D \leq d)$
-20.42	0.029	1.17	0.600
-11.08	0.057	1.75	0.629
-10.67	0.086	2.33	0.657
-9.92	0.114	3.50	0.686
-9.33	0.143	4.08	0.714
-8.75	0.200	4.67	0.743
-7.58	0.229	5.83	0.771
-7.00	0.286	6.42	0.800
-6.41	0.371	7.00	0.886
-6.00	0.400	12.83	0.914
-4.67	0.457	15.17	0.943
-0.58	0.543	15.75	0.971
-0.00	0.571	17.50	1.000

Permutation test example

- Observed $d = -15.75$
- Two-sided p-value

$$Pr(|D| \geq 15.75) = 3/35 = 0.086$$

- In R

```
library(exactRankTests)
perm.test(c(65,69,73),c(70,88,89,92),exact=T,tol=0.01)
```

Randomization-based inference

- Fisher Interval: The set of possible values of the causal quantity of interest corresponding to test statistics with p-values that fall within some range set by the researcher; usually interpreted to be a set of plausible values of the average causal effect

Example - explaining the Fisher interval

Jane is 23 years younger than her mother.

Jane's parents' ages sum to 58.

Jane's mother is two years younger than Jane's father.

How old is Jane?

Example - explaining the Fisher interval

- Start by assuming Jane is 30.
- This means Jane's mother must be 53, which means Jane's father must be 5. Contradiction!
- Our assumption that Jane is 30 must be wrong.

Example - explaining the Fisher interval

Try again!

- Assume Jane is 10.
- This means that Jane's mother must be 33, which means Jane's father must be 25.
- Our assumption that Jane is 10 must be wrong.

Example - explaining the Fisher interval

Keep repeating the process until you don't arrive at a contradiction.

- Assume Jane is 5.
- This means that Jane's mother must be 28, which means Jane's father must be 30.
- We cannot reject the assumption that Jane is five years old.

Randomization-based confidence interval

- Use p-values to determine an interval of plausible numbers for the treatment effect.
- Systematically go through hypothesized values for the treatment effect.
- Keep the values that are "plausible", for example, their corresponding p-values are between .05 and .95; discard others
- The ones we keep form a 90% Fisher interval.

Randomization-based confidence interval

- Suppose causal effect is additive

$$Y_i(1) - Y_i(0) = \delta \text{ for } i = 1, \ldots, n$$

- Consider test of $H_0 : Y_i(1) - Y_i(0) = \delta_0$ for some $\delta_0 \neq 0$
- Set of δ_0 where we fail to reject at the α significance level forms a $(1 - \alpha) \times 100\%$ confidence interval for δ.

Randomization-based confidence interval

- Want to test $H_0 : Y_i(1) - Y_i(0) = \delta_0$ for some $\delta_0 \neq 0$
- Instead of Y_i^{obs}, perform randomization test on $Y_i^{obs} + (1 - A_i)\delta_0$ which is independent of A because under H_0

  $$Y_i^{obs} + (1 - A_i)\delta_0 = (1 - A_i)Y_i(0) + A_i Y_i(1) + (1 - A_i)\delta_0 = Y_i(1)$$

- Confidence intervals are "exact" in the sense that the coverage probability is guaranteed to be at least $(1 - \alpha)$
- E.g., in R:

```
perm.test(c(65,69,73),c(70,88,89,92),exact=T,tol=0.01,
conf.int=TRUE)
```